ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

Identifying the roles of hub gene in keloid formation using single-cell transcriptomics

Xiangbing Zheng[1†], Zhengqiang Wan[2†], Bing Liu[1], Jun Yin[3], Cheng Chen[1], Yuzhen Ma[4], and Yong Zou[1*]

[1]Department of Burns and Plastic Surgery, The Second People's Hospital of Yibin, Sichuan, China

[2]Department of Thoracic Surgery, The First People's Hospital of Suining, Suining, Sichuan, China

[3]Department of Haematology, The Second People's Hospital of Yibin, Yibin, Sichuan, China

[4]Department of Oral and Maxillofacial Surgery, The First People's Hospital of Suining, Suining, Sichuan, China

Abstract

Introduction: Keloid, a fibroproliferative tumor characterized by excessive collagen deposition and fibroblast hyperplasia, lacks effective therapeutic strategies due to unclear molecular mechanisms.

Objective: This study aims to elucidate keloid pathogenesis and identify diagnostic biomarkers through multi-omics integration.

Methods: Single-cell RNA sequencing (ScRNA-seq) data (GSE163973) and bulk RNA sequencing datasets (GSE162904/GSE145725) were analyzed. Fibroblast subpopulations were identified using the Seurat R package, and cell–cell interactions were explored using the CellChat R package. Weighted gene co-expression network analysis (WGCNA) was employed to identify key gene modules in fibroblasts. Hub genes were screened using Lasso regression and validated through machine learning algorithms and a gene-immune convolutional neural network (CNN). Immune infiltration patterns were evaluated using the MCP-counter and Immuno-Oncology Biological Research R packages.

Results: ScRNA-seq analysis revealed eight distinct cell subtypes within keloid tissues, with fibroblasts significantly enriched compared to normal skin. Fibroblast clusters 1 and 5 exhibited elevated midkine–low-density lipoprotein receptor-related protein 1-mediated interactions and enhanced differentiation activity. WGCNA identified three critical modules—"brown," "cyan," and "yellow"—linked to fibroblast activation. Lasso regression produced an eight-gene signature that effectively distinguished keloid from normal skin (area under the curve = 0.885 – 0.889). Nonnegative matrix factorization classified keloids into four subtypes, each with distinct immune infiltration profiles correlated with hub gene expression. The gene-immune CNN model achieved 100% sensitivity and 88.9% specificity in diagnostic classification.

Conclusion: This study elucidates the molecular mechanisms underlying keloid formation through integrated single-cell and transcriptomic analysis, proposing an eight-gene signature as a potential diagnostic and therapeutic target. The identified keloid subtypes and associated immune infiltration patterns provide novel insights for advancing precision medicine approaches in keloid management.

Keywords: Cell communication; Deep learning; Differentiation trajectory; Immune infiltration; Keloid; Single-cell RNA

†*These authors contributed equally to this work.*

***Corresponding author:**
Yong Zou
(zy18090968815@163.com)

Citation: Zheng X, Wan Z, Liu B, Yin J, Chen C, Ma Y, Zou Y. Identifying the roles of hub gene in keloid formation using single-cell transcriptomics. *Eurasian J Med Oncol*. 2026;10(1):133-147. doi: 10.36922/EJMO025150098

Received: April 8, 2025

Revised: May 31, 2025

Accepted: June 13, 2025

Published online: July 3, 2025

1. Introduction

Keloids are pathological scars that extend beyond the original wound margins and continue to grow invasively. They are benign fibroproliferative mesenchymal tumors that often develop following surgical excision, characterized histologically by fibroblast proliferation, excessive collagen deposition, and mucopolysaccharide accumulation.[1] Clinically, keloids can cause pain, sensory abnormalities, itching, and functional impairment, in addition to disfiguring the skin, leading to both physical and psychological distress for affected individuals.[2]

Despite their clinical significance, the molecular mechanisms underlying keloid formation remain poorly understood.[3,4] Keloid tissue contains a significant accumulation of collagen fibers, which constitute the bulk of the lesion. During keloid formation, fibroblasts—the primary effector cells—exhibit disorganized distribution and may undergo senescence or necrosis.[5] Compared to fibroblasts in normal skin, keloid fibroblasts (KFs) produce higher levels of extracellular matrix (ECM) components[6] and readily differentiate into myofibroblasts.[7,8] These myofibroblasts express fibronectin, vimentin, and α-smooth muscle actin, with their nonmuscle actin-myosin microfilaments contributing to wound healing and scar formation.[9,10] Additionally, KFs influence keratinocyte proliferation and impair epithelial regeneration.[11] Given their role in keloid pathology, fibroblasts represent a key therapeutic target in managing keloid disorders.

Keloid is a distinct type of scar with several characteristics that set it apart from other scar types:

(i) Keloids exhibit uncontrolled growth and can infiltrate surrounding normal tissue, extending beyond the original wound margins
(ii) Keloid tissue is highly vascularized, with an abnormal abundance of capillaries
(iii) KFs demonstrate strong anti-apoptotic abilities
(iv) Keloids are closely associated with tumor-related genes and cytokines, such as tumor growth factor (TGF)-β, vascular endothelial growth factor, connective tissue growth factor, and various interleukins
(v) The underlying gene regulatory mechanisms of keloid pathogenesis show similarities to those observed in tumors
(vi) Several tumor treatment strategies have demonstrated therapeutic effects against keloid

These characteristics suggest that keloids share multiple characteristics with malignant tumors.[12]

Keloids also share several clinical features with tumors, including hyperplastic growth that protrudes above the skin surface, progressive invasion into the surrounding undamaged dermis, lack of clear boundaries with adjacent normal skin, persistence over time without regression, and a high rate of recurrence following surgical excision. Similar to tumors with hereditary and familial patterns, keloids also exhibit genetic predisposition. Keloid formation tends to aggregate within families and is consistent with an autosomal dominant mode of inheritance.[13]

Both tumors and keloids are characterized by excessive cell proliferation and may result from abnormal expression of oncogenes. For example, previous studies have shown that tumor-associated genes regulate keloid development through two primary mechanisms. The first involves mutations in tumor suppressor genes—such as *TP53*, *FAS*, *CDKN1B*, *RB1*, and *CDKN2A*—which result in the loss of control over fibroblast proliferation. The second involves the activation of oncogenes—such as *MYC*, *FOS*, *BCL2*, and *TNC*—that enhance fibroblast proliferation and confer resistance to apoptosis. Together, these dysregulated pathways promote fibroblast overproliferation and reduced apoptosis, ultimately contributing to keloid formation.[14-18]

Most existing studies have focused on the formation and invasive growth of KFs from the perspective of the fibroblasts themselves,[19] while relatively few have investigated the role of microenvironmental changes in keloid keratinized epithelial cells and their relationship with fibroblasts. The following key points summarize the current understanding:

(i) Epithelial–mesenchymal interactions may contribute to the altered local secretion of growth factors
(i) Proliferative pathways commonly observed in embryonic cells may be reactivated in scar tissue cells
(ii) Hypoxia within scar tissue may induce the overexpression of angiogenic and endothelial growth factors, thereby delaying wound maturation and enhancing collagen secretion by fibroblasts.[20]

In summary, the pathological mechanisms driving the persistent, tumor-like growth of keloids that infiltrate surrounding tissue remain unclear. This study aims to construct a hub gene signature associated with scar formation by integrating single-cell analysis and machine learning approaches. The objective is to facilitate the development of novel targeted therapies for keloid. The research workflow of this study is illustrated in Figure 1.

2. Materials and methods

2.1. Data download

Transcriptome data for keloid and normal skin tissue were retrieved from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/) using the keyword "keloid." Three datasets were selected for analysis: One

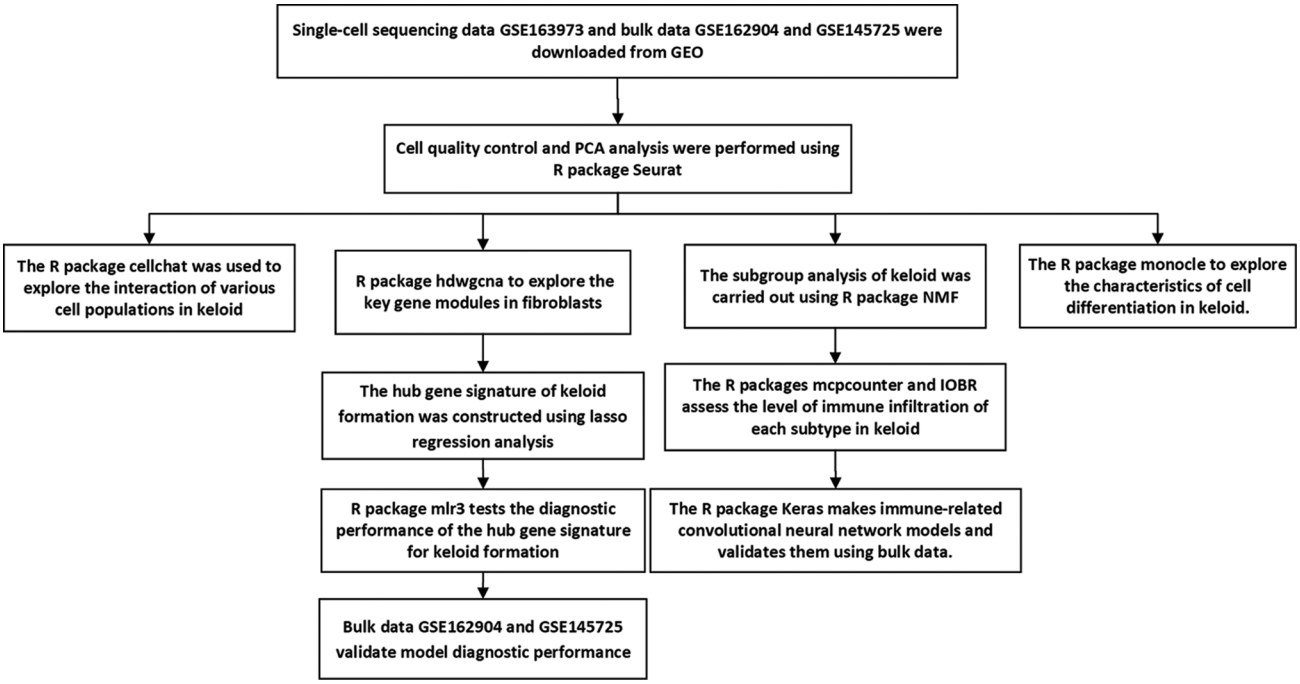

Figure 1. Schematic overview of the systematic workflow used in this study
Abbreviations: GEO: Gene Expression Omnibus; IOBR: Immuno-Oncology Biological Research; NMF: Nonnegative matrix factorization; PCA: Principal component analysis.

single-cell transcriptome dataset and two bulk transcriptome datasets. The single-cell dataset GSE163973 includes six samples—three keloid tissues and three normal skin tissues. The bulk transcriptome datasets include GSE162904, comprising six samples (three keloid tissues and three normal skin tissues), and GSE145725, which consists of 19 samples (nine keloid tissues and 10 normal skin tissues).

2.2. Single-cell data processing

In this study, analyses were conducted using R software (R Core Team, New Zealand). The R package Seurat[21] was used to construct single-cell transcriptome expression matrices and perform cellular quality control. Genes expressed in fewer than three cells were excluded, along with low-quality cells defined as those with fewer than 200 or more than 5,000 detected genes, a mitochondrial gene proportion exceeding 10%, or a hemoglobin gene proportion exceeding 5%. After filtering, a total of 32,542 high-quality cells were retained for subsequent analysis.

Subsequently, principal component analysis (PCA) was performed, and batch effects were corrected using the R package Harmony. Dimensionality reduction and clustering were performed using Uniform Manifold Approximation and Projection (UMAP) (Figure S1). Cell type annotation was performed with the R package SingleR, based on the "HumanPrimaryCellAtlasData" reference dataset.[22] Representative marker genes from this reference include *CD34*, *CD133*, and *ALDH1A1* for tissue stem cells (hematopoietic and epithelial stem cell markers), and *COL1A1*, *DCN*, and *ACTA2* (α-smooth muscle actin) for fibroblasts. A total of eight distinct cell subtypes were identified.

2.3. Weighted gene co-expression network analysis in high-dimensional data (hdWGCNA)

The R package hdWGCNA[23] was used to perform weighted gene co-expression network analysis (WGCNA) in high-dimensional transcriptomic data, such as single-cell RNA-sequencing or spatial transcriptomics. WGCNA is a systems biology method used to analyze gene expression patterns across multiple samples by clustering genes with similar expression profiles into modules and correlating these modules with specific traits or phenotypes. The hdWGCNA package was employed to identify key subpopulations in fibroblasts from the single-cell transcriptome dataset GSE163973 for subsequent analysis.

2.4. Construction of key gene signatures for keloid formation

The R package hdWGCNA was employed to select gene modules from key fibroblast subpopulations (clusters 1 and 5), focusing on the "brown," "cyan," and "yellow" modules. A total of 75 hub genes were identified and constructed as key gene signatures for keloid formation. Subsequently, the common genes among the 75 hub genes

were examined in the GSE162904 and GSE145725 datasets, resulting in the identification of 62 common genes. Lasso regression analysis was then applied to screen these hub genes, and eight hub genes were identified: *GUK1*, *BGN*, *MGST3*, *ADM*, *RAPGEF2*, *VPS28*, *NUMB*, and *SSFA2*. These hub genes were then used to construct the hub gene signatures for keloid formation.

2.5. Exploring cell interactions in keloids

Keloid cell subtype interactions were analyzed using the R package CellChat.[24] The functions in the CellChat—such as "createCellChat" for object construction, "identifyOverExpressedGenes" for identifying overexpressed ligand-receptor genes, and "identifyOverExpressedInteractions" for discovering these pairs in cellular subpopulations—were used. Expression values of these pairs were mapped to construct protein interaction networks using the "projectData" function, followed by the calculation of the overall interaction probability using "computeCommunProb" function. Additionally, pathways probabilities for each ligand–receptor pair were computed using "computeCommunProbPathway" function, and the results were visualized with "netVisual_bubble."

2.6. Exploring cell differentiation trajectories in keloid

Pseudotime trajectory analysis was conducted using Monocle software (Department of Genome Sciences, University of Washington, United States of America).[25] The "FindMarkers" function was utilized to construct differential gene sets between groups based on the criteria of |log2FC| >0.25 and min.pct = 0.25 for trajectory formation. The "newCellDataSet" function was employed to generate a CellDataSet object, excluding low-quality cells. Dimensionality reduction was performed using the "DDRTree" algorithm via the "reduceDimension" function. Cells were ordered with the "orderCells" function, and the trajectory was visualized using "plot_cell_trajectory" function. Gene expression changes across pseudotime were displayed using "plot_genes_in_pseudotime" function, illustrating expression patterns of copy number variation-altered genes. The starting point of the differentiation trajectory was determined based on biological relevance.

2.7. Nonnegative matrix factorization (NMF)

The principle of NMF[26] is that for any given nonnegative matrix V, it is possible to find nonnegative matrices W and H such that $V = W \times H$, thereby decomposing the original matrix into a product of two nonnegative matrices. In this study, R package NMF was used to perform cluster analysis on the bulk datasets GSE145725 and GSE162904, resulting in the construction of four subtypes. Differences in immune infiltration levels among these subtypes were further investigated.

2.8. Construction of an immune-related volume neural network model

A convolutional neural network (CNN) is a type of feed-forward neural network that incorporates convolutional operations within a deep structure and is one of the representative algorithms in deep learning**.** In this study, a gene-immunity CNN deep learning model was constructed using the R package Keras,[27] incorporating eight hub genes, 10 immune cells, and sample data. The diagnostic performance of the model was evaluated using GSE145725 as the training set and GSE162904 as the testing set.

2.9. Statistical analysis

The correlation between two continuous variables was assessed using Pearson's correlation analysis. The Kruskal-Wallis test was used to compare differences among the three groups. Receiver operating characteristic (ROC) curves for dichotomous variables were generated using the pROC package. All data preprocessing, statistical analyses, and visualizations were conducted in R version 4.1 software. All tests were two-sided, with $p<0.05$ considered statistically significant.

3. Results

3.1. Analysis of cell subtype differences between keloid and normal skin tissue

The R package Seurat was employed to perform quality control on the single-cell transcriptome dataset GSE163973, followed by data normalization for subsequent analysis. Dimensionality reduction and clustering using UMAP yielded 18 cell clusters (Figure 2A). Cell type annotation using the R package SingleR identified eight distinct cell subtypes (Figure 2B).

The results demonstrate that tissue stem cells and fibroblasts constitute the most abundant cell types in normal skin and keloid tissues, respectively. Comparative analysis reveals that fibroblasts are more prevalent in keloid tissue than in normal skin, while tissue stem cells are less abundant in keloid (Figure 2C-E). Previous studies have reported that keloids are driven by abnormal fibroblast proliferation and excessive collagen deposition.

Therefore, key gene modules from fibroblast subpopulations were selected for further analysis. Batch effects were removed using PCA, and UMAP was applied for dimensionality reduction and clustering, resulting in nine fibroblast subpopulations (Figure 2F and G). Comparative analysis of cell counts between keloid and normal skin tissues identifies clusters 1 and 5 as showing the most significant

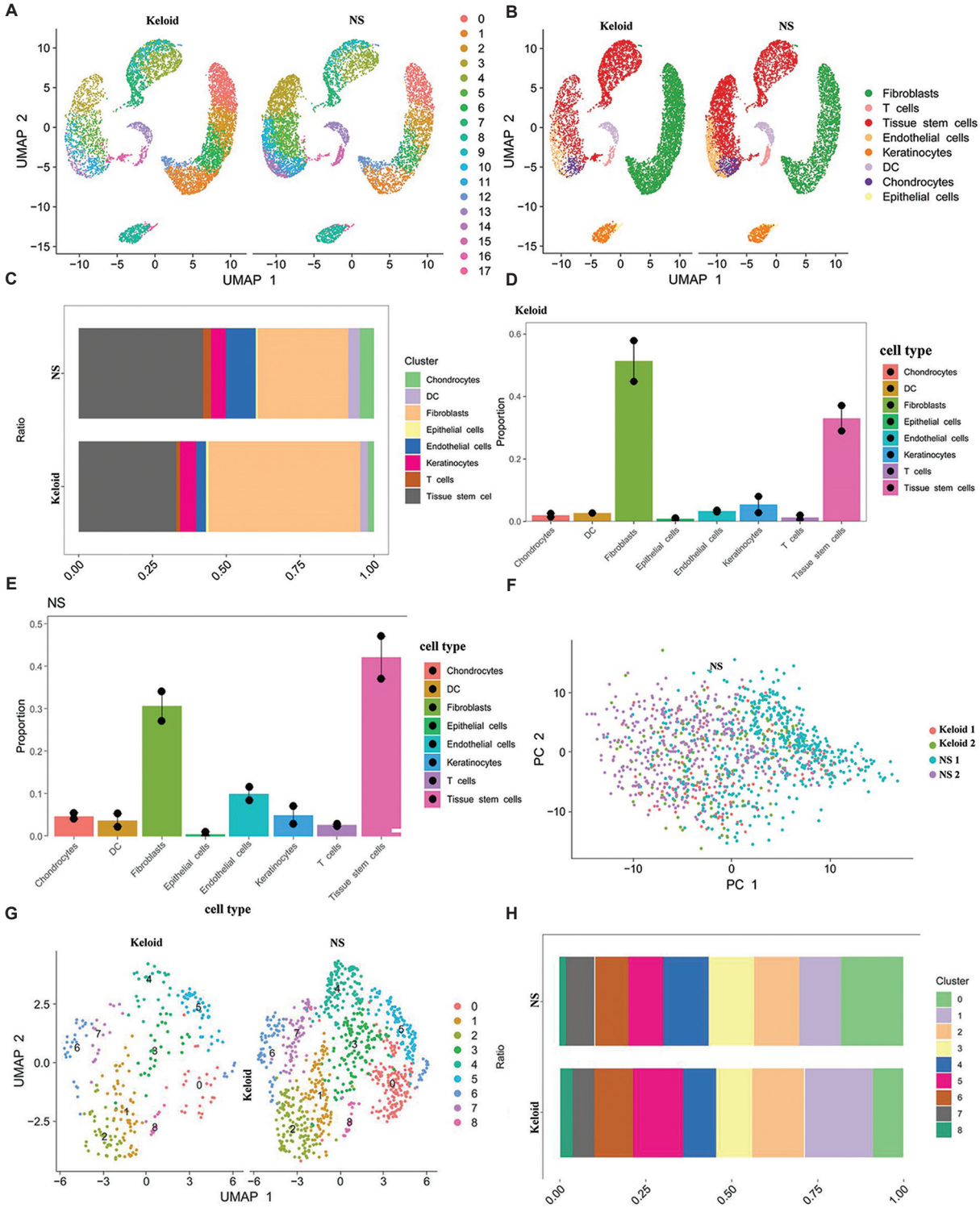

Figure 2. Identification of key cell subpopulations involved in keloid formation. (A) Scatter plot of dimensionality reduction and clustering in the GSE163973 dataset; (B) cell type annotation of the GSE163973 dataset; (C) proportion of each cell type in keloid and normal skin tissue; (D) proportion of each cell type within the keloid sample; (E) proportion of each cell type within the normal skin tissue; (F) principal component analysis scatter plot of fibroblasts in the GSE163973 dataset; (G) scatter plot of fibroblast clustering after dimensionality reduction; and (H) proportion of each fibroblast subpopulation in the GSE163973 dataset.

Abbreviations: DC: Dendritic cells; NS: Normal skin; UMAP: Uniform manifold approximation and projection.

differences, suggesting they may play critical roles in keloid formation. The expression matrices of clusters 1 and 5 were extracted for subsequent analysis (Figure 2H).

3.2. Analysis of cell-cell interactions and differentiation trajectories in keloid

The R package CellChat was employed to investigate interactions among the eight cell types identified in keloid tissue. The analysis reveals that endothelial cells exhibit the highest number and strength of interactions with other cells (Figure 3A and B). Given the critical role of fibroblasts in keloid formation,[8,9] their interactions with resident cells—such as endothelial cells and keratinocytes—were further examined. The results show that endothelial cells exhibit the strongest interactions with KFs (Figure 3C-J).

After exploring the ligand-receptor signaling involved in fibroblast interactions in keloid and other tissues, the midkine-low-density lipoprotein receptor-related protein 1 (MDK-LRP1) pair is identified as a key signaling pathway. The MDK-LRP1 ligand-receptor pair plays an extremely critical role in these interactions and shows significant involvement in mediating communication between keloid cells (Figure 3K).

The R package Monocle was employed to explore the characteristics of cell differentiation trajectories in keloids. The analysis identifies four key branches and nine differentiation states. Based on the characteristics of the simulated temporal scatter plot, most keloid cells are distributed at the beginning and end states (Figure 3L and M). Further analysis reveals that subpopulations 4 and 5 in keloid are highly active in differentiation and may represent key subpopulations in keloid formation (Figure 3N). Therefore, highly variable genes in keloid were extracted and examined for expression changes. The results, illustrated in a heatmap, demonstrate significant alterations in the expression of highly variable genes in clusters 1 and 5, which may represent key gene clusters involved in keloid formation (Figure 3O).

3.3. Identification of key gene modules in keloid

As keloid formation may result from abnormal fibroblast proliferation and excessive collagen deposition,[8] the R package hdWGCNA was employed to explore key gene modules in fibroblasts. The fibroblast subtype expression matrix was transposed to identify co-expressed gene modules using a soft threshold of $\beta = 7$. The analysis identifies 15 key gene modules (Figure 4A and B).

Hub genes for each module were determined by calculating the correlation between genes and their corresponding modules, with a threshold of $|kME| \geq 0.8$ (Figure 4C). The correlation between modules was also assessed, and the results demonstrate that the "green" module is significantly associated with the "black" module, while the "brown" module is significantly associated with the "red" module (Figure 4D).

Previous studies reported that fibroblast subpopulations 1 and 5 are key subpopulations for keloid formation. Therefore, this study further explored the enrichment of these key subpopulations in each keloid gene module. The findings identify remarkable enrichment of key subgroups in the "green," "dark blue," "purple," and "black" modules. Further analysis reveals that the gene expression level of the "brown" module is significantly higher in cluster 1 compared to other clusters. In cluster 5, the gene expression levels of the "cyan" and "yellow" modules are higher than those of other modules (Figure 4E and F).

These findings suggest that the "cyan," "yellow," and "brown" modules are key modules for keloid formation. Therefore, 75 hub genes from these modules were extracted as the key gene signatures for keloid formation (Table S1). Gene enrichment analysis of the 75 hub genes was then performed using the online Metascape database to investigate their biological functions. The results indicate close associations with osteoarthropathy, anemia, and keloid, as well as involvement in ribosomal and cytoplasmic biological processes (Figure 4G and H).

3.4. Analysis of hub gene signatures for keloid formation using machine learning

In the bulk datasets GSE162904 and GSE145725, the accuracy of key gene signatures for keloid formation was evaluated using multiple machine learning algorithms. PCA was performed on the GSE162904 and GSE145725 datasets to remove inter-sample differences (Figure S2) and normalize the data. Common genes among the 75 hub genes in both datasets were identified, resulting in a total of 62 common genes. Subsequently, Lasso regression analysis was applied to screen for hub genes. The analysis identifies eight hub genes: *GUK1*, *BGN*, *MGST3*, *ADM*, *RAPGEF2*, *VPS28*, *NUMB*, and *SSFA2*. These hub genes were then constructed as the hub gene signatures for keloid formation (Table S2) (Figure 5A and B).

To evaluate the diagnostic efficacy of the hub gene signatures for keloid formation using machine learning, the R package mlr3 was employed. The GSE145725 dataset was used as the training group, while GSE162904 was used as the testing group. In the training group, seven machine learning algorithms were applied, including "kknn," "lda,"

"log_rog," "navie_bayes," "ranger," "rpart," and "svm." The results indicate that the hub gene signature for keloid formation demonstrates excellent diagnostic efficacy across multiple algorithms, with a strong ability to distinguish between keloid and normal skin tissue, particularly in the "kknn" algorithm (Figure 5C and D).

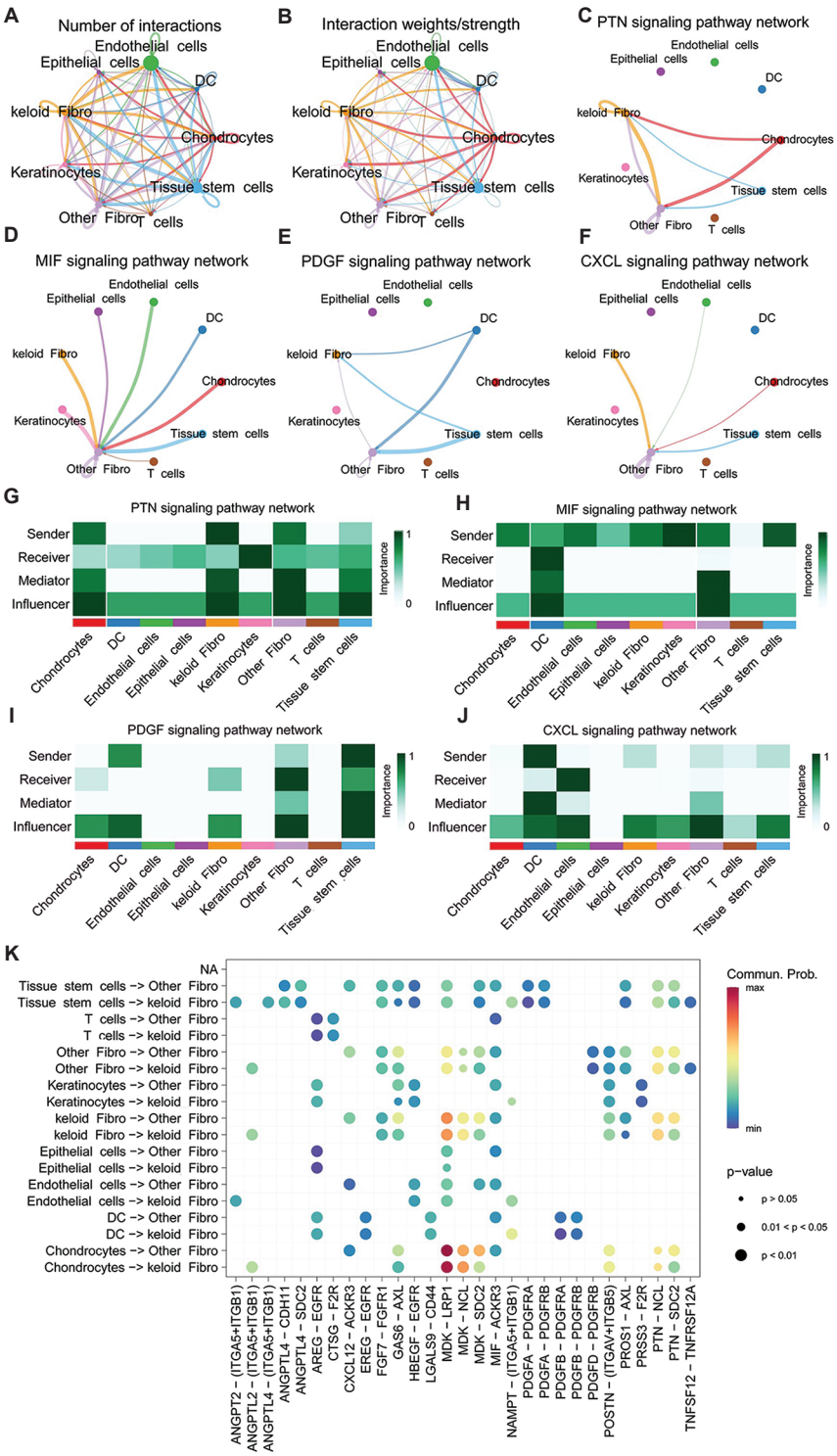

Figure 3. *(Continued)*

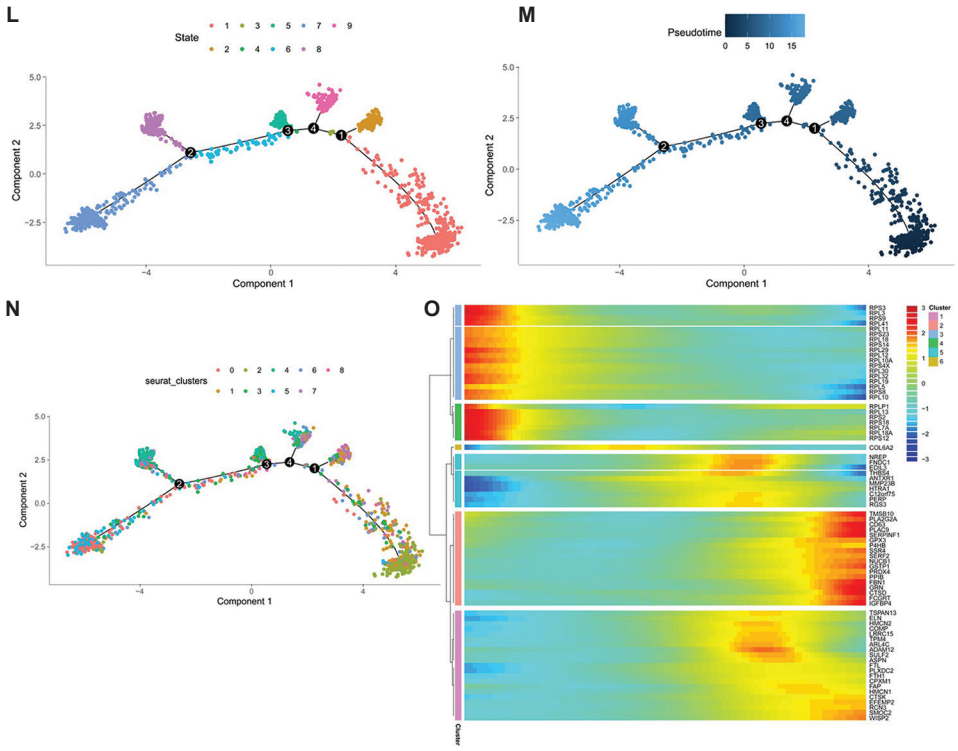

Figure 3. Analysis of cell–cell interactions and developmental trajectories in keloid. (A) Number of interactions among keloid cell types; (B) interaction strength between keloid subpopulations; (C-F) key signaling pathways involved in fibroblast interactions with other cell types in keloid tissue; (G-J) heatmaps depicting intracellular interaction in different signaling pathways; (K) bubble plot showing ligand–receptor interactions between individual keloid cell types; (L) scatter plot illustrating the differentiation status of individual keloid cells; (M) pseudotime trajectory analysis of keloid cells; (N) scatter plot of differentiation status across keloid cell subpopulations; and (O) heatmap of the changes in the expression of highly variable genes during keloid cell differentiation.
Abbreviations: CXCL: CXC Chemokine ligand; EGFR: Epidermal growth factor receptor; MIF: Macrophage migration inhibitory factor; PTN: Pleiotrophin.

Based on this, the "kknn" algorithm was further employed to plot the ROC curve and calculate diagnostic efficacy. The area under the curve in the training group is 0.885, indicating excellent performance. Similarly, in the testing group GSE162904, the hub gene signature shows strong diagnostic ability with an area under the curve of 0.889 (Figure 5E and F).

To determine whether the hub gene signature is also differentially expressed in the bulk datasets, its expression levels in GSE162904 and GSE145725 were analyzed. The results show that the signature is differentially expressed between keloid and normal skin tissue, with higher expression observed in keloids (p<0.05) (Figure 5G).

The lack of effective drugs for treating keloid may be due to its heterogeneity and the presence of different subtypes. In this study, subtype analysis of keloid was performed using the R package NMF. The analysis identifies four keloid subtypes, with significant heterogeneity observed among them (Figure 5H and I).

The R packages MCPcounter and Immuno-Oncology Biological Research were employed to assess immune infiltration levels for each subtype. The results demonstrate significant differences in immune infiltration levels among the subtypes, with the highest infiltration observed in subtype 1 and the lowest in subtype 4 (p<0.05) (Figure 5J). The heatmap of immune infiltration levels shows a notable difference between subtypes 2 and 4 (Figure 5K). Additionally, the correlation between hub genes involved in keloid formation and immune infiltration was analyzed. The findings show that these hub genes are significantly and positively correlated with fibroblast infiltration levels in keloids (Figure 5L).

A CNN model was developed using the R package Keras, with GSE145725 as the training set and GSE162904 as the testing set. A two-dimensional array of eight hub genes and 10 immune cell types was first constructed and combined with 19 samples from the GSE145725 dataset to generate a three-dimensional array for the gene-immunity CNN model. Convolution was carried out over 200 iterations. The results demonstrate that the model performs well

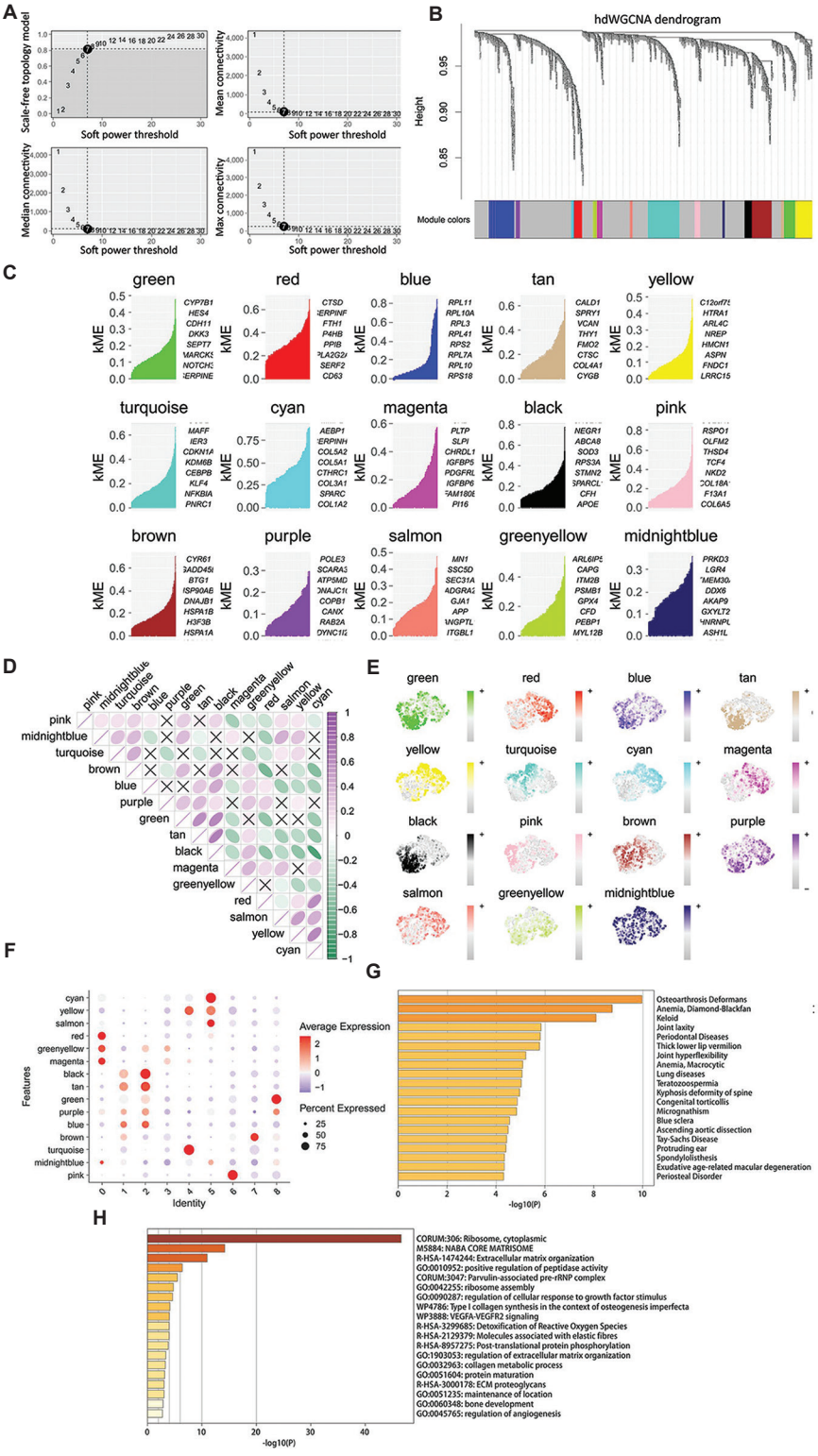

Figure 4. Construction of key gene signatures for keloid formation. (A) Determination of the soft threshold (β = 7); (B) hierarchical clustering dendrogram of fibroblast module signatures in keloid; (C) identification of hub genes across modules; (D) heatmap showing correlations between modules; (E) scatter plot of enrichment levels for key fibroblast subpopulations in keloid across modules; (F) bubble plot of enrichment levels for key keloid fibroblast subpopulations in each module; and (G and H) gene enrichment analysis of 75 hub genes.
Abbreviation: ECM: Extracellular matrix; VEGFA: ascular endothelial growth factor A; VEGFR2: Vascular endothelial growth factor receptor 2.

in the training set, with a bias value of 0.3 and a correct classification rate of 0.8 (Figure 5M). Similarly, the model shows strong performance in the testing set, achieving a sensitivity of 1.00 and a specificity of 0.889 at a threshold of 0.498 (Figure 5N).

4. Discussion

Keloid is a distinct type of scarring that primarily arises from abnormal proliferation of dermal fibroblasts following traumatic injury to the skin, accompanied by

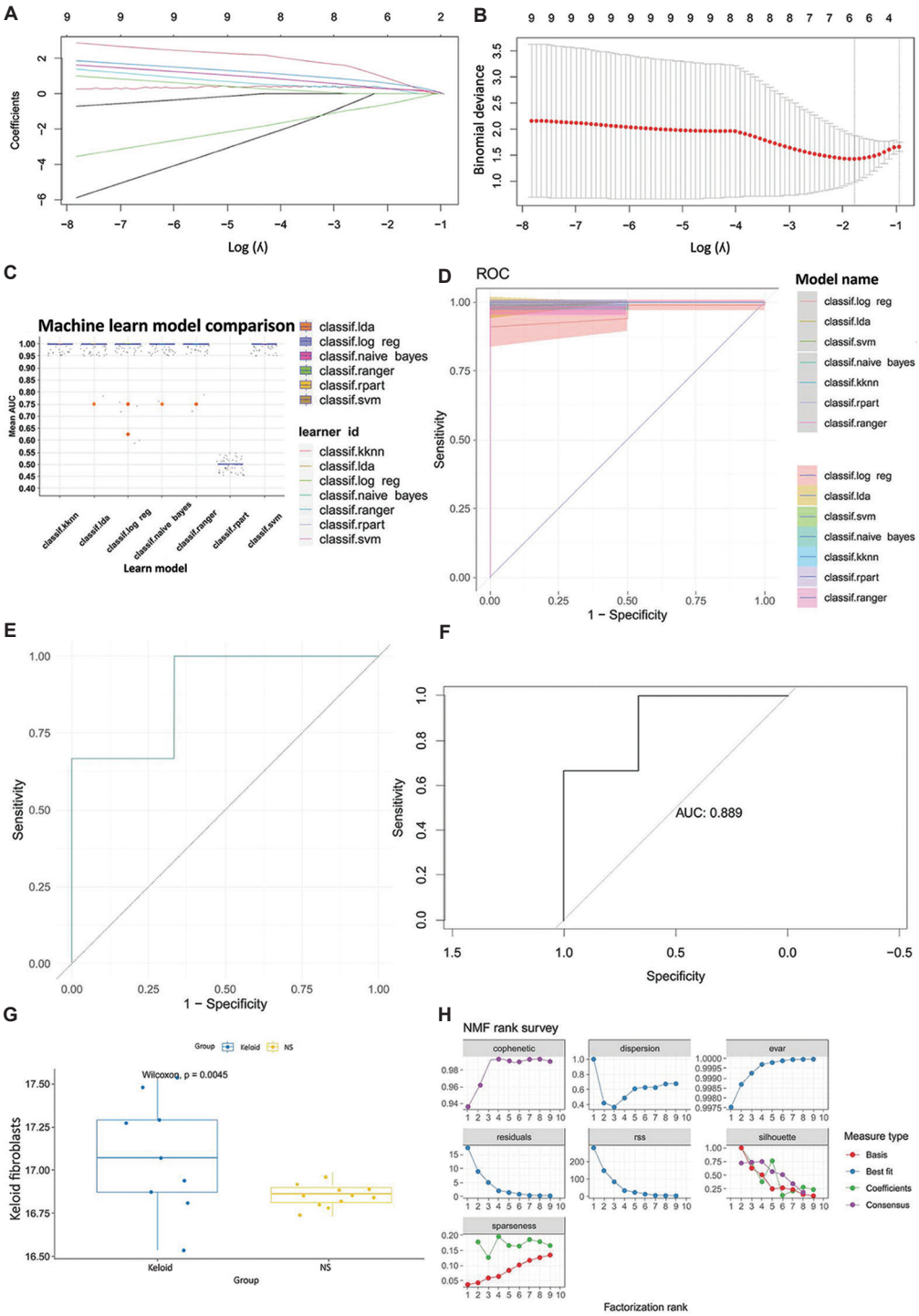

Figure 5. *(Continued)*

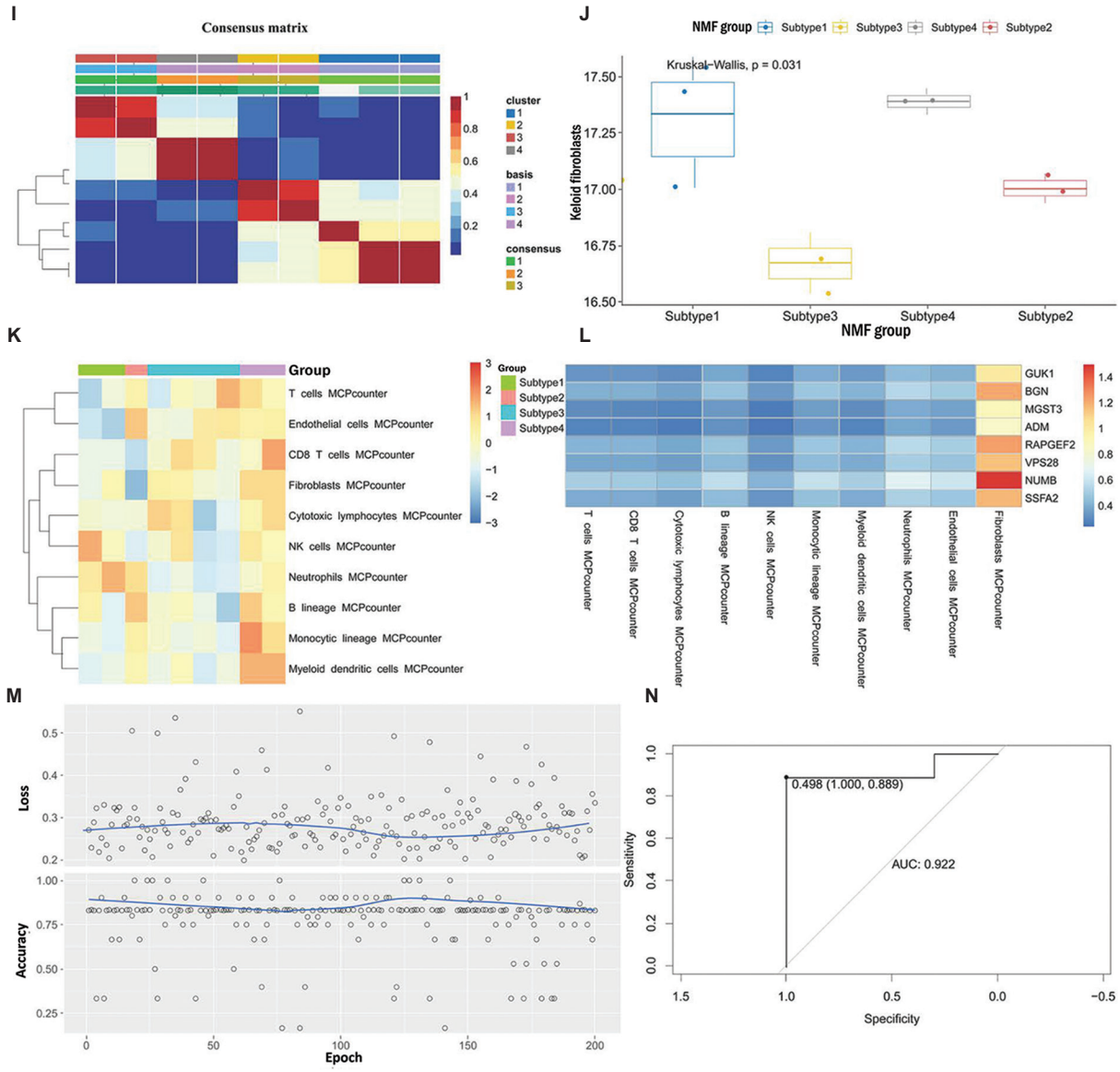

Figure 5. Diagnostic performance of hub gene signature for keloid formation using machine learning. (A and B) Lasso regression analysis of hub genes; (C) box plots of diagnostic performance across machine learning algorithms; (D) Receiver operating characteristic curves evaluating performance of machine learning models; (E) diagnostic performance of the hub gene signature in the GSE145725 training dataset; (F) diagnostic performance of the hub gene signature in the GSE162904 testing dataset; (G) expression levels of hub gene signatures for keloid formation in GSE162904 and GSE145725; (H) selection of optimal cluster number; (I) heatmap of non-negative matrix factorization clustering results; (J) box plots illustrating immune infiltration levels for each subtype; (K) heatmap showing immune infiltration levels for each subtype; (L) correlation heatmap between hub genes and immune cell infiltration; (M) performance of the gene-immune convolutional neural network deep learning model in the GSE145725 training dataset; (N) Receiver operating characteristic curve of the gene-immunity convolutional neural network deep learning model in the GSE162904 testing dataset.
Abbreviations: NMF: Non-negative matrix factorization; NS: Normal skin; ROC: Receiver operating characteristic.

excessive deposition of connective tissue and hyaline-like degeneration. It commonly develops in areas prone to injury, such as the anterior chest, earlobes, upper arms, and scapulae, with the anterior chest—particularly the sternum–being the most frequently affected site.[28] Keloid is often associated with pain and itching, and its disfiguring appearance can significantly affect an individual's quality of life, both physically and psychologically.

Keloid typically develops secondary to skin trauma caused by various factors, including burns, surgery, tattoos, ear piercing, or accidental injury. A key distinction between keloid and proliferative keloid lies in their

etiology and presentation. Keloid can form spontaneously in response to minimal or even no apparent trauma and clinically present as excessive growth that extends beyond the original wound boundaries, invading adjacent normal tissue in a tumor-like manner. Due to these characteristics, keloid is often referred to as benign skin tumor.[29] It may continue to grow for several years post-injury and does not exhibit contracture. In contrast, proliferative keloid is typically caused by deep dermal trauma, appears red, does not have a preferred anatomical site, and generally remains confined within the boundaries of the original wound. It often develops within a few months of injury and may be accompanied by keloid contracture.[13,30]

Although the pathogenesis of keloid remains unclear, commonly used treatment strategies—including hormone therapy, radiation therapy, cryotherapy, laser therapy, compression therapy, silicone membrane therapy, antitumor therapy, surgical resection, and immunological agents—can achieve some therapeutic benefit. However, the recurrence rate following treatment remains high, and overall outcomes are often unsatisfactory. Previous studies have suggested that keloid formation may be attributed to several key mechanisms:

(i) Gene dysregulation: The development of keloid is closely associated with reduced apoptosis and excessive fibroblast proliferation. The *FAS* gene, a member of the tumor necrosis factor receptor family, plays a crucial role in apoptosis regulation. The overexpression of nonfunctional Fas protein in keloid may contribute to insufficient fibroblast apoptosis. Chen *et al.*[31] reported that the low expression of oncogenes such as *CDKN1B* and *CDKN2A* is associated with keloid development, and that the anti-apoptotic gene *BCL2* may inhibit apoptosis in fibroblasts.[32]

(ii) Abnormal ECM remodeling: Keloid is characterized by an imbalance in ECM production and degradation, resulting in excessive ECM deposition. Overexpression of ECM proteins is a hallmark feature of keloid tissue. Collagen, the main component of the ECM, exhibits a disrupted ratio in keloid, with significantly increased type I collagen and reduced type III collagen. This altered composition is believed to contribute to pathological scar formation.[33]

(iii) Involvement of immune mechanisms: Keloid tissue shows marked infiltration of leukocytes, including dendritic cells, lymphocytes, plasma cells, and immunoglobulins. This indicates that an immune response is actively involved in keloid development and progression. Initially, keloid grows slowly and then gradually expands, a pattern that Placik *et al.*[34] suggest may be linked to local immune response (Figure 5K and 5L).

However, Niessen *et al.*[35] proposed that Langerhans cells and interleukin-4 play a more prominent role in keloid formation than mast cells, T lymphocytes, and macrophages. The findings of this study are consistent with several aspects of this hypothesis. Single-cell transcriptomic analysis identifies keratinocytes and chondrocytes—both of which contribute to collagen production—as key cell types involved in keloid formation. The number of keratinocytes and chondrocytes is higher in keloid tissue than in normal skin tissue. This suggests that ECM deposition in the keloid is associated with an abnormal increase in keratin-producing cells and chondrocytes, leading to excessive collagen accumulation.

Additionally, the results show immune cell infiltration in keloid tissue—such as T cells—though the difference is not statistically significant compared to normal skin tissue. Interestingly, the number of tissue stem cells in keloids is significantly lower than in normal skin. Fibroblasts act as the primary effector cells in keloid fibrosis, while tissue stem cells support repair and fibroblast replenishment. Keloid tissue demonstrates an expansion of fibroblasts alongside a depletion of stem cells, suggesting a possible reciprocal regulation between the two. These findings align with those of Han,[36] who reported that adipose-derived mesenchymal stem cells inhibit fibrosis in KFs, potentially through downregulation of the Notch1/Jagged-1 and TGF-β1/SMAD3 signaling pathways. Current research on the role of tissue stem cells in keloid formation remains limited, highlighting a promising direction for future investigation.

This study also reveals key features of keloid formation that may offer insights into its underlying pathological mechanisms (Figures 3K and J). Analysis of cell–cell interactions indicates that endothelial cells are highly active in intercellular signaling, exhibiting the highest number and intensity of interactions. Notably, the MDK-LRP1 ligand-receptor pair plays a critical role across all cellular interactions in keloid, particularly in signaling between keratinocytes and fibroblasts—cells whose abnormal proliferation underlies keloid pathology. Within the tumor-like stromal microenvironment of keloids, endothelial cells and keratinocytes may secrete growth factors such as platelet-derived growth factor and TGF-β, promoting fibroblast activation and ECM deposition. Future studies may employ conditioned media experiments to further clarify these paracrine signaling pathways.

Given the poor efficacy and high recurrence rate of current keloid treatments, this study also explored the possibility of disease heterogeneity. Subtype analysis identifies four distinct heterogeneous subtypes of keloid. Further analysis reveals significant variation in immune infiltration levels among these subtypes, which may

underlie differential responses to immunotherapy and potential resistance mechanisms (Figure 5H-J). In addition, a hub gene signature for keloid formation is established, offering a potential molecular target for therapeutic development aimed at halting the transition from normal wound healing to pathological keloid formation.

However, several limitations remain in this study. The sample sizes from the single-cell transcriptome dataset GSE163973 and the bulk transcriptome datasets GSE162904 and GSE145725 are relatively small, and some findings require validation with larger cohorts. In the single-cell transcriptome analysis of GSE163973, the number of tissue stem cells in keloid is significantly lower than in normal skin tissue; however, further classification of these stem cells is not possible due to their cellular complexity. As tissue stem cells encompass multiple subtypes, the specific cell population involved in keloid formation remains unidentified. In addition, several findings in this study require experimental validation. Collaborative efforts with research teams at Sichuan University are currently underway to further confirm these findings. It is also important to note that the inferred cell–cell interactions based on ligand-receptor analysis rely on transcriptomic data and prior knowledge. These predictions may not fully reflect actual biological interactions, as they are subject to modulation by post-transcriptional and posttranslation mechanisms.

Despite these limitations, this study establishes a hub gene signature for keloid formation, consisting of eight hub genes identified from single-cell transcriptomic data. This signature demonstrates strong diagnostic performance across multiple machine learning algorithms and may serve as a promising therapeutic target for future intervention strategies in keloid treatment.

5. Conclusion

This study identifies eight hub genes signature for keloid diagnosis, unveils fibroblast subtypes and immune infiltration heterogeneity, offering novel therapeutic targets and precision medicine strategies.

Acknowledgments

We thank the authors of GSE163973, GSE162904, and GSE145725 for their contribution and the Metascape database for providing the platform for data analysis.

Funding

This study was funded by the Medical Youth Innovation Research Project Plan of Sichuan Province of the Sichuan Medical Association (S20091).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Yong Zou
Formal analysis: Bing Liu, Jun Yin
Funding acquisition: Zhengqiang Wan
Investigation: Yuzhen Ma
Methodology: Cheng Chen, Jun Yin
Writing–original draft: Xiangbing Zheng, Zhengqiang Wan
Writing–review & editing: Yong Zou, Xiangbing Zheng

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Second People's Hospital of Yibin (Approval no. 2024-007-01). The data used in this study were obtained from the public Gene Expression Omnibus database; therefore, patients were not required to sign an informed consent form for this study.

Consent for publication

Not applicable.

Availability of data

The authors confirm that the data supporting the findings of this study are available in the article and its supplementary file.

References

1. Jing P. *The Role of Splicing Factor SRSF2 in Keloid Fibroblasts*. [Master's Thesis]; 2019. p. 79.
2. Asilian A, Darougheh A, Shariati F. New combination of triamcinolone, 5-fluorouracil, and pulsed-dye laser for treatment of keloid and hypertrophic scars. *Dermatol Surg*. 2006;32:907-915.

   doi: 10.1111/j.1524-4725.2006.32195.x
3. Shih B, Bayat A. Genetics of keloid scarring. *Arch Dermatol Res*. 2010;302:319-339.

   doi: 10.1007/s00403-009-1014-y
4. Supp DM, Hahn JM, Glaser K, McFarland KL, Boyce ST. Deep and superficial keloid fibroblasts contribute differentially to tissue phenotype in a novel *in vivo* model of keloid scar. *Plast Reconstr Surg*. 2012;129:1259-1271.

   doi: 10.1097/PRS.0b013e31824ecaa9
5. Lim IJ, Phan TT, Bay BH, *et al*. Fibroblasts cocultured with keloid keratinocytes: Normal fibroblasts secrete collagen in a keloidlike manner. *Am J Physiol Cell Physiol*. 2002;283:C212-C222.

   doi: 10.1152/ajpcell.00555.2001

6. Iqbal SA, Sidgwick GP, Bayat A. Identification of fibrocytes from mesenchymal stem cells in keloid tissue: A potential source of abnormal fibroblasts in keloid scarring. *Arch Dermatol Res*. 2012;304:665-671.

doi: 10.1007/s00403-012-1225-5

7. Ueda K, Furuya E, Yasuda Y, Oba S, Tajima S. Keloids have continuous high metabolic activity. *Plast Reconstr Surg*. 1999;104:694-698.

doi: 10.1097/00006534-199909030-00012

8. Van De Water L, Varney S, Tomasek JJ. Mechanoregulation of the myofibroblast in wound contraction, scarring, and fibrosis: Opportunities for new therapeutic intervention. *Adv Wound Care (New Rochelle)*. 2013;2:122-141.

doi: 10.1089/wound.2012.0393

9. Tholpady SS, DeGeorge BR Jr., Campbell CA. The effect of local rho-kinase inhibition on murine wound healing. *Ann Plast Surg*. 2014;72:S213-S219.

doi: 10.1097/SAP.0000000000000150

10. Kao HK, Chen B, Murphy GF, Li Q, Orgill DP, Guo L. Peripheral blood fibrocytes: Enhancement of wound healing by cell proliferation, re-epithelialization, contraction, and angiogenesis. *Ann Surg*. 2011;254:1066-1074.

doi: 10.1097/SLA.0b013e3182251559

11. Wolfram D, Tzankov A, Pulzl P, Piza-Katzer H. Hypertrophic scars and keloids--a review of their pathophysiology, risk factors, and therapeutic management. *Dermatol Surg*. 2009;35:171-181.

doi: 10.1111/j.1524-4725.2008.34406.x

12. Van Leeuwen MC, Stokmans SC, Bulstra AJ, Meijer OW, Van Leeuwen PA, Niessen FB. High-dose-rate brachytherapy for the treatment of recalcitrant keloids: A unique, effective treatment protocol. *Plast Reconstr Surg*. 2014;134:527-534.

doi: 10.1097/PRS.0000000000000415

13. Halim AS, Emami A, Salahshourifar I, Kannan TP. Keloid scarring: Understanding the genetic basis, advances, and prospects. *Arch Plast Surg*. 2012;39:184-189.

doi: 10.5999/aps.2012.39.3.184

14. Wu Y, Wang B, Li YH, *et al*. Meta-analysis demonstrates association between arg72pro polymorphism in the p53 gene and susceptibility to keloids in the Chinese population. *Genet Mol Res*. 2012;11:1701-1711.

doi: 10.4238/2012.June.29.2

15. Messadi DV, Doung HS, Zhang Q, *et al*. Activation of nfkappab signal pathways in keloid fibroblasts. *Arch Dermatol Res*. 2004;296:125-133.

doi: 10.1007/s00403-004-0487-y

16. Lu F, Gao J, Ogawa R, Hyakusoku H, Ou C. Fas-mediated apoptotic signal transduction in keloid and hypertrophic scar. *Plast Reconstr Surg*. 2007;119:1714-1721.

doi: 10.1097/01.prs.0000258851.47193.06

17. Hu Z, Lou L, Luo S. [Experimental study of the expression of c-myc, c-fos and proto-oncogenes on hypertrophic and scars]. *Zhonghua Zheng Xing Wai Ke Za Zhi*. 2002;18:165-167.

18. Teofoli P, Barduagni S, Ribuffo M, Campanella A, De Pita'O, Puddu P. Expression of bcl-2, p53, c-jun and c-fos protooncogenes in keloids and hypertrophic scars. *J Dermatol Sci*. 1999;22:31-37.

doi: 10.1016/s0923-1811(99)00040-7

19. Jin Z. Increased c-met phosphorylation is related to keloid pathogenesis: Implications for the biological behaviour of keloid fibroblasts. *Pathology*. 2014;46:25-31.

doi: 10.1097/PAT.0000000000000028

20. Xiaoyang M. Study on hypoxia/HIF-1α-induced EMT in keloid keratinocytes and its effect on invasiveness [Doctoral dissertation]. Peking Union Medical College; 2015.

21. Aran D, Looney AP, Liu L, *et al*. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. 2019;20:163-172.

doi: 10.1038/s41590-018-0276-y

22. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. Hdwgcna identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods*. 2023;3:100498.

doi: 10.1016/j.crmeth.2023.100498

23. Jin S, Guerrero-Juarez CF, Zhang L, *et al*. Inference and analysis of cell-cell communication using cellchat. *Nat Commun*. 2021;12:1088.

doi: 10.1038/s41467-021-21246-9

24. Trapnell C, Cacchiarelli D, Grimsby J, *et al*. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32:381-386.

doi: 10.1038/nbt.2859

25. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999;401:788-791.

doi: 10.1038/44565

26. van Merriënboer B, Bahdanau D, Dumoulin V, et al. Blocks and Fuel: Frameworks for Deep Learning. *arXiv preprint* arXiv:1506.00619; 2015.

doi: 10.48550/arXiv.1506.00619

27. Heping H. Study on the Expression of Hedgehog Signaling Pathway and EMT-related lncRNAs in Keloids [Doctoral Dissertation]. Nanchang University; 2018.

28. Vincent AS, Phan TT, Mukhopadhyay A, Lim HY, Halliwell B, Wong KP. Human skin keloid fibroblasts display bioenergetics of cancer cells. *J Invest Dermatol*.

2008;128:702-709.

doi: 10.1038/sj.jid.5701107

29. Murray JC. Keloids and hypertrophic scars. *Clin Dermatol*. 1994;12:27-37.

doi: 10.1016/0738-081x(94)90254-2

30. Slemp AE, Kirschner RE. Keloids and scars: A review of keloids and scars, their pathogenesis, risk factors, and management. *Curr Opin Pediatr*. 2006;18:396-402.

doi: 10.1097/01.mop.0000236389.41462.ef

31. Yahong C, Xiaoli W. Research progress in the mechanism of invasive growth of keloid. *J Tissue Eng Reconstr Surg*. 2015;11:335-338.

32. Gang Z, Shaojun L, Shaoming T, Jie L, Qiang Z. Expression of BCL-2 in different parts of keloid and its significance. *Guangdong Med*. 2006:1811-1812.

33. Sidgwick GP, Bayat A. Extracellular matrix molecules implicated in hypertrophic and keloid scarring. *J Eur Acad Dermatol Venereol*. 2012;26:141-152.

doi: 10.1111/j.1468-3083.2011.04200.x

34. Placik OJ, Lewis VJ Jr. Immunologic associations of keloids. *Surg Gynecol Obstet*. 1992;175:185-193.

35. Niessen FB, Schalkwijk J, Vos H, Timens W. Hypertrophic scar formation is associated with an increased number of epidermal langerhans cells. *J Pathol*. 2004;202:121-129.

doi: 10.1002/path.1502

36. Han B. Study on the Effect and Molecular Mechanism of Paracrine Factors from Adipose-Derived Mesenchymal Stem Cells against Fibrosis of Pathological Scar Fibroblasts [Doctoral Dissertation]. Peking Union Medical College; 2018.