


# Identification of potential therapeutic target SPP1 and related RNA regulatory pathway in keloid based on bioinformatics analysis

Ruxin Xie\* , Jiao Yun\*, Chenyu Li\*, Shiwei Zhang\*, Ai Zhong, Junliang Wu, Ying Cen, Zhengyong Li and Junjie Chen

Department of Burn and Plastic Surgery, West China Hospital of Sichuan University, Chengdu, Sichuan, China

## ABSTRACT

**Objective:** To explore the complex mechanisms of keloid, new approaches have been developed by different strategies. However, conventional treatment did not significantly reduce the recurrence rate. This study aimed to identify new biomarkers and mechanisms for keloid progression through bioinformatics analyses.

**Methods:** In our study, microarray datasets for keloid were downloaded from the GEO database. Differentially expressed genes (DEGs) were identified by R software. Multiple bioinformatics tools were used to identify hub genes, and reverse predict upstream miRNAs and lncRNA molecules of target hub genes. Finally, the total RNA-sequencing technique and miRNA microarray were combined to validate the identified genes.

**Results:** Thirty-one DEGs were screened out and the upregulated hub gene SPP1 was finally identified, which was consistent with our RNA-sequencing analysis results and validation dataset. In addition, a ceRNA network of mRNA (SPP1)-miRNA (miR-181a-5p)-lncRNA (NEAT1, MALAT1, LINC00667, NORAD, XIST and MIR4458HG) was identified by the bioinformatics databases. The results of our miRNA microarray showed that miR-181a-5p was upregulated in keloid, also we found that the lncRNA NEAT1 could affect keloid progression by retrieving the relevant literature.

**Conclusions:** We speculate that SPP1 is a potential candidate biomarker and therapeutic target for patients with keloid, and NEAT1/miR-181a-5p/SPP1 might be the RNA regulatory pathway that regulates keloid formation.

## KEY MESSAGES

- Identify new biomarkers in keloid, potentially improve disease diagnosis and treatment.
- Through a variety of bioinformatics analysis tools, we found that the miRNA pathway NEAT1/miR-181a-5p/SPP1 may participate in controlling disease progression in the keloid.
- Providing insight into the mechanisms of disease development in the keloid at the transcriptome level.

## ARTICLE HISTORY

Received 27 January 2024

Revised 29 March 2024

Accepted 5 June 2024

## KEYWORDS



Keloid; bioinformatics analysis; biomarker; therapeutic target; ceRNA

## 1. Introduction


Keloid is usually considered a kind of benign fibroproliferative skin tumor that presents with raised dermal lesions due to an impaired wound healing process and infiltration into adjacent normal tissues [1]. Deposition of extracellular matrix (ECM) and excessive proliferation of fibroblasts (FBs) eventually lead to the emergence of keloid progression [2]. Multiple methods, including surgical excision, silicone gel products, hydrogel scaffolds, laser treatment, local injections of insoluble steroids, cryosurgery treatment, and local

chemotherapeutic agents (e.g. bleomycin and 5-fluorouracil) [3–9], have been used in clinical therapies. However, considering its high rate of recurrence even after curative treatments, keloid is an intractable disease, which has a lack of effective therapeutic targets. Hence, there is an urgent need to explore new molecular mechanisms and therapeutic targets for keloid.

With the integration of computer science and biology, we can dig and obtain huge biological information data in the era of big data, which provides a new

**CONTACT** Junjie Chen  [cjemail@163.com](mailto:cjemail@163.com)  Department of Burn and Plastic Surgery, West China Hospital of Sichuan University, No. 37 Guoxue Alley, Wuhou District, Chengdu, Sichuan, 610041, China.

\*These authors contributed to the work equally and should be regarded as co-first authors to this work

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07853890.2024.2382949>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

research method in many fields such as genetics and molecular biology. Previous studies have found that there are a large number of differentially expressed genes (DEGs) between keloid and healthy skin, and they are concentrated in genes related to collagen production and angiogenesis, which is indicated by the excessive collagen deposition and vascular proliferation of keloid [10, 11]. FBs are the primary source of ECM and collagen proteins. FBs isolated from keloid tissues are more sensitive to exogenous TGF- $\beta$ 1, PDGF, and insulin-like growth factor I compared to FBs from healthy skin, resulting in the upregulation of collagen and ECM-related genes [12, 13]. Additionally, several studies have suggested that during keloid formation, the abnormal function of endothelial cells (ECs) and their interaction with immune cells and FBs may influence the deposition of ECM and the progression of inflammation [14]. Pathological changes such as vascular dysregulation, low microvessel density, microvascular occlusion, and flattened lumens could potentially lead to tissue hypoxia in keloid tissues, thereby resulting in higher expression of collagen and ECM-related genes in keloid FBs compared to healthy skin FBs [15]. Therefore, we aim to explore the key pathogenic genes influencing keloid through the core cell populations in these two types of keloid tissues.

Further understanding of the reported genes and molecular networks will enhance a comprehensive understanding of keloid pathology. High false positive rates were observed in studies using only one microarray platform and a small number of samples, which may lead to inconsistent results. Thus, it is necessary to identify new therapeutic targets and biomarkers to overcome the inconsistencies observed in previous studies [16, 17]. Herein, we compared the mRNA expression profiles of FBs and ECs in healthy and keloid skin from GEO database. The intersection DEGs in different datasets were used for subsequent analysis. Then, the STRING database was utilized to establish a protein-protein interaction (PPI) network for further analysis and identification of hub genes. It is hoped that the comprehensive investigation of hub genes will facilitate our understanding of keloid pathogenesis and progression. However, new therapeutic targets remain to be validated by further comprehensive

analysis. Moreover, our bioinformatics analysis only determines the gene-level correlation differences, and further experimental studies are needed to explore the proteomic differences.

## 2. Methods

### 2.1. Data set acquisition and DEG filtration

The expression profiles of keloid and healthy skin samples were obtained from GEO (<http://www.ncbi.nlm.nih.gov/geo/>), a public database comprising of many gene profiles in various diseases [18]. The datasets GSE7890 [19], GSE145725 [20] and GSE121618 [21] (including 18 keloid samples and 19 healthy skin samples) as training sets to find hub DEGs. In addition, the GSE92566 [22] based on GPL570 included 4 keloid samples and 3 healthy skin samples, as validation set (Table 1). We selected the gene expression dataset GSE92566 from whole tissue biopsy samples of keloid as the validation set for this study, aiming to confirm the expression of key genes in the overall tissue.

The raw data were retrieved as MINiML files. All data were standardized, and the batch effect was removed. Identification of DEGs was achieved with LIMMA v3.40.2 in R language [23]. Significance was defined as a P-value of  $<0.05$ , and  $|\text{Log}_2\text{Fold Change(FC)}|>1$  as the thresholds for the screening of DEGs. The volcano plots were generated by the ggplot2 package in R software [24]. The heatmap in R was used to display the heatmap [25]. Both were visualized through an online database Assistant for Clinical Bioinformatics ([www.aclbi.com](http://www.aclbi.com)). Analyses were performed independently for the keloid and healthy specimens, and the intersection among 3 training datasets was used to determine the DEGs. A Venn diagram of DEGs was drawn, and the overlapping DEGs were retained for further analysis [26].

### 2.2. Functional enrichment analysis of DEGs

Functional enrichment analysis was conducted to further assess the functions of the potential target. Gene Ontology (GO) is the most widely used ontology and involves biological process (BP), molecular function

**Table 1.** Details for GEO data sets.

Category	GEO accession	Platform	Specimen	Keloid	Healthy skin	DEG	
						Up	Down
Training sets	GSE7890 <sup>19</sup>	GPL570	FB	5	5	91	260
	GSE145725 <sup>20</sup>	GPL16043	FB	9	10	212	247
	GSE121618 <sup>21</sup>	GPL21185	EC	5	6	247	415
Validation set	GSE92566 <sup>22</sup>	GPL570	Whole skin biopsies	4	3	–	–

FB: fibroblast; EC: endothelial cell; GEO: Gene Expression Omnibus.

(MF), and cellular component (CC) of human and model organism genes. To analyze gene function and related advanced genomic functions, Kyoto Encyclopedia of Genes and Genomes (KEGG) [27] is a useful resource. The DAVID6.8 bioinformatics database was used to excavate the enrichment pathways of GO and KEGG to further confirm the biological role of DEGs [28]. Bubble diagrams (displayed by R software package ggplot2), and chord diagram from the free online platform Bioinformatics (<http://www.bioinformatics.com.cn>) are both visualization tools useful for highlighting the significantly enriched GO and KEGG terms of DEGs.

### **2.3. Establishment of the PPI network and identification of the hub genes**

The STRING interaction network analysis tool (<https://string-db.org/>) was applied to generate PPI networks [29]. All the identified DEGs were uploaded to STRING's official website to determine the interrelationships among the proteins in order to elucidate the mechanisms of keloid development. Then, the interaction data were downloaded and the PPI network was constructed and displayed with Cytoscape v3.8.0 [30]. The CytoHubba [31] plugin was employed for the identification of hub genes in this network. Density of Maximum Neighborhood Component (DMNC), Maximal Clique Centrality (MCC), Degree, Edge Percolated Component (EPC), Maximum neighborhood component (MNC), Eccentricity, BottleNeck, Radiality, Closeness, Stress, Betweenness, and Clustering Coefficients of twelve algorithms were used to measure the top five hub genes. The intersection genes of the 12 calculation methods were used as the hub genes.

### **2.4. Verification of the identified hub gene expression by total RNA-sequencing (RNA-seq) techniques and a validation microarray data from the GEO database GSE92566**

The hub genes for keloid were screened for further confirmation. For this, a total of 8 participants, including 4 keloid specimens and 4 healthy control specimens extracted from non-lesion sites, were recruited from West China Hospital. There were no significant differences in age, gender and parts of the material between them ( $p > 0.05$ ) (Supplementary Table 1). Our research adheres to the Declaration of Helsinki. The study began on April 29th, 2014, recruitment began on May 16, 2014, and research to this day. Ethical approval was obtained from the Ethics Committee of West China Hospital of Sichuan University, the number

is 2014 REVIEW (NO.65), and informed consent was signed by the patients before specimen collection. Keloid patient inclusion criteria: (a) duration of illness more than two years; (b) clinical diagnosis by an experienced clinician and confirmed by pathological examination; (c) no local infection or ulcer; (d) no nonsurgical treatment received before operation. The exclusion criteria were connective tissue diseases or other fibrous, systemic immune, genetic, or infectious skin diseases.

According to the manufacturer's protocol, total RNA was extracted from tissues using Trizol reagent. The quantity and integrity of the RNA were assessed using Agilent 2200 TapeStation. The ribosomal RNA was digested using the TruSeq Stranded Total RNA with Ribo-Zero Gold kit, and the mRNA was fragmented to approximately 200 bp. Following the TruSeq Stranded Total RNA library preparation kit with Ribo-Zero Gold, the RNA fragments underwent first and second strand cDNA synthesis through adapter ligation and low-cycle enrichment. The purified library products were evaluated using the Agilent 2100 Bioanalyzer. These libraries were sequenced by Ribobio's Illumina at 150 bp paired ends.

Besides, the validation set GSE92566 was also employed to verify the expression levels of hub genes.

### **2.5. Analyses of mRNA-miRNA targets and establishment of the ceRNA network**

Investigating the miRNAs of target mRNAs is critical for elucidating the regulatory mechanism and function of mRNAs. Herein, the upstream regulatory miRNAs of the target hub genes were predicted by employing the online predictive databases miRTarBase v8.0 [32], StarBase3.0 [33], miRDB [34], and miRWalk [35]. The predicted miRNAs are shown in a Venn diagram. In addition, StarBase3.0 was used to predict the upstream molecule lncRNAs of target miRNAs. The screening criteria were: human, mammalian h19 genome, with or without degradome data, and strict stringency ( $\geq 5$ ). Finally, the network of the mRNA-miRNA-lncRNA pairs was visualized using Cytoscape software.

### **2.6. Quantitative analysis of the identified miRNA by miRNA microarray**

Similarly, miRNA microarray data analysis was performed using the four keloid and four healthy control samples collected above to evaluate the expression levels of identified hub miRNA. The clinical baseline data of the samples are shown in Supplementary Table 1. Total RNA was extracted by Trizol reagent.

RNA quantification was performed using NanoDrop ND-2000 (Thermo Scientific) and RNA integrity was detected by Agilent Bioanalyzer 2100 (Agilent Technologies). After the RNA was qualified, the samples were labeled, hybridized and eluted according to the standard process of Agilent Human miRNA, Release 21.0 (8\*60K, Design ID:070156). First, the total RNA is dephosphorylated, denatured, and further labeled with Cyanine-3-CTP(Cy3). The labeled RNA was purified and hybridized with the chip, and the original image was obtained by scanning with Agilent Scanner G2505C (Agilent Technologies) after elution. Feature Extraction software version 10.7.1.1 (Agilent Technologies) was used to process the original image and extract the original data. Next, Genespring software (version 13.1, Agilent Technologies) is used for quantile standardization and subsequent processing.

## 2.7. Statistical analysis

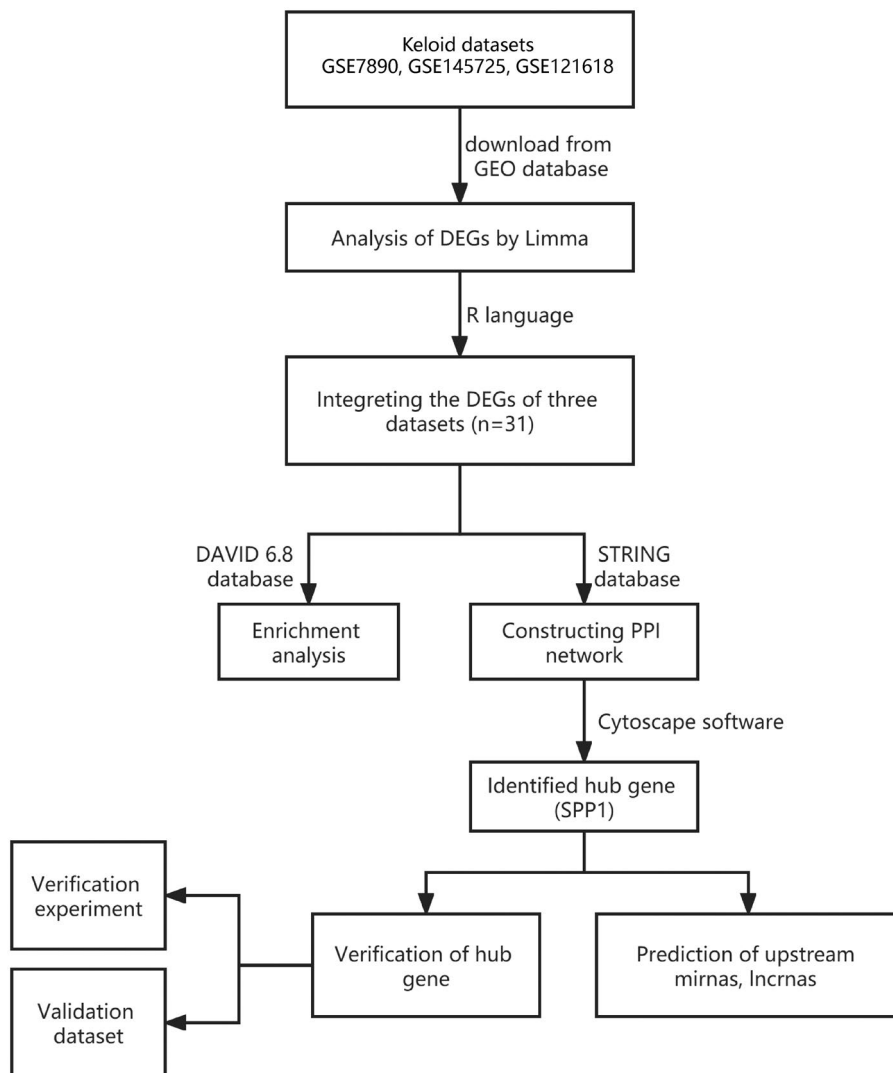
Statistical tests were performed with SPSS v26.0 and GraphPad Prism v8 software. Unpaired t test was utilized for comparing the group's differences.  $p < 0.05$  was deemed statistically significant.

## 3. Results

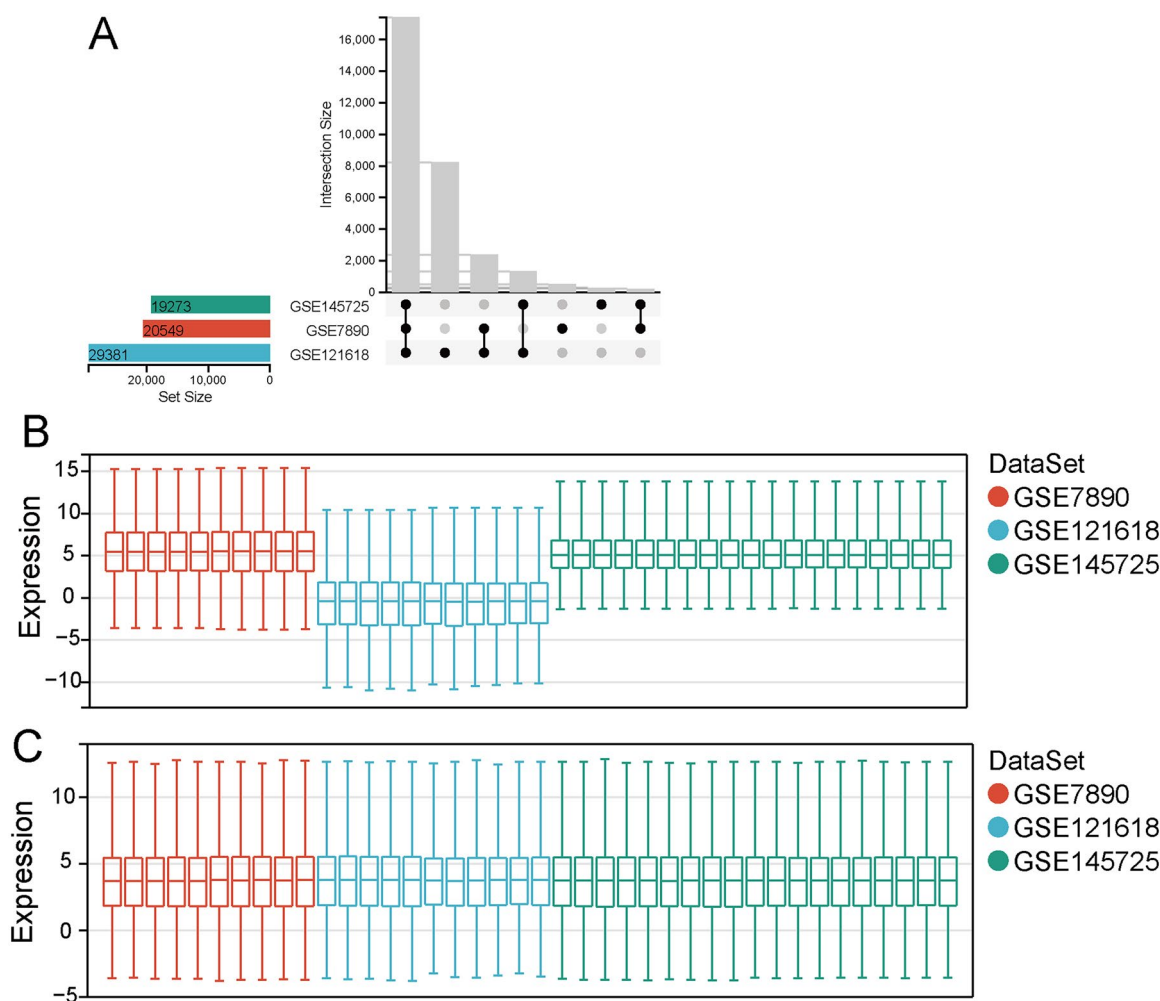
The flowchart of this study is detailed in Figure 1.

### 3.1. Screening of DEGs

For the statistical evaluation of microarray data, the changes in gene expression were analyzed in the R programming language. The robust multiarray average expression measured in R was applied for the



**Figure 1.** Summary graphic illustration of selection of studies. GEO: Gene expression omnibus; DEGs: differentially expressed genes; SPP1: secreted phosphoprotein 1.



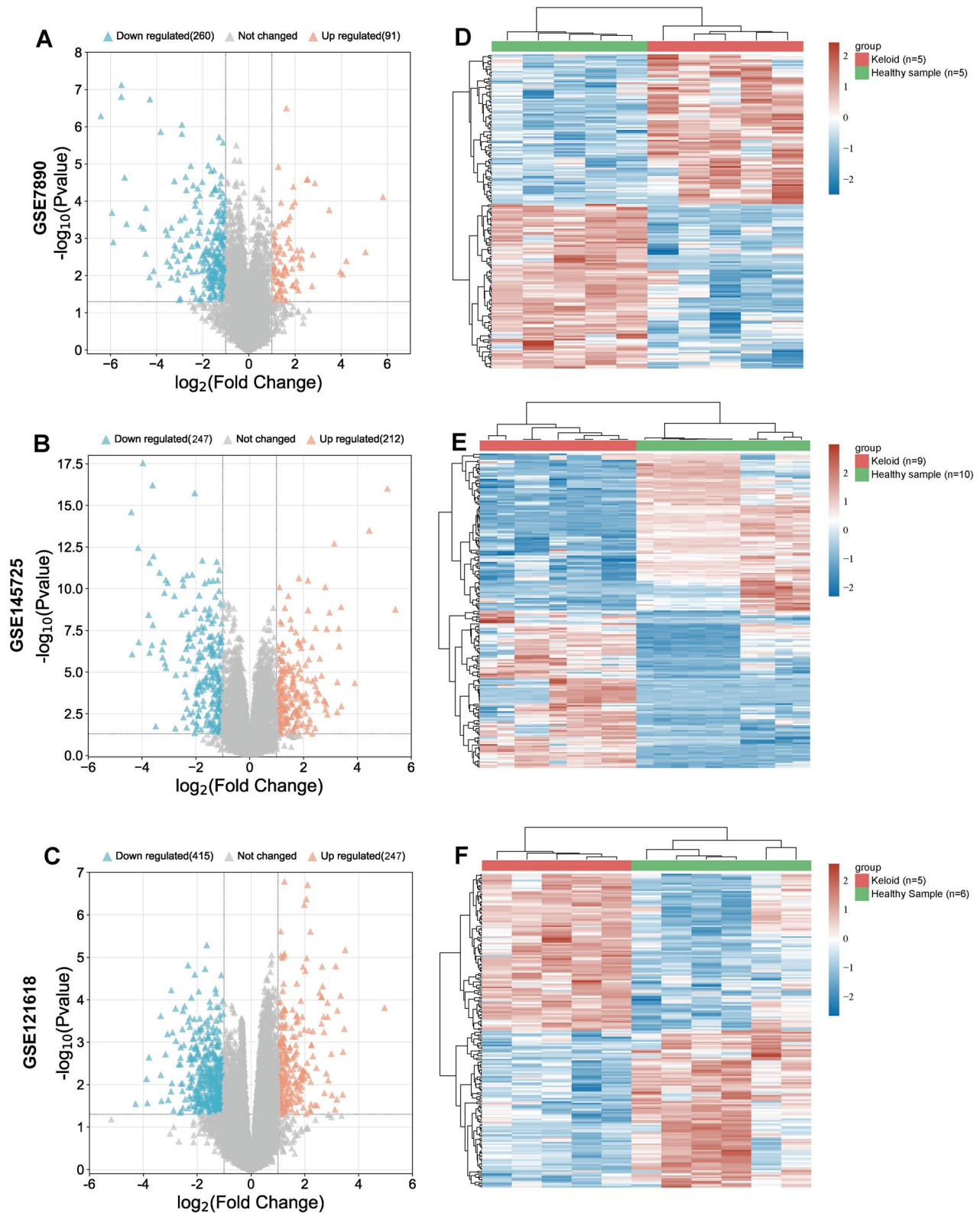
**Figure 2.** Multiple dataset standardization. (a) Intersection between genes in three datasets. (b) The box plots of sample distributions for the three datasets before removing batch effects. (c) The box plots of sample distributions for the three datasets after removing batch effects.

normalization of the obtained data. After normalization, the mean gene expression values of the samples were basically the same (Figure 2). From the box plots, we can observe significant differences in sample distributions among the datasets before batch effect removal, indicating the presence of batch effects. After removing batch effects, the data distributions among the datasets tend to align, with the median values lying on a single line. Overall, 351 DEGs, of which 260 were downregulated and 91 were upregulated, were screened in GSE7890 dataset; 459 DEGs were screened in the GSE145725 dataset, comprising of 247 downregulated genes and 212 upregulated genes; and 662 DEGs were identified in the dataset GSE121618, including 415 downregulated genes and 247 upregulated genes (Supplementary Table 2). Heatmaps and volcano plots of genes with significantly changed expression were generated using the R program (Figure 3). The candidate DEGs acquired from the three datasets were

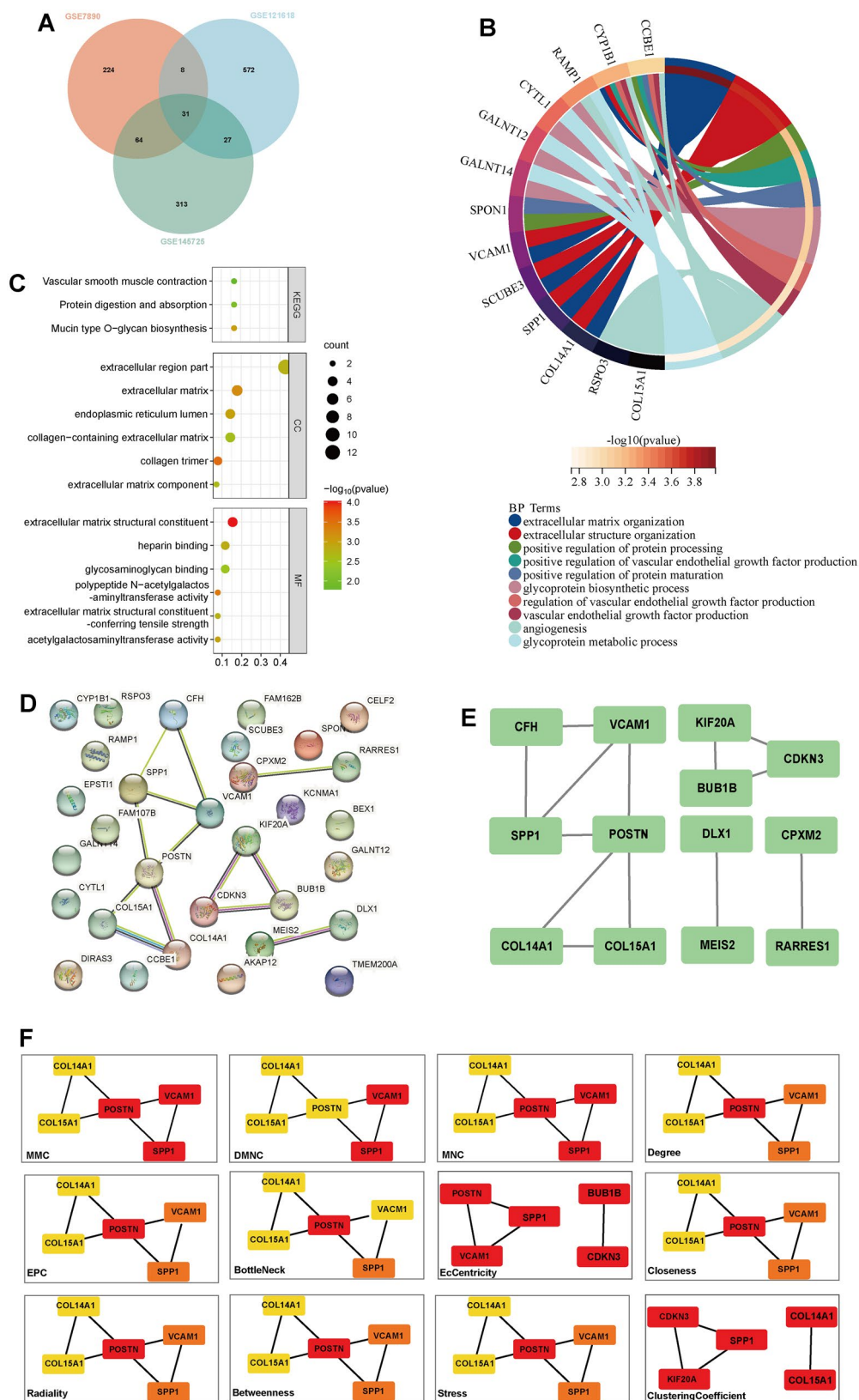
intersected using a Venn diagram (Figure 4(a)). Finally, 31 DEGs were identified.

### 3.2. Results of functional and pathway enrichment analyses for the DEGs

Gene enrichment analysis, including GO and KEGG, was carried out using DAVID to further understand the biological characteristics of these DEGs in a thorough and comprehensive way. The top related enriched terms were selected (Table 2). The significantly enriched GO term analysis included ECM organization, positive regulation of protein processing and mutation etc. KEGG analysis revealed that there were mainly three pathways, such as 'protein digestion and absorption', 'mucin type O-glycan biosynthesis' and 'vascular smooth muscle contraction', in the DEG-enriched pathways (Figure 4(b,c)).



**Figure 3.** Identification Of DEGs. Volcano plots were constructed using fold-change values and adjusted P. The red point in the plot represents the over-expressed DEGs and the blue point indicates the down-expressed DEGs with statistical significance. In the heat maps, red represents up-regulated DEGs and blue represents down-regulated DEGs. Due to the large number of different genes, 50 up-regulated genes and 50 down-regulated genes with the greatest difference were shown here. (a–c) Volcano plot of DEGs from GSE7890, GSE145725 and GSE121618. (d–f) The heat map of DEGs from GSE7890, GSE145725 and GSE121618. DEGs: differentially expressed genes.



**Figure 4.** (a) The Venn diagram shows the DEGs obtained from three dataset (GSE7890, GSE145725 and GSE121618). (b) GO chord plot of the BP for DEGs. The genes are linked to their assigned terms *via* colored ribbons. Genes are ordered according to the observed log-fold change (logFC), which is displayed in descending intensity of red squares displayed next to the selected genes. (c) The enriched GO chord plot of the CC, MF and KEGG signaling pathways of DEGs. The abscissa indicates gene ratio and the enriched pathways were presented in the ordinate. (d) PPI network was processed with string online database. Network nodes represent proteins. The edges represent protein-protein associations. The thickness of the edges represents the strength of the association. (e) PPI network constructed with the DEGs in cytoscape software. (f) 12 CytoHubba plugins were used to analyze significant genes, the top five genes with high score. GO: Gene ontology; BP: biological processes; CC: cellular component; MF: molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes; Protein-protein interaction.

**Table 2.** The significantly enriched KEGG terms top five GO terms for DEGs.

Category	ID	Description	p Value	Count
BP	GO:0030198	Extracellular matrix organization	0.00010432	5
BP	GO:0043062	Extracellular structure organization	0.00020765	5
BP	GO:0010954	Positive regulation of protein processing	0.00074843	2
BP	GO:0010575	Positive regulation of vascular endothelial growth factor production	0.00080316	2
BP	GO:1903319	Positive regulation of protein maturation	0.00085976	2
CC	GO:0005581	Collagen trimer	0.00028422	3
CC	GO:0031012	Extracellular matrix	0.00059028	5
CC	GO:0005788	Endoplasmic reticulum lumen	0.00106099	4
CC	GO:0044421	Extracellular region part	0.00172439	12
CC	GO:0044420	Extracellular matrix component	0.00224766	2
MF	GO:0005201	Extracellular matrix structural constituent	0.00009224	4
MF	GO:0004653	Polypeptide N-acetylgalactosaminyltransferase activity	0.00037998	2
MF	GO:0008376	Acetylgalactosaminyltransferase activity	0.00108855	2
MF	GO:0008201	Heparin binding	0.00145949	3
MF	GO:0030020	Extracellular matrix structural constituent conferring tensile strength	0.00178478	2
KEGG	hsa00512	Mucin type O-glycan biosynthesis	0.00095648	2
KEGG	hsa04974	Protein digestion and absorption	0.00870223	2
KEGG	hsa04270	Vascular smooth muscle contraction	0.01633424	2

GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: biological process; CC: cellular component; MF: molecular function.

**Table 3.** Screening hub gene by twelve CytoHubba algorithms.

Algorithm	Gene name
MCC	VCAM1, <b>SPP1</b> ,POSTN, COL14A1, COL15A1
DMNC	VCAM1, <b>SPP1</b> , COL14A1,POSTN, COL15A1
MNC	VCAM1, <b>SPP1</b> , COL14A1,POSTN, COL15A1
Degree	POSTN,VCAM1, <b>SPP1</b> ,COL14A1, COL15A1
EPC	POSTN,VCAM1, <b>SPP1</b> ,COL14A1, COL15A1
BottleNeck	POSTN, <b>SPP1</b> ,VCAM1,COL14A1, COL15A1
EcCentricity	VCAM1, <b>SPP1</b> ,POSTN, BUB1B, CDKN3
Closeness	POSTN,VCAM1, <b>SPP1</b> ,COL14A1, COL15A1
Radiality	POSTN,VCAM1, <b>SPP1</b> ,COL14A1, COL15A1
Betweenness	POSTN,VCAM1, <b>SPP1</b> ,COL14A1, COL15A1
Stress	POSTN,VCAM1, <b>SPP1</b> ,COL14A1, COL15A1
Clustering Coefficient	COL14A1, COL15A1, <b>SPP1</b> , CDKN3, KIF20A

The significance of the bold values were that different algorithms all find the same key gene 'SPP1'

### 3.3. PPI network establishment and hub gene identification

Thirty-one DEGs were uploaded to the STRING online database to construct the PPI network (Figure 4(d)). The minimum required interaction score was 0.4, the local clustering coefficient was 0.376, and the PPI enrichment P-value was 4.01e-05. The obtained PPI network contains 31 nodes and 13 edges. CytoHubba screened five genes with high clustering coefficients plugin by twelve different algorithms in Cytoscape (Figure 4(e), Table 3). Finally, we crossed the results of the 12 algorithms, and secreted phosphoprotein 1 (SPP1) was selected as a hub gene because it was identified among 12 results (Figure 4(f), Table 4). Therefore, we speculate that the SPP1 may be one of the important factors influencing the development of keloid.

### 3.4. Validation of the identified hub gene

We utilized three independent training sets to assess the mRNA expression levels of SPP1 and found a

significant increase in SPP1 expression in keloids (Figure 5(a-c)). Consistent with the results from the training sets, the mRNA levels of SPP1 in keloid samples were significantly increased in both ourselves RNA-sequencing data ( $p=0.028$ ) and the validation dataset ( $p=0.022$ ), shown in Figure 5(d-e). The receiver operating characteristic (ROC) curves were established to discriminate keloid patients from healthy controls for the gene, and the areas under the ROC curves (AUCs) were used to evaluate the diagnostic accuracy of the identified hub gene. We combined the data from our RNA-sequencing techniques and the GSE92566 validation dataset, followed by ROC curve analysis using ROCR software (Figure 5(f)). These findings demonstrate that SPP1 (AUC = 0.759) has good accuracy as a biomarker for keloid.

### 3.5. Further miRNA prediction and construction of ceRNA networks

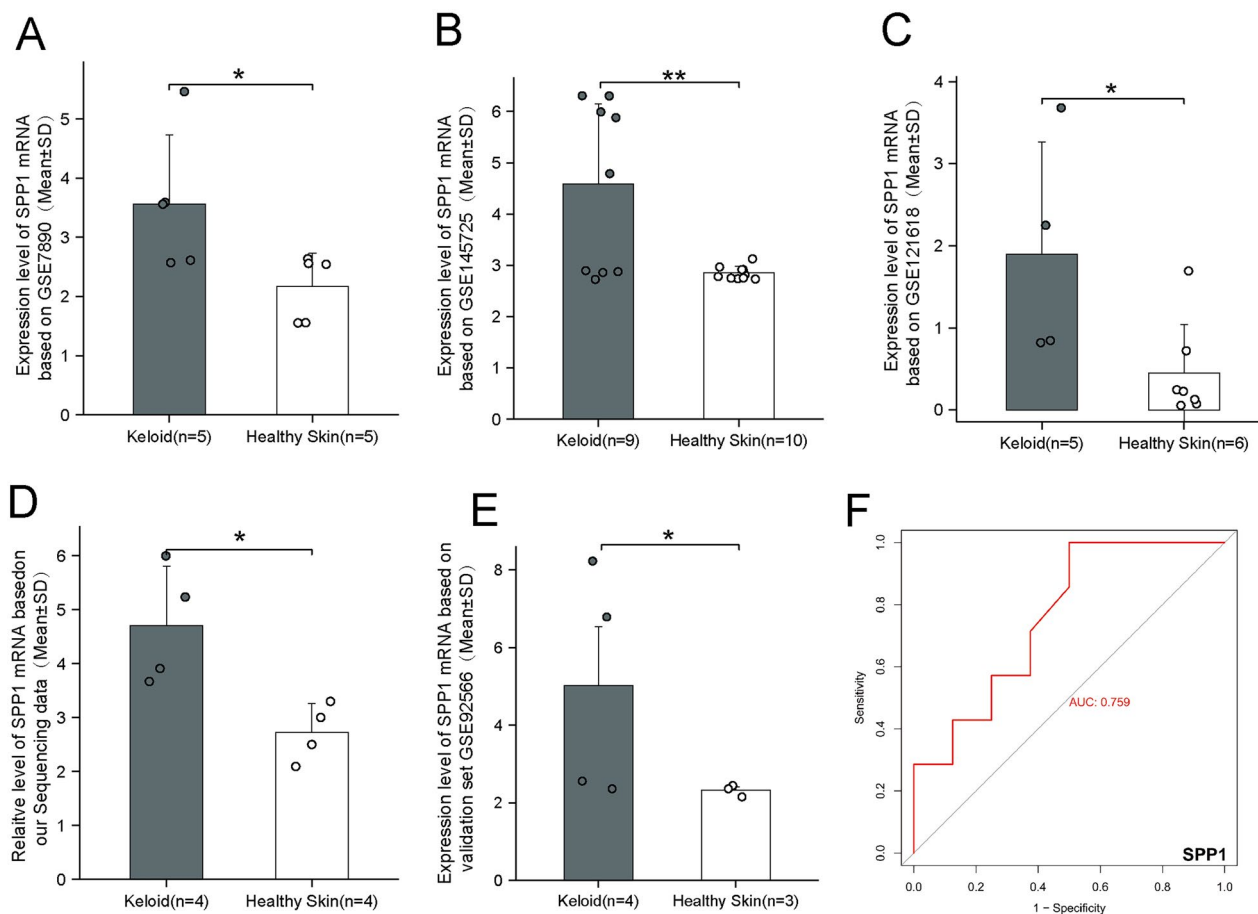
By combining the results of four databases (Supplementary Table 3), only upstream miRNA hsa-miRNA-181a-5p was obtained (Figure 6(a)). In addition, the results of miRNA microarray expression analysis showed that miRNA-181a-5p was upregulated in keloid tissue compared with NS tissue (Figure 6(b)). Recent studies found that lncRNA TRHDE-AS1 could inhibit hypertrophic scar fibroblast (HSF) proliferation and promote HSF apoptosis by downregulating miRNA-181a-5p expression [36]. miRNA-181a has been shown to promote cell proliferation, inhibit apoptosis, and ultimately accelerate the formation of pathological scars [37]. Therefore, miRNA-181a-5p plays a vital role in keloid pathogenesis. Then, StarBase was used to predict the upstream lncRNA molecules that could interact with the selected miRNAs.



**Table 4.** Identified hub gene.

Gene symbol	Full name	GSE7890 <sup>19</sup>		GSE145725 <sup>20</sup>		GSE121618 <sup>21</sup>	
		Log <sub>2</sub> (FC)	p Value	Log <sub>2</sub> (FC)	p Value	Log <sub>2</sub> (FC)	p Value
SPP1	Secreted phosphoprotein 1	2.303	0.003	1.733	0.003	2.226	0.430

FC: fold change.



**Figure 5.** (a) SPP1 Expression in dataset GSE7890. (b) SPP1 expression in dataset GSE145725. (c) SPP1 expression in dataset GSE121618. (d) The expression of SPP1 by our RNA-sequencing techniques. (e) Verification of SPP1 by GSE92566. Student's t-tests were utilized for all the comparison and error bars are mean ± SD. (f) The ROC curve analysis of SPP1 in keloid tissues from validation data of GSE92566 and our sequencing data, AUC = 0.759.

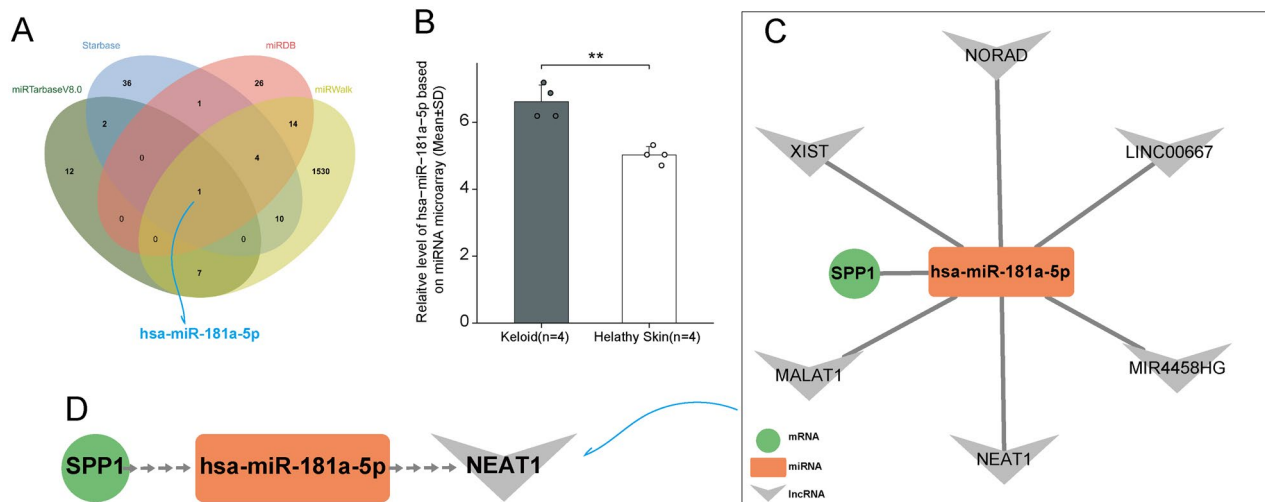
Six target lncRNAs (NEAT1, MALAT1, LINC00667, NORAD, XIST and MIR4458HG) were identified (Figure 6(c)). Finally, a literature search was conducted. NEAT1 was highly expressed in keloid tissues, and NEAT1 knockdown suppressed the formation of keloid [38]. Hence, we propose that NEAT1/miRNA-181a-5p/SPP1 may be a RNA regulatory pathway involved in keloid progression (Figure 6(d)).

#### 4. Discussion

Keloids characterized by fibroblast proliferation and excessive accumulation of ECM have a high incidence around the world, and the global impact is enormous. Although there are a variety of keloid treatment methods, such as

surgical resection, glucocorticoid injection, laser treatment and radiotherapy, the treatment effect is not satisfactory and the recurrence rate is extremely high. In addition, the pathogenesis of keloid has not been fully elucidated, and there is still a lack of radical treatment [10]. Therefore, it is of great clinical significance to find new targets for early treatment and prognostic prediction of keloid. This study aimed to integrate the DEGs between keloid and healthy skin samples from public database, so as to explore the key factors affecting keloid collagen synthesis and angiogenesis. Overall, 31 DEGs were identified, and in-depth analysis was carried out.

In our study, the upregulated hub genes were found to be highly enriched in ECM organization-related



**Figure 6.** (a) Four online predictive miRNA databases: miRTarBaseV8.0, StarBase3.0, miRDB and miRWalk were used to screen the target miRNA of hub gene SPP1. The Venn diagram showed an overlap of a miRNA: hsa-miR-181a-5p. (b) The expression level of has-miR-181a-5p by the miRNA microarray ( $p=0.001$ ). (c) The co-expressed network and ceRNA network. The green ellipse represents the hub gene, orange round rectangle represent the target miRNA and gray V represents the target lncRNAs. (d) The potential RNA regulated pathway NEAT1/has-miR-181a-5p/SPP1. MCC: Maximal Clique Centrality; DMNC: Density of Maximum Neighborhood Component; MNC: Maximum neighborhood component; EPC: Edge Percolated Component. miRNA: microRNA; ceRNA: competing endogenous RNAs; lncRNA: long non-coding RNA.

biological functions and pathways, and the main component of the ECM was collagen. Previous studies have found that the protein encoded by the SPP1 gene is responsible for the attachment of osteoclasts to the mineralized bone matrix, and is associated with FB activation [39]. Numerous studies have confirmed that SPP1 participates in promoting tumor cell proliferation, migration, and invasion [40, 41]. Kramerova et al. reported that SPP1 could upregulate collagen expression in FBs [42]. Similarly, Bi et al. demonstrated that the function of hub gene was highly enriched in keloids and collagen synthesis [43]. At the same time, the latest research shows that SPP1 stimulation *in vitro* can increase the expression of collagen in FBs and the proportion of CTHRC1+ FBs [44]. Herein, we found that the up-regulated gene SPP1 might activate FB during the formation of keloids, resulting in the synthesis and deposition of a large number of collagen fibers, thus promoting the synthesis of ECM. Our RNA-sequencing results indicated that the expression of SPP1 was increased in keloid patients, which is consistent with our conjectured results. Thus, we hold the opinion that SPP1 plays a vital role in keloid pathogenesis and may be a novel biomarker and therapeutic target, which can be useful for understanding the pathogenesis of keloids. Inhibition of scar formation by blocking the interaction between SPP1 and its receptor through molecular targeted drugs of SPP1.

By using predictive online databases, the upstream miRNAs of SPP1 were screened. It was found that hsa-miRNA-181a-5p directly targeted SPP1. Then, our

miRNA microarray analysis revealed that the expression of hsa-miRNA-181a-5p was remarkably upregulated ( $p=0.0013$ ) in keloid tissue. Wei et al. confirmed that the upregulated level of hsa-miRNA-181a-5p promoted FB proliferation and inhibited it by directly binding to PTEN mRNA and downregulating PTEN expression. In addition, several studies have suggested that hsa-miRNA-181a-5p is involved in cellular proliferation, invasion, and migration [45, 46]. Although hsa-miRNA-181a-5p has not been reported in keloid, we speculate that miRNA-181a-5p can regulate SPP1 by suppressing the translation of SPP1, thereby affecting keloid formation. Finally, we gradually reverse predicted the upstream lncRNAs (NEAT1, MALAT1, LINC00667, NORAD, XIST, and MIR4458HG) from hsa-miR181a-5. There have been reports of the upregulated expression of NEAT1 in keloid, suggesting the potential role of NEAT1 as an effective therapeutic target. Previous research has shown that lncRNAs act as sponges for miRNAs, thus interfering with mRNA expression [47]. Collectively, we predict that lncRNA NEAT1 is directly connected to miRNA-181a-5p as a molecular sponge, eliminating its inhibitory effect on SPP1 mRNA translation, thereby promoting collagen synthesis and deposition and inducing scar formation. Furthermore, we successfully constructed a ceRNA regulatory network, in which NEAT1/miRNA-181a-5p/SPP1 may be a potential regulatory mechanism of keloid formation, and we can inhibit keloid progression by regulating gene expression on this axis.

To the best of our knowledge, this is the first work to determine the association between SPP1 and keloid. Compared with other studies, by integrating the differences in the transcription levels of keloid fibroblast and endothelial cell line, the potential therapeutic targets and the pathways of target genes are identified, which regulate the occurrence and development of keloids. There are, however, few limitations that need to be addressed. First, this study was conducted mainly *via* bioinformatic analyses, which requires to be verified by further experimental studies. Second, the number of samples for validation was limited. Therefore, follow-up studies are warranted to validate our findings.

## 5. Conclusions

In our study, we analysed three keloid microarray datasets to comprehensively analyse the genes that drive keloid development and minimize the high false-positive rate that is often reported when using single microarray analyses. We speculate that the SPP1 gene might be a novel biomarker for the treatment of keloid and that NEAT1/miR-181a-5p/SPP1 might be a potential RNA regulatory pathway that controls disease progression, providing insight into the mechanisms of keloid disease development at the transcriptome level.

## Acknowledgements

The authors would like to express their gratitude to EditSprings (<https://www.editsprings.cn>) for the expert linguistic services provided. And the authors thank all the investigators and the patients who participated in the studies.

## Authors' contributions

Ruxin Xie, Chenyu Li, Jiao Yun and Shiwei Zhang drafted the manuscript, participated in the design of the study and performed the statistical analysis. Junjie Chen, Zhengyong Li and Ying Cen conceived of the study, revised it critically for intellectual content and final approval of the version to be published. Chenyu Li, Jiao Yun, Ai Zhong and Junliang Wu helped to draft the manuscript. All authors agree to be accountable for all aspects of the work.

## Disclosure statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This work was supported by the Natural Science Foundation of Sichuan Province of China under Grant 23NSFSC0834; 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University under Grant ZYPY20001 and ZYPY20002.

## ORCID

Ruxin Xie  <http://orcid.org/0000-0003-3561-5884>

## Data availability statement

The analysis data in all datasets in this paper were downloaded from the public free data platform GEO. In addition, our experimental data during the current study are available from the corresponding author on reasonable request.

## References

- [1] Yu X, Li Z, Chan MT, et al. MicroRNA deregulation in keloids: an opportunity for clinical intervention? *Cell Prolif.* 2015;48(6):626–630. doi: [10.1111/cpr.12225](https://doi.org/10.1111/cpr.12225).
- [2] Renz P, Hasan S, Gresswell S, et al. Dose effect in adjuvant radiation therapy for the treatment of resected keloids. *Int J Radiat Oncol Biol Phys.* 2018;102(1):149–154. doi: [10.1016/j.ijrobp.2018.05.027](https://doi.org/10.1016/j.ijrobp.2018.05.027).
- [3] Mustoe TA. Scars and keloids. *BMJ.* 2004;328(7452):1329–1330. doi: [10.1136/bmj.328.7452.1329](https://doi.org/10.1136/bmj.328.7452.1329).
- [4] Mamalis AD, Lev-Tov H, Nguyen DH, et al. Laser and light-based treatment of keloids—a review. *J Eur Acad Dermatol Venereol.* 2014;28(6):689–699. doi: [10.1111/jdv.12253](https://doi.org/10.1111/jdv.12253).
- [5] Liu T, Ma X, Ouyang T, et al. Efficacy of 5-aminolevulinic acid-based photodynamic therapy against keloid compromised by downregulation of SIRT1-SIRT3-SOD2-mROS dependent autophagy pathway. *Redox Biol.* 2019;20:195–203. doi: [10.1016/j.redox.2018.10.011](https://doi.org/10.1016/j.redox.2018.10.011).
- [6] Marttala J, Andrews JP, Rosenbloom J, et al. Keloids: animal models and pathologic equivalents to study tissue fibrosis. *Matrix Biol.* 2016;51:47–54. doi: [10.1016/j.matbio.2016.01.014](https://doi.org/10.1016/j.matbio.2016.01.014).
- [7] Berman B, Garikaparthi S, Smith E, et al. A novel hydrogel scaffold for the prevention or reduction of the recurrence of keloid scars postsurgical excision. *J Am Acad Dermatol.* 2013;69(5):828–830. doi: [10.1016/j.jaad.2013.06.025](https://doi.org/10.1016/j.jaad.2013.06.025).
- [8] Coentro JQ, Pugliese E, Hanley G, et al. Current and upcoming therapies to modulate skin scarring and fibrosis. *Adv Drug Deliv Rev.* 2019;146:37–59. doi: [10.1016/j.addr.2018.08.009](https://doi.org/10.1016/j.addr.2018.08.009).
- [9] Har-Shai Y, Mettanes I, Zilberstein Y, et al. Keloid histopathology after intralesional cryosurgery treatment. *J Eur Acad Dermatol Venereol.* 2011;25(9):1027–1036. doi: [10.1111/j.1468-3083.2010.03911.x](https://doi.org/10.1111/j.1468-3083.2010.03911.x).
- [10] Li C, Jin M, Luo Y, et al. Integrated bioinformatics analysis of core regulatory elements involved in keloid formation. *BMC Med Genom.* 2021;14(1):239. doi: [10.1186/s12920-021-01087-7](https://doi.org/10.1186/s12920-021-01087-7).

- [11] Wang J, Wu H, Xiao Z, et al. Expression profiles of lncRNAs and circRNAs in keloid. *Plast Reconstr Surg Glob Open*. 2019;7(6):e2265. doi: [10.1097/gox.0000000000002265](https://doi.org/10.1097/gox.0000000000002265).
- [12] Niessen FB, Spauwen PH, Schalkwijk J, et al. On the nature of hypertrophic scars and keloids: a review. *Plast Reconstr Surg*. 1999;104(5):1435–1458. doi: [10.1097/00006534-199910000-00031](https://doi.org/10.1097/00006534-199910000-00031).
- [13] Bettinger DA, Yager DR, Diegelmann RF, et al. The effect of TGF-beta on keloid fibroblast proliferation and collagen synthesis. *Plast Reconstr Surg*. 1996;98(5):827–833. doi: [10.1097/00006534-199610000-00012](https://doi.org/10.1097/00006534-199610000-00012).
- [14] Matsumoto NM, Peng W-X, Aoki M, et al. Histological analysis of hyalinised keloidal collagen formation in earlobe keloids over time: collagen hyalinisation starts in the perivascular area. *Int Wound J*. 2017;14(6):1088–1093. doi: [10.1111/iwj.12763](https://doi.org/10.1111/iwj.12763).
- [15] Kurokawa N, Ueda K, Tsuji M. Study of microvascular structure in keloid and hypertrophic scars: density of microvessels and the efficacy of three-dimensional vascular imaging. *J Plast Surg Hand Surg*. 2010;44(6):272–277. doi: [10.3109/2000656x.2010.532923](https://doi.org/10.3109/2000656x.2010.532923).
- [16] Arjunan S, Gan SU, Choolani M, et al. Inhibition of growth of Asian keloid cells with human umbilical cord Wharton's jelly stem cell-conditioned medium. *Stem Cell Res Ther*. 2020;11(1):78. doi: [10.1186/s13287-020-01609-7](https://doi.org/10.1186/s13287-020-01609-7).
- [17] Wang Q, Wang P, Qin Z, et al. Altered glucose metabolism and cell function in keloid fibroblasts under hypoxia. *Redox Biol*. 2021;38:101815. doi: [10.1016/j.redox.2020.101815](https://doi.org/10.1016/j.redox.2020.101815).
- [18] Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol*. 2016;1418:93–110. doi: [10.1007/978-1-4939-3578-9\\_5](https://doi.org/10.1007/978-1-4939-3578-9_5).
- [19] Smith JC, Boone BE, Opalenik SR, et al. Gene profiling of keloid fibroblasts shows altered expression in multiple fibrosis-associated pathways. *J Invest Dermatol*. 2008;128(5):1298–1310. doi: [10.1038/sj.jid.5701149](https://doi.org/10.1038/sj.jid.5701149).
- [20] Kang Y, Roh MR, Rajadurai S, et al. Hypoxia and HIF-1 $\alpha$  regulate collagen production in keloids. *J Invest Dermatol*. 2020;140(11):2157–2165. doi: [10.1016/j.jid.2020.01.036](https://doi.org/10.1016/j.jid.2020.01.036).
- [21] Matsumoto NM, Aoki M, Okubo Y, et al. Gene expression profile of isolated dermal vascular endothelial cells in keloids. *Front Cell Dev Biol*. 2020;8:658. doi: [10.3389/fcell.2020.00658](https://doi.org/10.3389/fcell.2020.00658).
- [22] Fuentes-Duculan J, Bonifacio KM, Suárez-Fariñas M, et al. Aberrant connective tissue differentiation towards cartilage and bone underlies human keloids in African Americans. *Exp Dermatol*. 2017;26(8):721–727. doi: [10.1111/exd.13271](https://doi.org/10.1111/exd.13271).
- [23] Dalangood S, Zhu Z, Ma Z, et al. Identification of glycogene-type and validation of ST3GAL6 as a biomarker predicts clinical outcome and cancer cell invasion in urinary bladder cancer. *Theranostics*. 2020;10(22):10078–10091. doi: [10.7150/thno.48711](https://doi.org/10.7150/thno.48711).
- [24] Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis. *Meas-Interdiscip Res*. 2019;17(3):160–167. doi: [10.1080/15366367.2019.1565254](https://doi.org/10.1080/15366367.2019.1565254).
- [25] Li GM, Zhang CL, Rui RP, et al. Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy. *Eur Rev Med Pharmacol Sci*. 2018;22(11):3553–3569. doi: [10.26355/eurev\\_201806\\_15182](https://doi.org/10.26355/eurev_201806_15182).
- [26] Bardou P, Mariette J, Escudié F, et al. jvenn: an interactive Venn diagram viewer. *BMC Bioinf*. 2014;15(1):293. doi: [10.1186/1471-2105-15-293](https://doi.org/10.1186/1471-2105-15-293).
- [27] Kanehisa M, Sato Y, Furumichi M, et al. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019;47(D1):D590–D595. doi: [10.1093/nar/gky962](https://doi.org/10.1093/nar/gky962).
- [28] Huang DW, Sherman BT, Tan Q, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. 2007;35(Web Server issue):W169–75. doi: [10.1093/nar/gkm415](https://doi.org/10.1093/nar/gkm415).
- [29] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–D613. doi: [10.1093/nar/gky1131](https://doi.org/10.1093/nar/gky1131).
- [30] Otasek D, Morris JH, Bouças J, et al. Cytoscape automation: empowering workflow-based network analysis. *Genome Biol*. 2019;20(1):185. doi: [10.1186/s13059-019-1758-4](https://doi.org/10.1186/s13059-019-1758-4).
- [31] Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8 Suppl 4(Suppl 4):S11. doi: [10.1186/1752-0509-8-s4-s11](https://doi.org/10.1186/1752-0509-8-s4-s11).
- [32] Huang H-Y, Lin Y-C-D, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48(D1):D148–D154. doi: [10.1093/nar/gkz896](https://doi.org/10.1093/nar/gkz896).
- [33] Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–7. doi: [10.1093/nar/gkt1248](https://doi.org/10.1093/nar/gkt1248).
- [34] Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48(D1):D127–D131. doi: [10.1093/nar/gkz757](https://doi.org/10.1093/nar/gkz757).
- [35] Dweep H, Sticht C, Pandey P, et al. miRWalk-database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform*. 2011;44(5):839–847. doi: [10.1016/j.jbi.2011.05.002](https://doi.org/10.1016/j.jbi.2011.05.002).
- [36] Wei Y, Wang T, Zhang N, et al. LncRNA TRHDE-AS1 inhibit the scar fibroblasts proliferation via miR-181a-5p/PTEN axis. *J Mol Histol*. 2021;52(2):419–426. doi: [10.1007/s10735-021-09968-y](https://doi.org/10.1007/s10735-021-09968-y).
- [37] Rang Z, Wang ZY, Pang QY, et al. MiR-181a targets PHLPP2 to augment AKT signaling and regulate proliferation and apoptosis in human keloid fibroblasts. *Cell Physiol Biochem*. 2016;40(3–4):796–806. doi: [10.1159/000453139](https://doi.org/10.1159/000453139).
- [38] Yang J, Deng P, Qi Y, et al. NEAT1 knockdown inhibits keloid fibroblast progression by miR-196b-5p/FGF2 axis. *J Surg Res*. 2021;259:261–270. doi: [10.1016/j.jss.2020.09.038](https://doi.org/10.1016/j.jss.2020.09.038).
- [39] Morse C, Tabib T, Sembrat J, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J*. 2019;54(2):1802441. doi: [10.1183/13993003.02441-2018](https://doi.org/10.1183/13993003.02441-2018).
- [40] Deng G, Zeng F, Su J, et al. BET inhibitor suppresses melanoma progression via the noncanonical NF- $\kappa$ B/SPP1 pathway. *Theranostics*. 2020;10(25):11428–11443. doi: [10.7150/thno.47432](https://doi.org/10.7150/thno.47432).

- [41] Wang Y, Su J, Wang Y, et al. The interaction of YBX1 with G3BP1 promotes renal cell carcinoma cell metastasis via YBX1/G3BP1-SPP1-NF- $\kappa$ B signaling axis. *J Exp Clin Cancer Res.* 2019;38(1):386. doi: [10.1186/s13046-019-1347-0](https://doi.org/10.1186/s13046-019-1347-0).
- [42] Kramerova I, Kumagai-Cresse C, Ermolova N, et al. Spp1 (osteopontin) promotes TGF $\beta$  processing in fibroblasts of dystrophin-deficient muscles through matrix metalloproteinases. *Hum Mol Genet.* 2019;28(20):3431–3442. doi: [10.1093/hmg/ddz181](https://doi.org/10.1093/hmg/ddz181).
- [43] Bi S, Liu R, Wu B, et al. Bioinformatic analysis of key genes and pathways related to keloids. *Biomed Res Int.* 2021;2021:5897907–5897911. doi: [10.1155/2021/5897907](https://doi.org/10.1155/2021/5897907).
- [44] Liu J, Huang Y, Gong Y, et al. CTHRC1+ fibroblasts are stimulated by macrophage-secreted SPP1 to induce excessive collagen deposition in keloids. *Clin Transl Med.* 2022;12(12):e11115. doi: [10.1002/ctm2.1115](https://doi.org/10.1002/ctm2.1115).
- [45] Wu L, Song W-Y, Xie Y, et al. miR-181a-5p suppresses invasion and migration of HTR-8/SVneo cells by directly targeting IGF2BP2. *Cell Death Dis.* 2018;9(2):16. doi: [10.1038/s41419-017-0045-0](https://doi.org/10.1038/s41419-017-0045-0).
- [46] Nakamura A, Rampersaud YR, Nakamura S, et al. MicroRNA-181a-5p antisense oligonucleotides attenuate osteoarthritis in facet and knee joints. *Ann Rheum Dis.* 2019;78(1):111–121. doi: [10.1136/annrheumdis-2018-213629](https://doi.org/10.1136/annrheumdis-2018-213629).
- [47] Xiao H, Tang K, Liu P, et al. LncRNA MALAT1 functions as a competing endogenous RNA to regulate ZEB2 expression by sponging miR-200s in clear cell kidney carcinoma. *Oncotarget.* 2015;6(35):38005–38015. doi: [10.18632/oncotarget.5357](https://doi.org/10.18632/oncotarget.5357).