

# Secondary data mining of GEO database for long non-coding RNA and Competing endogenous RNA network in keloid-prone individuals

Yu Deng<sup>1,\*</sup>, Yangbin Xu<sup>1,\*</sup>, Shuqia Xu<sup>1,\*</sup>, Yujing Zhang<sup>1</sup>, Bing Han<sup>1</sup>, Zheng Liu<sup>1</sup>, Xiangxia Liu<sup>1</sup>, Zhaowei Zhu<sup>1</sup>

<sup>1</sup>Department of Plastic Surgery, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou 510080, China

\*Co-first authors

**Correspondence to:** Zhaowei Zhu, Xiangxia Liu, Yangbin Xu; **email:** [nmtmemoir@aliyun.com](mailto:nmtmemoir@aliyun.com), <https://orcid.org/0000-0003-4469-5417>; [liuxsha@mail.sysu.edu.cn](mailto:liuxsha@mail.sysu.edu.cn), <https://orcid.org/0000-0003-0887-5170>; [xuyangbin@hotmail.com](mailto:xuyangbin@hotmail.com), <https://orcid.org/0000-0002-2587-9619>

**Keywords:** keloid-prone, lncRNA, GEO database

**Received:** April 14, 2020

**Accepted:** August 25, 2020

**Published:** November 16, 2020

**Copyright:** © 2020 Deng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

This study aimed to identify long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and messenger RNAs (mRNAs) differentially expressed (DE) during keloid formation, predict DElncRNA-DEmiRNA-DEmRNA interactions, and construct a competing endogenous RNA (ceRNA) network through secondary data mining of keloid-related sequencing and microarray data in the open-source Gene Expression Omnibus (GEO) database. The GSE113621 dataset was downloaded from the GEO database,  $|\log_2FC| > 1$  and  $p < 0.05$  were set as screening criteria, genes expressed only in keloid-prone individuals were selected as research objects, and DElncRNAs, DEmiRNAs, and DEmRNAs before injury and 6 weeks after injury were screened. A Pearson correlation coefficient (PCC) of  $> 0.95$  was selected as the index to predict the targeting relationships among lncRNAs, miRNAs, and mRNAs; and a network diagram was constructed using Cytoscape. The expression of 2356 lncRNAs was changed in the keloid-prone group—1306 were upregulated and 1050 were downregulated. Six lncRNAs, namely, 2 upregulated (DLEU2 and AP000317.2) and 4 downregulated (ADIRF-AS1, AC006333.2, AL137127.1 and LINC01725) lncRNAs, were expressed only in the keloid-prone group and were used to construct a ceRNA network. DLEU2 may regulate fibroblast proliferation, differentiation, and apoptosis through hsa-miR-30a-5p/hsa-miR-30b-5p. In-depth mining of GEO data indicated that lncRNAs and a ceRNA regulatory network participate in the wound healing process in keloid-prone individuals, possibly providing novel intervention targets and treatment options for keloid scars.

## INTRODUCTION

Both hypertrophic scars and keloids are pathological scars. Due to abnormal proliferation of dermal fibrous tissue, excessive deposition of collagen during the wound healing process often affects the appearance of patients and even affects local function [1]. Although many studies have addressed the mechanism underlying keloid occurrence and development and the treatment of keloids, the therapeutic effect of existing treatments on

keloids is still not ideal [2]. Gene chips and sequencing are powerful tools currently used for studying gene expression profiles; however, due to limitations of the manufacturing process and algorithms for analysis of gene chips during the early stage of this technology, the required data could not be acquired completely [3]. As bioinformatic analysis methods have continually developed and the understanding of epigenetics has increased, many scholars have mined new transcriptome information from old microarray or sequencing data by

improving the original algorithms and upgrading the databases to offer new ideas for clinical and scientific research [3–7].

In this study, sequencing and microarray data related to keloid formation were downloaded from the open-source National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database, and secondary data analysis was conducted by combining the R language with network database bioinformatic analysis technology. This analysis aimed to identify the coding RNAs (messenger RNAs, mRNAs) and non-coding RNAs (ncRNAs), including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), expressed during keloid formation and to predict their expression and relationships. Through this research, we intended to verify the possibility of secondary bioinformatic analysis of sequencing and microarray data to extract lncRNAs and construct a competing endogenous RNA (ceRNA) network and simultaneously mine these data for information related to regulation of the pathophysiological process of keloid formation. This study offers new research methods for more comprehensive exploration of keloid pathogenesis and treatment.

## RESULTS

### Changes in mRNAs and ncRNAs in skin wounds of keloid-prone individuals and control individuals 6 weeks after injury

Secondary mining of the RNA sequencing data and miRNA array data was conducted for different individuals. In the 8 keloid-prone individuals, 3039 DEmRNAs were identified—1479 were upregulated and 1560 were downregulated 6 weeks after injury compared with before injury (Figure 1B). Heatmap of lncRNA expression before and after injury in Keloid-prone groups were shown in Figure 1C. In the 6 control group individuals, 2802 DEmRNAs, namely, 1436 upregulated (Figure 1A) and 1366 downregulated mRNAs. A total of 2356 DElncRNAs, namely, 1306 upregulated and 1050 downregulated lncRNAs were identified in the keloid-prone group; 2547 DElncRNAs—1479 upregulated and 1068 downregulated lncRNAs were identified in the control group. Analysis of the miRNA array data revealed 252 DEmiRNAs (151 upregulated and 101 downregulated miRNAs) in keloid-prone patients and 177 DEmiRNAs in the control group (95 upregulated and 82 downregulated miRNAs).

### GO enrichment analysis of mRNAs expressed only in keloid-prone individuals 6 weeks after injury

To more comprehensively study the effect of DEmRNAs on biological activities, BP category analysis was

performed. The DEmRNAs were enriched mainly in the BP terms muscle contraction, presynapse assembly, regulation of membrane potential, and presynapse organization. The top 10 enriched terms are listed in Table 1-BP. In the CC category analysis, the DEmRNAs were enriched mainly in the CC terms synaptic membrane, postsynaptic membrane, intrinsic component of synaptic membrane, and glutamatergic synapse. The top 10 enriched terms are listed in Table 1-CC. In the MF category analysis, the DEmRNAs were enriched mainly in the MF terms channel activity, substrate-specific channel activity, passive transmembrane transporter activity, and ligand-gated ion channel activity. The top 10 enriched terms are listed in Table 1-MF. The GO enrichment analysis results are summarized in Figure 1C.

### KEGG enrichment analysis of mRNAs expressed only in keloid-prone individuals 6 weeks after injury

To more comprehensively analyse the roles played by the DEmRNAs, we performed KEGG pathway analysis. The DEmRNAs were enriched mainly in the neuroactive ligand-receptor interaction, nicotine addiction, calcium signalling pathway, and glutamatergic synapse pathways. The enrichment of DEmRNAs is shown in Figure 1D and Table 2. The gene names and expression heatmaps for each pathway are shown in Figure 2.

### Analysis of the lncRNA-miRNA-mRNA network present only in keloid-prone individuals after injury

Volcano plots and heatmap of lncRNA expression in keloid-prone individuals were shown in Figure 3. The StarBase database was used to predict and compare the targets of DElncRNAs and DEmiRNAs, with  $PCC > 0.95$  as the criterion, to construct a lncRNA-miRNA-mRNA molecular interaction diagram (Figure 4A–4C). Six DElncRNAs—2 upregulated (DLEU2 and AP000317.2) and 4 downregulated (ADIRF-AS1, AC006333.2, AL137127.1, and LINC01725)—were screened (Table 3). Additionally, 65 DEmiRNAs were altered (31 were upregulated and 5 were downregulated) as the lncRNAs were altered, and 43 DEmRNAs were altered (13 were upregulated and 30 were downregulated) as the lncRNAs were altered.

### Expression of DElncRNAs in fibroblasts in keloid scars and normal skin

Expression of all 6 DElncRNAs except for AP000317.2 was detected in fibroblasts in both keloid scars and normal skin. Specifically, the expression of DLEU2 and ADIRF-AS1 were in significantly different expression levels between fibroblasts and keloid. In contrast, the differences in the expression levels of AC006333.2, AL137127.1, and LINC01725

**Table 1. GO Analysis of mRNAs specifically expressed in keloid-prone individuals 6 weeks after injury.**

	Term ID	Term description	Ct.	Adjusted p-value
BP	GO:0006936	muscle contraction	35	0.001473
	GO:0099054	presynapse assembly	11	0.001702
	GO:0042391	regulation of membrane potential	37	0.001702
	GO:0099172	presynapse organization	11	0.001702
	GO:0035270	endocrine system development	18	0.001702
	GO:0035418	protein localization to synapse	12	0.002476
	GO:0051952	regulation of amine transport	13	0.002476
	GO:0071772	response to BMP	19	0.003641
	GO:0071773	cellular response to BMP stimulus	19	0.003641
	GO:0015837	amine transport	13	0.003706
CC	GO:0097060	synaptic membrane	45	2.00E-06
	GO:0045211	postsynaptic membrane	37	4.25E-06
	GO:0099240	intrinsic component of synaptic membrane	21	0.001069
	GO:0098978	glutamatergic synapse	34	0.001069
	GO:0005858	axonemal dynein complex	6	0.00128
	GO:0098936	intrinsic component of postsynaptic membrane	17	0.001407
	GO:1902495	transmembrane transporter complex	28	0.001456
	GO:0098793	presynapse	38	0.001534
	GO:1990351	transporter complex	28	0.001534
	GO:0034703	cation channel complex	21	0.001567
MF	GO:0015267	channel activity	37	0.000261
	GO:0022838	substrate-specific channel activity	35	0.000261
	GO:0022803	passive transmembrane transporter activity	37	0.000261
	GO:0015276	ligand-gated ion channel activity	17	0.000513
	GO:0022834	ligand-gated channel activity	17	0.000513
	GO:0005216	ion channel activity	33	0.000517
	GO:0008509	anion transmembrane transporter activity	27	0.000614
	GO:0005261	cation channel activity	27	0.000708
	GO:0015077	monovalent inorganic cation transmembrane transporter activity	27	0.000879
	GO:0046873	metal ion transmembrane transporter activity	32	0.001416

Ct.: number of genes

between the two groups were not statistically significant (Figure 4D).

## DISCUSSION

Keloid scarring is a clinical condition that is difficult to cure. Although keloids are benign tumours, they often extend past the extent of the damage, invading the surrounding normal skin, and seriously affecting the appearance of the skin after healing [4]. Keloids are most common in people between the ages of 10 and 30 years, are especially common in individuals of African, Hispanic, and Asian ethnicity, and frequently occur on the chest, earlobes, shoulders, and back [26]. Although different treatment methods, such as radiotherapy, hormone therapy, and surgical resection are available, keloids still have a relatively high recurrence rate [8]. Many scholars believe that fibroblasts are a main participant in the occurrence and development of

keloids. After skin injury, a complicated regulatory signalling network is activated to regulate the proliferation, migration, and secretion of fibroblasts. Therefore, the biological behaviour of fibroblasts regarding the mechanism of keloid formation is a trending research topic [9]. Gene chips and sequencing are important methods for studying gene transcription profiles and are widely used in regenerative medicine [30], disease [27, 10–11], and cancer [7] research. However, the focus of transcriptome research has recently shifted gradually from protein-coding genes to epigenetics involving ncRNAs. In 2018, Onoufriadis et al. [26] compared patients with keloids with patients without keloids. They sampled wounds on the buttocks of all participants and sampled the untreated wounds 6 weeks later. Two specimens from each participant were obtained for transcriptome analysis, and the results revealed differences in mRNA and miRNA expression between the 2 groups and identified the miRNA-mRNA

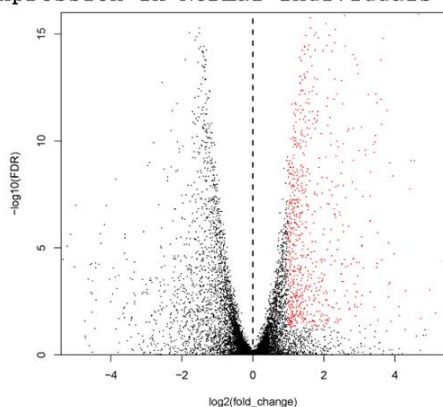
regulatory relationships. In the current study, the open-source GEO database to conduct secondary mining of ncRNAs and mRNAs through bioinformatics algorithms, aiming to identify suitable ncRNA molecules for further exploration of the epigenetic regulation mechanism of keloids.

In this study, the genes of keloid-prone patients were separated from those of control individuals and were analysed separately according to the wound healing process. Considering that the wound healing process is common to all populations, we excluded the intersecting set of genes between the keloid-prone patients and the control individuals, retaining only the genes differentially expressed in keloid-prone patients for further study. Because differential gene screening is generally based on counts, errors often occur. To reduce the false positive and false negative rates, the most recent version of DESeq2 software was used herein to obtain the required DEMRNAs and DELncRNAs in

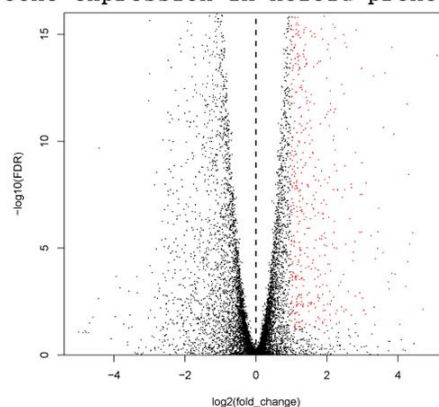
order to ensure the accuracy of the data analysis [12]. For typical prediction of miRNA target genes, more than 2 miRNA target gene prediction software programs are selected for prediction, and the intersection of the prediction results is then obtained to identify the most reliable miRNA target gene prediction results [26]. In this study, we used the StarBase database for analysis. The miRNA data in this database were verified by Ago2-CLIP-Seq [31] with a relatively high degree of reliability.

GO is an international standardized gene functional classification system that provides a set of dynamically updated controlled terms to comprehensively describe the attributes of genes and gene products in organisms. GO has 3 ontologies, which individually describe the MF, CC, and BP of genes. The basic unit of GO is the term, and each term corresponds to an attribute. Each gene is associated with 1 or more GO terms (aka GO functions) [13]. KEGG pathway analysis is a gene

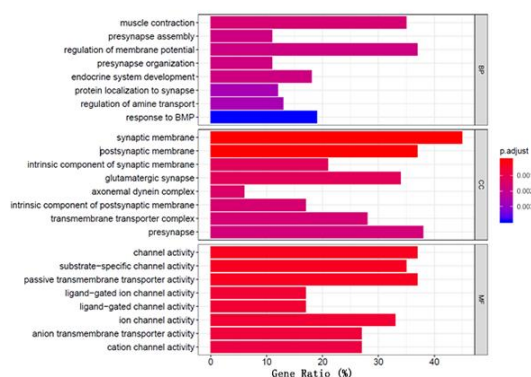
**A** Gene expression in Normal individuals



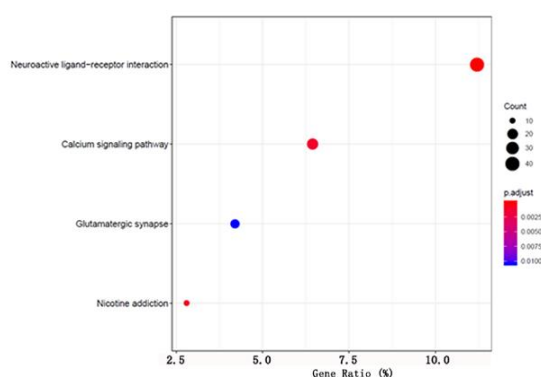
**B** Gene expression in keloid-prone individuals



**C** GO Enrichment Analysis



**D** KEGG Enrichment Analysis

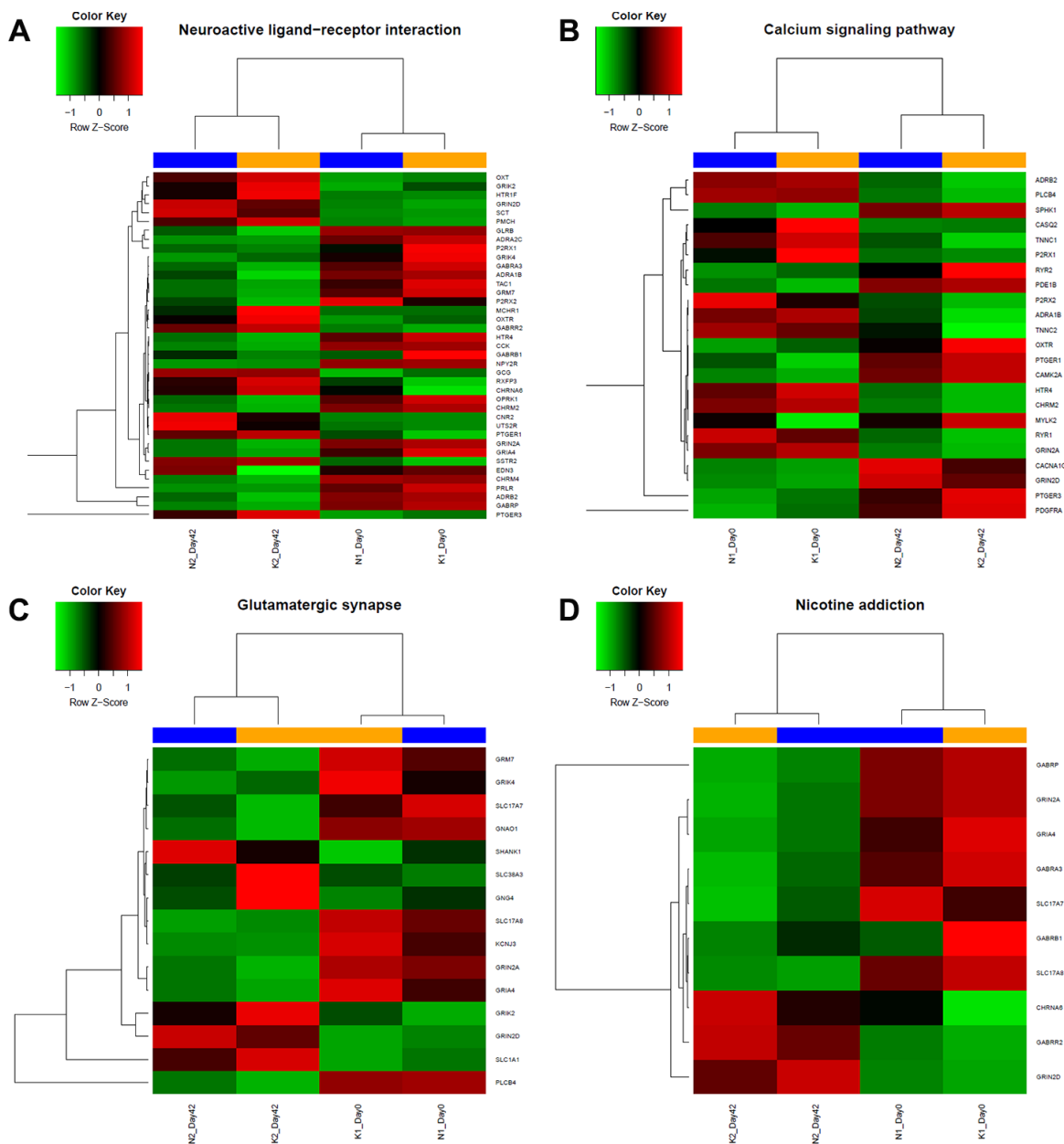


**Figure 1. Volcano plots of mRNA expression in the two groups; GO and KEGG enrichment analyses of specific genes expressed in keloid-prone individuals. (A)** mRNA expression in normal individuals. **(B)** mRNA expression in keloid-prone individuals. **(C)** GO enrichment analysis. **(D)** KEGG enrichment analysis. In A and B, the left side of the dashed line represents the uninjured state (day 0); the right side represents the injured state (day 42).

**Table 2. KEGG analysis of mRNAs specifically expressed in keloid-prone individuals 6 weeks after injury.**

mRNA	Term description	Ct.	Adjusted <i>p</i> -value
hsa04080	neuroactive ligand-receptor interaction	40	3.42×10 <sup>-6</sup>
hsa05033	nicotine addiction	10	0.000953
hsa04020	calcium signaling pathway	23	0.001322
hsa04724	glutamatergic synapse	15	0.010286

Ct.: number of genes.



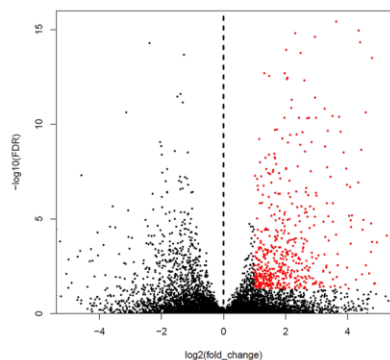
**Figure 2. Heatmap of specific gene expression in different groups from 4 different KEGG enrichment pathway categories.** According to results from KEGG analysis, expression level of genes of each group from (A) Neuroactive ligand-receptor related pathways, (B) Calcium signalling related pathway, (C) Glutamatergic synapse related pathway and (D) Nicotine addiction related pathway were shown. Red indicates upregulation; green indicates downregulation; blue indicates the normal group, including day 0 and day 42; yellow indicates keloid-prone individuals, including day 0 and day 42.

annotation database-based method for detecting pathways significantly enriched with differentially expressed genes. Therefore, a comprehensive database and complete pathway annotations are the keys to pathway analysis. Pathway analysis results reveal direct interactions between genes in the actual signalling pathways. Compared with GO analysis, pathway analysis allows researchers to study the biological effects of differentially expressed genes; moreover, it is complementary to GO analysis [30].

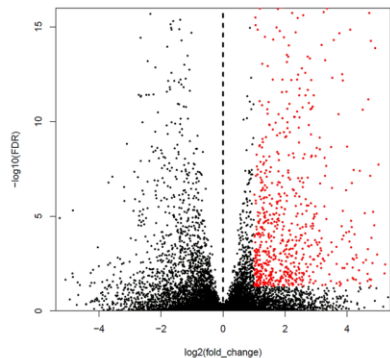
This study showed that during wound healing, DE mRNAs expressed only in keloid-prone individuals were mainly enriched in various ion channels in GO MF categories. KEGG pathway analysis further showed that these DE mRNAs were enriched mainly in the calcium signalling pathway. Liang et al. reported a similar finding in a microarray study focusing on DE lncRNAs and DE mRNAs between keloid tissue and normal skin tissue and found that the lncRNA CACNA1G-AS, an antisense RNA to CACNA1G, which encodes a calcium channel subtype, was upregulated in keloid tissue in independent pairs of samples [14]. Calcium channels

are associated with reactions to mechanical forces in different cell types, and elevated intracellular calcium levels were observed in fibroblasts and keratinocytes under application of hydraulic pressure [15]. Promotion of calcium channel expression could positively affect the migration of keloid fibroblast cells [16]. L-type calcium channel blockers, such as verapamil, have already been used in clinical treatment for keloids and have been proven to be safe and effective [17]. These results collectively suggest that dysregulation of a mechanical signal-related network, with mediation of calcium channel activation, could be pivotal in keloid development. KEGG pathway analysis also showed that the glutamatergic synapse pathway might also play an important role. Evidence has shown immunoreactivity of L-glutamate and its ionotropic receptor NMDAR in the human epidermis [18–20]. These effects may cooperatively regulate the calcium levels in related tissues and cells, as previously reported [21]. Moreover, our findings showed that neuroactive ligand-receptor interactions might also contribute. A 2018 bioinformatic analysis of the molecular mechanism of pathological scarring by Zhang et al. noted that genes expressed in

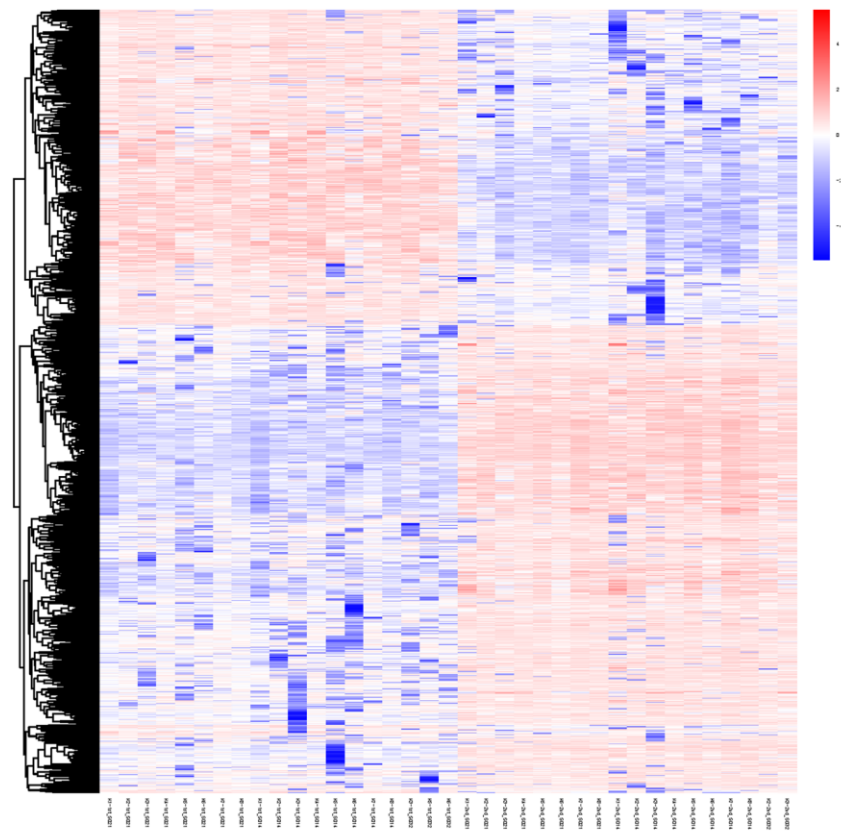
**A. Normal Individuals**



**B. Keloid-prone Individuals**



**C. Heatmap of lncRNA expression in Keloid-prone individuals**

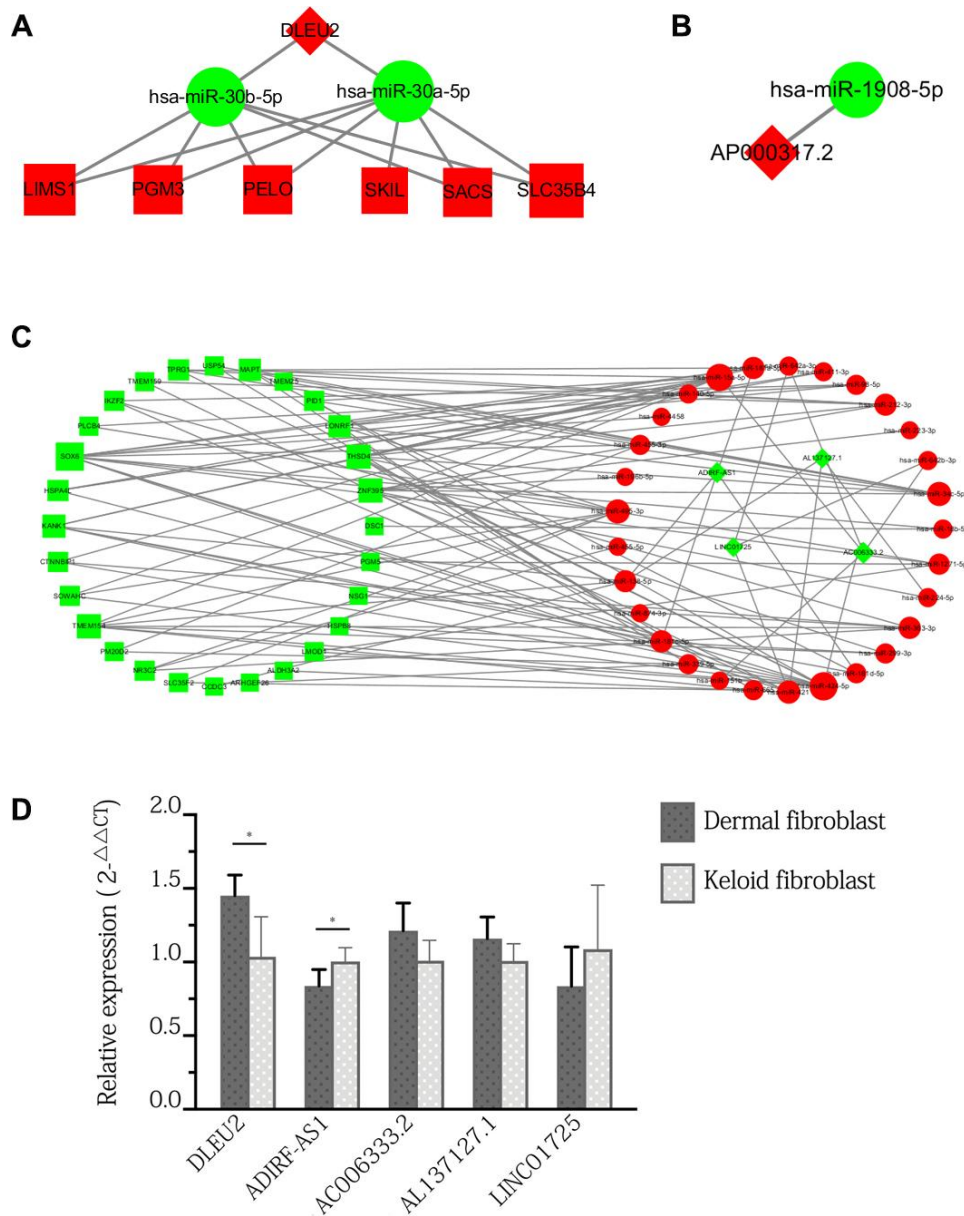


**Figure 3. Volcano plots and heatmap of lncRNA expression in keloid-prone individuals.** (A) lncRNA expression in normal individuals. (B) lncRNA expression in keloid-prone individual group. In A and B, the left side of the dashed line represents the uninjured state (day 0); the right side represents the injured state (day 42). (C) Heatmap of lncRNA expression before and after injury in Keloid-prone groups.

scar tissues were enriched in this channel [22]. The neuroactive ligand-receptor interaction pathway includes hundreds of genes involving many types of neuroreceptors, suggesting a broader, yet-to-be-discovered, complex neuroactive ligand-receptor mediated network; therefore, further investigations are needed. Moreover, GO analysis showed that the genes were enriched not only in the BP terms muscle contraction, presynapse assembly, and regulation of membrane potential but also CC terms related to different structural parts of synapses. Last, by KEGG

analysis, we also found DEMRNAs enriched in the nicotine addiction pathway, which has not been reported in previous studies and awaits experimental verification.

ncRNAs are RNAs that are not directly translated into proteins and were once thought to be non-functional components in gene expression and transcription [10]. As epigenetics and other research methods for genes and proteomic approaches have recently been developed, ncRNAs have been found to be associated with gene expression, not only regulating gene



**Figure 4. ceRNA sub-network of specific lncRNAs expressed only in keloid-prone individuals and qPCR verification of DElncRNAs in keloid and normal fibroblasts. (A) Sub-network of DLEU2. (B) Sub-network of AP000317.2. (C) Sub-network of downregulated lncRNAs. Green indicates downregulated molecules, red indicates upregulated molecules, diamonds indicate lncRNAs, circles indicate miRNAs, and squares indicate mRNAs. (D) qPCR verification of DElncRNAs in keloid and normal fibroblasts. \*p<0.05.**

**Table 3. Differentially expressed lncRNAs in keloid-prone individuals 6 weeks after injury.  $p < 0.05$  indicates a statistically significant difference.**

Gene Id	Gene_Name	KDay0	KDay42	log2FC	Adjusted-P	NDay0	NDay42	log2FC	Adjusted-P
ENSG00000231607	DLEU2	306.79	641.38	1.06	$1.33 \times 10^{-17}$	74.88	66.26	-0.18	1.00
ENSG00000272657	AP000317.2	75.61	183.43	1.28	$3.53 \times 10^{-17}$	66.42	119.35	0.85	$7.08 \times 10^{-3}$
ENSG00000272734	ADIRF-AS1	1350.45	552.62	-1.29	$2.46 \times 10^{-29}$	1251.46	646.56	-0.95	$1.24 \times 10^{-5}$
ENSG00000272686	AC006333.2	82.88	39.50	-1.07	$8.29 \times 10^{-9}$	70.48	41.97	-0.75	$1.46 \times 10^{-2}$
ENSG00000272084	AL137127.1	661.42	240.08	-1.46	$2.66 \times 10^{-18}$	407.22	223.42	-0.87	$7.60 \times 10^{-5}$
ENSG00000233008	LINC01725	257.15	121.20	-1.09	$1.87 \times 10^{-13}$	211.00	134.79	-0.65	$4.60 \times 10^{-2}$

transcription and post-transcriptional modifications and translation but also forming ceRNA regulatory networks to participate in mutual reciprocal regulation, which in turn affects the biological functions of cells, tissues, and organisms [20]. According to their length, ncRNAs can be divided into small non-coding RNAs (sncRNAs), lncRNAs, and circular RNAs (circRNAs). Wang et al. studied the expression and effects of lncRNA-H19 in keloid fibroblasts and found that H19 regulated the viability and apoptosis of fibroblasts through the miR-29a/COL1A1 axis [19]. However, the functions of most lncRNAs are still unclear. Therefore, we used the existing gene data for keloid fibroblasts to identify the ncRNA molecules differentially expressed during keloid formation and to more comprehensively explore the epigenetic regulatory mechanism in keloids.

We found that a total of 6 lncRNAs were involved in the network: ADIRF-AS1, DLEU2, AP000317.2, AC006333.2, AL137127.1, and LINC01725. Among these DElncRNAs, DLEU2 and AP000317.2 were upregulated, and the others were downregulated. The ceRNA sub-network in which these lncRNAs are involved is shown in Figure 4A–4C. The experimental design for acquisition of the original GEO data involved invasive surgery on uninjured skin of patients with keloid scars. Therefore, no matching data can be found in public databases. To verify the expression of the DElncRNAs, we used pathological specimens of keloid scars as a substitute to compare with normal skin. The results showed that only the expression of DLEU2 and ADIRF-AS1 in keloid fibroblasts showed significant difference and that the expression of other genes was not changed significantly. We cannot exclude the possibility that our results were affected by differences in the sequencing models or qPCR analysis. Although these lncRNAs have not been reported in previous keloid-related literature, molecules such as DLEU2 and ADIRF-AS1 have been detected in other tissues and diseases [23, 24]. DLEU2, a gene located on chromosome 13q14, has frequently been observed to be deleted or epigenetically suppressed in leukaemia and is considered a tumour suppressor [24, 25]. Wu et al.

found that the pattern of high lncRNA DLEU2 expression or low miR-30a-5p expression was an unfavourable prognostic factor for survival and tumour recurrence in patients with non-small cell lung cancer and showed that the lncRNA DLEU2 accelerates the development of cancer by sponging miR-30a-5p [24]. In this study, we predicted that DLEU2 acted on target genes, including development-related genes (such as LIMS3), apoptosis-related genes (such as SKIL), differentiation-related genes (such as PELO), and protein transporter-related genes (such as SLC35B4), through miR-30a-5p or miR-30b-5p, suggesting that DLEU2-associated ceRNAs may play an important regulatory role in keloid formation and wound healing. This possibility awaits verification by more basic and clinical experimental data.

Through data mining of the open-source GEO database, we acquired new regulatory information from previous sequencing and microarray data. However, our study has limitations. First, our study was based on secondary analysis of online sequencing data with limited amounts of original clinical samples. More clinical samples are needed to provide supportive evidence. The differential expression of lncRNAs and their targets in our ceRNA network also requires confirmation via qRT-PCR. Second, our hypotheses were inspired by similar findings in tumour studies, because keloids are thought—but have not been confirmed—to be tumour-like. Therefore, further well-designed experiments are needed to verify these hypotheses.

## CONCLUSION

This study used bioinformatics methods to perform in-depth data mining of the ncRNA data in GEO microarray datasets. We found that the calcium signalling pathway, glutamatergic synapse pathway, neuroactive ligand-receptor interaction pathway and nicotine addiction pathway may have potential regulatory effects that influence keloid development. We also identified lncRNAs and a ceRNA regulatory network that exist only in keloid-prone patients during the wound healing



process, possibly providing novel intervention targets and therapeutic options for patients with keloid scars.

## MATERIALS AND METHODS

### Acquisition of GEO sequencing data

The GSE113621 dataset, including sequencing data from the GSE113619 dataset and raw miRNA microarray data from the GSE113620 dataset, was acquired from the NCBI GEO database (<http://www.ncbi.nlm.nih.gov/geo>). This dataset included data from 8 keloid-prone patients (K1-K8) and 6 non-keloid-prone patients (N1-N6). The specific information for the patients is listed in Table 4 [26]. The data can be divided into 4 groups, where K<sub>1-8</sub>-1st refer to tissues from keloid-prone patients before injury, K<sub>1-8</sub>-2nd refer to tissues from keloid-prone patients 6 weeks after injury, N<sub>1-6</sub>-1st refer to tissues from non-keloid-prone patients before injury, and N<sub>1-6</sub>-2nd refer to tissues from non-keloid-prone patients 6 weeks after injury. The data were acquired on the Illumina HiSeq 2500 (*Homo sapiens*) and Illumina HiSeq 3000 (*Homo sapiens*) sequencing platforms, and the Affymetrix Multispecies miRNA-4 Array microarray data platform was used. The analysis process is shown in Figure 5A and 5B.

### Quality control and annotation of sequencing data

The raw RNA sequencing (RNA-Seq) data in FASTQ format were acquired from the GEO database and initially filtered using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to obtain the sequencing quality distribution, base content distribution, and proportion of repetitive sequencing fragments [27]. Because this portion of the data was basically normal, thus having no direct impact on the results, no QC results were provided. After quality control, Hisat2 software was used to compare the clean reads with the reference genome (hg38), and the count rate of each sample was compared [28]. According to the comparison results, the fragments per kilobase per million mapped reads (FPKM) value was used as the standard for quantitation of gene expression, and mRNAs and lncRNAs were quantitatively analysed based on the annotation information in the Genecode V25 database.

### Screening and statistical analysis of mRNAs differentially expressed between different groups

To more comprehensively study phenotype-related gene expression, the differentially expressed genes between an experimental group and the control group must be identified and correlated with a phenotype. For differential gene expression analysis, the most recent version of DESeq2 software was used, and the

thresholds for significant differential gene expression were set as follows:  $|\log_2\text{FoldChange (FC)}| > 1$  and  $\text{FDR} < 0.05$  [29]. As shown in Figure 5B, differentially expressed mRNAs and lncRNAs (DEmRNAs and DELncRNAs) in the K<sub>1-8</sub>-1st/K<sub>1-8</sub>-2nd and N<sub>1-6</sub>-1st/N<sub>1-6</sub>-2nd groups were screened first and designated as the keloid group and the non-keloid-prone (control) group, respectively. DEmRNAs and DELncRNAs that were also expressed in the control group were excluded from the keloid group, and genes that were expressed only in the keloid group were selected for subsequent analysis.

### Gene Ontology (GO)/Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses of DEmRNAs

The analysis method [30] previously described by our group was used to perform GO and KEGG pathway enrichment analysis. In brief, DEmRNAs were used as the analysis objects, and to obtain all GO terms corresponding to each DEmRNA, GO annotation analysis based on the GO database was performed for 3 ontologies: biological process (BP), molecular function (MF), and cellular component (CC). Fisher's exact test was used to calculate the significance level of each GO term. The test criterion  $\alpha$  was set to 0.05, and the significantly enriched GO terms were screened.

Additionally, KEGG pathway annotation was performed for the DEmRNAs in the database to obtain all pathway terms corresponding to each gene. The significance level of each pathway term was calculated using Fisher's exact test. The test criterion  $\alpha$  was set to 0.05, and significantly enriched pathway terms were screened. Heatmaps of pathway terms highly enriched with the DEmRNAs were generated using the R language.

### Secondary mining of miRNA microarray data

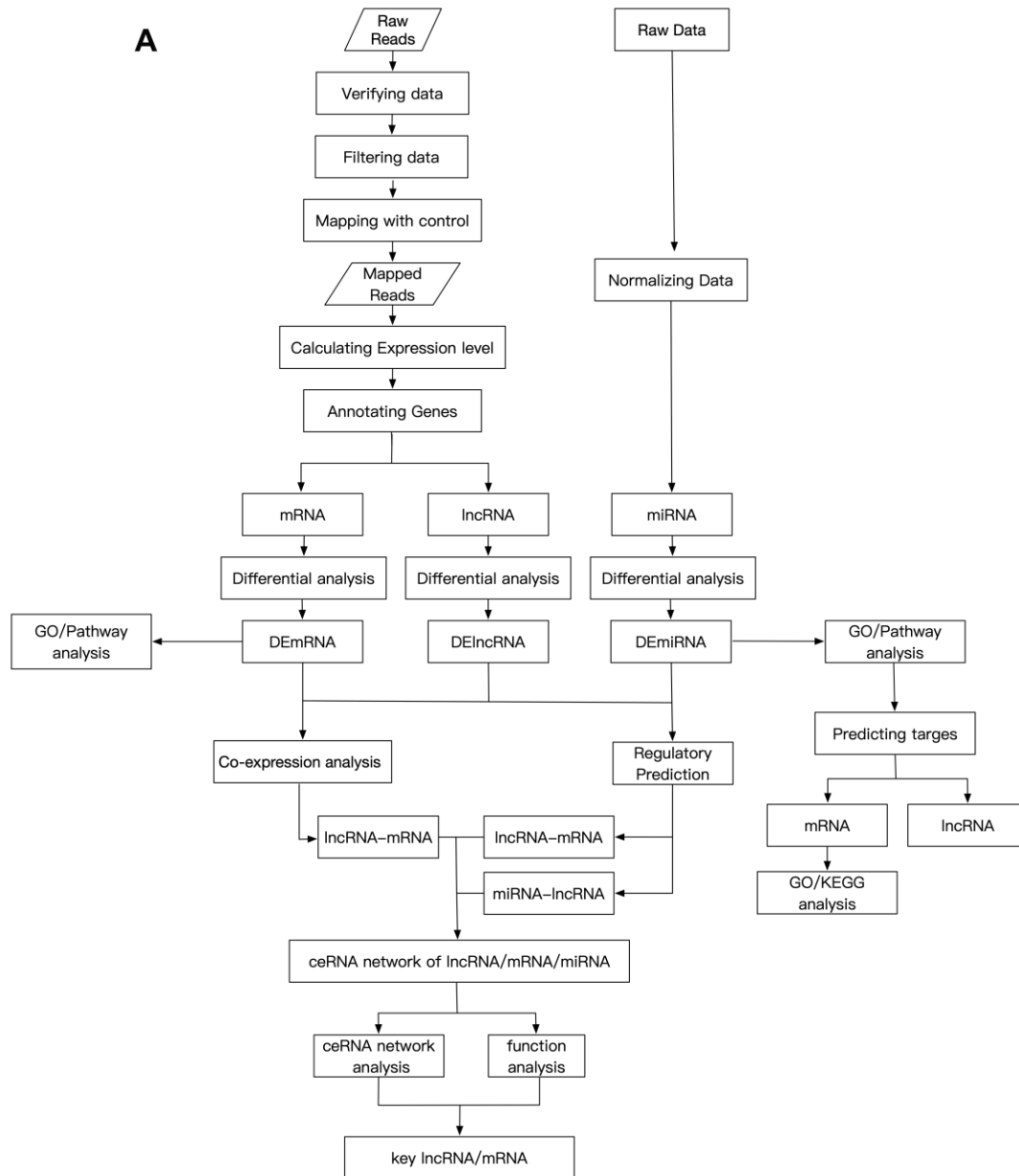
The raw data for miRNA expression analysis were obtained from a GEO dataset (GEO accession number GSE113621). In brief, the CEL file was pre-processed and normalized with the R packages affy (version 1.64.0) and simpleaffy (version 2.62.0). Differential expression analysis of miRNAs was performed with the R package limma (version 3.42.2). MiRNAs for which  $|\text{FC}| > 1.5$  and  $P < 0.5$  were considered significant DE miRNAs. MiRNA target prediction was performed with StarBase version 3.0 (<http://starbase.sysu.edu.cn>) [31, 32].

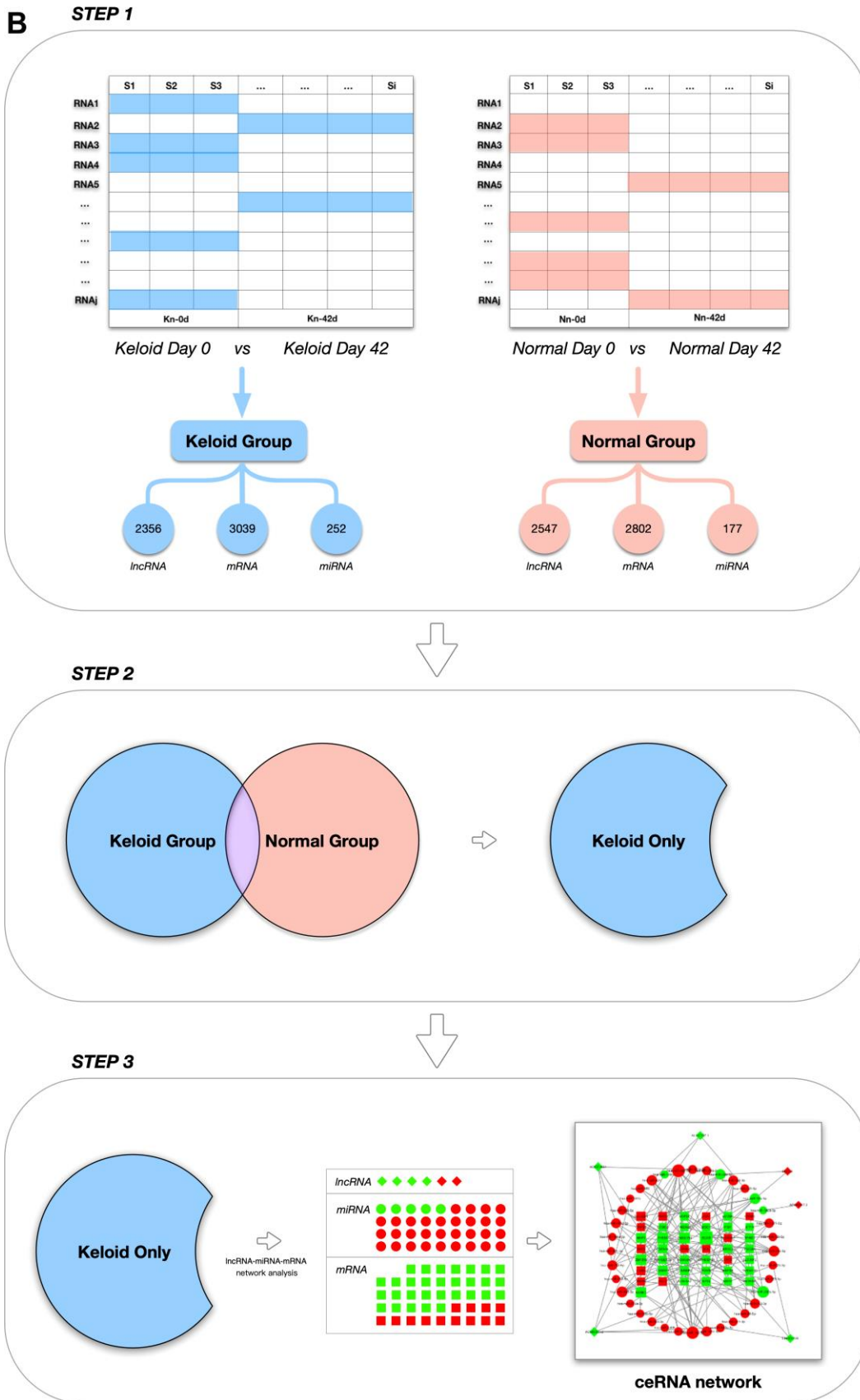
### Data acquisition and construction of the lncRNA-miRNA-mRNA ceRNA network

After the DELncRNA, DE miRNA, and DEmRNA analysis results and miRNA target gene prediction

**Table 4. Details of patients in GEO dataset SRP137071.**

Keloid-prone ID	Age (years)	Sex	Control ID	Age (years)	Sex
K-1	32	male	N-1	24	male
K-2	23	male	N-2	58	male
K-3	28	female	N-3	28	male
K-4	23	male	N-4	22	male
K-5	24	male	N-5	31	male
K-6	30	male	N-6	34	male
K-7	57	female			
K-8	25	male			





**Figure 5. Schematic design of the study. (A) Data processing and analysis procedures. (B) Grouping diagram.**

**Table 5. qPCR primer sequences of DElncRNAs specifically expressed in keloid-prone individuals.**

LncRNA	Forward (5'-3')	Reverse (5'-3')
DLEU2	TCC GAG AGT ATA GCG CCA CT	ACT GCC CTT TGC TCC AAG TA
ADIRF-AS1	GCC CAC TGA ATT CCC CTG AA	AAT CAG AGT GAC TGC CCC AC
AC006333.2	TCC AAC TTC GTA CTC TGG CC	GCT TCG CAA AGG TGT ACG TC
LINC01725	TCC AGC TCT TCT CCC CTG AA	GGG GGA ACT TAG AAA GGC CA
AP000317.2	AGT GGA TGG CAA GCT TCC TT	GGA ACC CTC GTC TTT GGG AA
AL137127.1	TCC CTT GAA TGC ACA GCC AT	CAA TGT GGT GCC CCA ACT TG

results were obtained, the lncRNA-miRNA-mRNA ceRNA network diagram was constructed [33, 34] by the following method. First, the Pearson correlation coefficient (PCC) was used to evaluate the correlations between the expression of DElncRNAs and DEMRNAs. PCC > 0.95 was used as the screening criterion, and the screened lncRNA-mRNA combinations were considered correlation pairs. Next, miRNA target genes were screened from the lncRNA-mRNA correlation pairs. For lncRNA-mRNA co-expression pairs that were simultaneously negatively correlated with an miRNA and were also target genes of the miRNA, the lncRNA-mRNA-miRNA axis was considered a competitive triad. Finally, Cytoscape (version 3.6.1) was used to visualize the ceRNA competitive triad in the network.

#### qPCR verification of DElncRNA expression in fibroblasts from keloid scars and normal skin

The expression levels of deleted in lymphocytic leukaemia 2 (DLEU2), AP000317.2, ADIRF-AS1, AC006333.2, AL137127.1, and LINC01725 were measured at different time points using real-time qPCR. In brief, keloid tissue and normal skin were collected from different patients (n=5). This study conformed to the guidelines established by the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University, and written informed consent was obtained from all included patients. Via the extraction method described in Han et al. [2], fibroblasts were extracted from both groups of tissues and were subcultured in an incubator. Total RNA was extracted using TRIzol (Thermo Fisher Scientific, USA). The RNA concentration and purity were measured using a NanoDrop 2000 (Thermo Fisher Scientific), and the final concentration was adjusted to 200 ng/μl. RNA (1 μg from each specimen) was reverse transcribed with a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). An appropriate amount of cDNA was amplified using FastStart Universal SYBR Green Master Mix (Roche, Switzerland) in a StepOnePlus Real-Time PCR instrument (Thermo Fisher Scientific). The specific procedure is shown in the system manual. Three replicates per specimen were analysed. The expression levels of all target mRNAs were

normalized to those of Gapdh. The information about the detected genes and the amplification primers are shown in Table 5. The specificity of the RT-PCR results was confirmed via routine agarose gel electrophoresis and melting curve analysis. The 2-ΔΔCt method was used to calculate relative gene expression levels.

#### AUTHOR CONTRIBUTIONS

YD and SX downloaded the GEO data and finish bioinformatic analysis. YZ analysed and interpreted these data. XL, BH and ZL were responsible for the statistical analyses. ZZ and YX designed the research. YD, SX and ZZ edited the manuscript. ZZ and YX revised the manuscript. All authors read and approved the final manuscript. Special thanks to Dr. Yuanqi Feng for consultant in bioinformation analysis.

#### CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

#### FUNDING

This research was supported by the Natural Science Foundation of Guangdong Province (2018A0303130217, 2020A151501159); the National Natural Science Foundation of China (81901024), Guangdong Basic and Applied Basic Research Foundation (2019A1515110462).

#### REFERENCES

- Han B, Fan J, Liu L, Tian J, Gan C, Yang Z, Jiao H, Zhang T, Liu Z, Zhang H. Faciocervical reconstruction using a large expanded forehead island flap grafted using a microsurgical technique for burned cicatricial contracture correction. *J Craniofac Surg.* 2018; 29:1848–50. <https://doi.org/10.1097/SCS.0000000000004647> PMID:29863568
- Han B, Fan J, Liu L, Tian J, Gan C, Yang Z, Jiao H, Zhang T, Liu Z, Zhang H. Adipose-derived mesenchymal stem

- cells treatments for fibroblasts of fibrotic scar via downregulating TGF- $\beta$ 1 and notch-1 expression enhanced by photobiomodulation therapy. *Lasers Med Sci.* 2019; 34:1–10.  
<https://doi.org/10.1007/s10103-018-2567-9>  
PMID:30367294
3. Zhang XQ, Wang ZL, Poon MW, Yang JH. Spatial-temporal transcriptional dynamics of long non-coding RNAs in human brain. *Hum Mol Genet.* 2017; 26:3202–11.  
<https://doi.org/10.1093/hmg/ddx203> PMID:28575308
  4. Chen R, Zhang Z, Xue Z, Wang L, Fu M, Lu Y, Bai L, Zhang P, Fan Z. Protein-protein interaction network of gene expression in the hydrocortisone-treated keloid. *Int J Dermatol.* 2015; 54:549–54.  
<https://doi.org/10.1111/ijd.12743> PMID:25660986
  5. Wang H, Quan L, Liang J, Shi J, Qiu T, Zhang Y, Wang Y, Hui Q, Zhang Y, Tao K. Gene expression profiling analysis of keloids with and without hydrocortisone treatment. *Exp Ther Med.* 2017; 14:5283–88.  
<https://doi.org/10.3892/etm.2017.5263>  
PMID:29285054
  6. Zhang L, Qin H, Wu Z, Chen W, Zhang G. Gene expression profiling analysis: the effect of hydrocortisone on keloid fibroblasts by bioinformatics. *J Dermatolog Treat.* 2019; 30:200–05.  
<https://doi.org/10.1080/09546634.2018.1484559>  
PMID:29863417
  7. Liu H, Chen D, Liu P, Xu S, Lin X, Zeng R. Secondary analysis of existing microarray data reveals potential gene drivers of cutaneous squamous cell carcinoma. *J Cell Physiol.* 2019; 234:15270–78.  
<https://doi.org/10.1002/jcp.28172> PMID:30697722
  8. Ogawa R. Keloid and hypertrophic scars are the result of chronic inflammation in the reticular dermis. *Int J Mol Sci.* 2017; 18:606.  
<https://doi.org/10.3390/ijms18030606>  
PMID:28287424
  9. Tsai CH, Ogawa R. Keloid research: current status and future directions. *Scars Burn Heal.* 2019; 5:2059513119868659.  
<https://doi.org/10.1177/2059513119868659>  
PMID:31452957
  10. Wang Z, Feng C, Song K, Qi Z, Huang W, Wang Y. lncRNA-H19/miR-29a axis affected the viability and apoptosis of keloid fibroblasts through acting upon COL1A1 signaling. *J Cell Biochem.* 2020; 121:4364–76.  
<https://doi.org/10.1002/jcb.29649> PMID:31930556
  11. He Z, Yang D, Fan X, Zhang M, Li Y, Gu X, Yang M. The roles and mechanisms of lncRNAs in liver fibrosis. *Int J Mol Sci.* 2020; 21:1482.  
<https://doi.org/10.3390/ijms21041482>  
PMID:32098245
  12. Wang Q, Zhang H, Liang Y, Jiang H, Tan S, Luo F, Yuan Z, Chen Y. A novel method to efficiently highlight nonlinearly expressed genes. *Front Genet.* 2020; 10:1410.  
<https://doi.org/10.3389/fgene.2019.01410>  
PMID:32082366
  13. Yi M, Horton JD, Cohen JC, Hobbs HH, Stephens RM. WholePathwayScope: a comprehensive pathway-based analysis tool for high-throughput data. *BMC Bioinformatics.* 2006; 7:30.  
<https://doi.org/10.1186/1471-2105-7-30>  
PMID:16423281
  14. Liang X, Ma L, Long X, Wang X. lncRNA expression profiles and validation in keloid and normal skin tissue. *Int J Oncol.* 2015; 47:1829–38.  
<https://doi.org/10.3892/ijo.2015.3177> PMID:26397149
  15. Goto M, Ikeyama K, Tsutsumi M, Denda S, Denda M. Calcium ion propagation in cultured keratinocytes and other cells in skin in response to hydraulic pressure stimulation. *J Cell Physiol.* 2010; 224:229–33.  
<https://doi.org/10.1002/jcp.22121> PMID:20432375
  16. Li Y, Liang X, Wang P, Long X, Wang X, Meng Z. Long non-coding RNA CACNA1G-AS1 promotes calcium channel protein expression and positively affects human keloid fibroblast migration. *Oncol Lett.* 2018; 16:891–97.  
<https://doi.org/10.3892/ol.2018.8717> PMID:29963160
  17. Che K, Lyu Q, Ma G. Comparative Efficacy and Safety of Common Therapies in Keloids and Hypertrophic Scars: A Systematic Review and Meta-analysis. *Aesthetic Plast Surg.* 2020. [Epub ahead of print].  
<https://doi.org/10.1007/s00266-020-01639-9>  
PMID:32040601
  18. Nordlind K, Johansson O, Lidén S, Hökfelt T. Glutamate- and aspartate-like immunoreactivities in human normal and inflamed skin. *Virchows Arch B Cell Pathol Incl Mol Pathol.* 1993; 64:75–82.  
<https://doi.org/10.1007/BF02915098>  
PMID:8106099
  19. Kasuya A, Tokura Y. Attempts to accelerate wound healing. *J Dermatol Sci.* 2014; 76:169–72.  
<https://doi.org/10.1016/j.jdermsci.2014.11.001>  
PMID:25468357
  20. Genever PG, Skerry TM. Non-neuronal glutamate signalling pathways. *Emerging Therapeutic Targets.* 2000; 4:333–45.  
<https://doi.org/10.1517/14728222.4.3.333>
  21. Skeberdis VA, Chevaleyre V, Lau CG, Goldberg JH, Pettit DL, Suadicani SO, Lin Y, Bennett MV, Yuste R, Castillo

- PE, Zukin RS. Protein kinase a regulates calcium permeability of NMDA receptors. *Nat Neurosci.* 2006; 9:501–10.  
<https://doi.org/10.1038/nn1664> PMID:[16531999](https://pubmed.ncbi.nlm.nih.gov/16531999/)
22. Zhang L, Qin H, Wu Z, Chen W, Zhang G. Identification of the potential targets for keloid and hypertrophic scar prevention. *J Dermatolog Treat.* 2018; 29:600–05.  
<https://doi.org/10.1080/09546634.2017.1421309> PMID:[29271272](https://pubmed.ncbi.nlm.nih.gov/29271272/)
23. Yu H, Xu Q, Liu F, Ye X, Wang J, Meng X. Identification and validation of long noncoding RNA biomarkers in human non-small-cell lung carcinomas. *J Thorac Oncol.* 2015; 10:645–54.  
<https://doi.org/10.1097/JTO.0000000000000470> PMID:[25590602](https://pubmed.ncbi.nlm.nih.gov/25590602/)
24. Wu W, Zhao Y, Gao E, Li Y, Guo X, Zhao T, He W, Zhang H. LncRNA DLEU2 accelerates the tumorigenesis and invasion of non-small cell lung cancer by sponging miR-30a-5p. *J Cell Mol Med.* 2020; 24:441–50.  
<https://doi.org/10.1111/jcmm.14749> PMID:[31721438](https://pubmed.ncbi.nlm.nih.gov/31721438/)
25. Zhou Y, Shi H, Du Y, Zhao G, Wang X, Li Q, Liu J, Ye L, Shen Z, Guo Y, Huang Y. LncRNA DLEU2 modulates cell proliferation and invasion of non-small cell lung cancer by regulating miR-30c-5p/SOX9 axis. *Aging (Albany NY).* 2019; 11:7386–401.  
<https://doi.org/10.18632/aging.102226> PMID:[31541993](https://pubmed.ncbi.nlm.nih.gov/31541993/)
26. Onoufriadis A, Hsu CK, Ainali C, Ung CY, Rashidghamat E, Yang HS, Huang HY, Niazi U, Tziotzios C, Yang JC, Nuamah R, Tang MJ, Saxena A, et al. Time series integrative analysis of RNA sequencing and MicroRNA expression data reveals key biologic wound healing pathways in keloid-prone individuals. *J Invest Dermatol.* 2018; 138:2690–93.  
<https://doi.org/10.1016/j.jid.2018.05.017> PMID:[29870686](https://pubmed.ncbi.nlm.nih.gov/29870686/)
27. Saddala MS, Lennikov A, Mukwaya A, Fan L, Hu Z, Huang H. Transcriptome-wide analysis of differentially expressed chemokine receptors, SNPs, and SSRs in the age-related macular degeneration. *Hum Genomics.* 2019; 13:15.  
<https://doi.org/10.1186/s40246-019-0199-1> PMID:[30894217](https://pubmed.ncbi.nlm.nih.gov/30894217/)
28. Kalbfleisch TS, Rice ES, DePriest MS Jr, Walenz BP, Hestand MS, Vermeesch JR, O Connell BL, Fiddes IT, Vershinina AO, Saremi NF, Petersen JL, Finno CJ, Bellone RR, et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol.* 2018; 1:197.  
<https://doi.org/10.1038/s42003-018-0199-z> PMID:[30456315](https://pubmed.ncbi.nlm.nih.gov/30456315/)
29. Koh MJ, Shin DH, Lee SJ, Hwang CS, Lee HJ, Kim A, Park WY, Lee JH, Choi KU, Kim JY, Lee CH, Sol MY. Gastric-type gene expression and phenotype in non-terminal respiratory unit type adenocarcinoma of the lung with invasive mucinous adenocarcinoma morphology. *Histopathology.* 2020; 76:898–905.  
<https://doi.org/10.1111/his.14077> PMID:[31985086](https://pubmed.ncbi.nlm.nih.gov/31985086/)
30. Liu JH, Tang Q, Liu XX, Qi J, Zeng RX, Zhu ZW, He B, Xu YB. Analysis of transcriptome sequencing of sciatic nerves in sprague-dawley rats of different ages. *Neural Regen Res.* 2018; 13:2182–90.  
<https://doi.org/10.4103/1673-5374.241469> PMID:[30323151](https://pubmed.ncbi.nlm.nih.gov/30323151/)
31. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014; 42:D92–7.  
<https://doi.org/10.1093/nar/gkt1248> PMID:[24297251](https://pubmed.ncbi.nlm.nih.gov/24297251/)
32. Namkung J. Statistical methods for identifying biomarkers from miRNA profiles of cancers. *Methods Mol Biol.* 2019; 1882:261–86.  
[https://doi.org/10.1007/978-1-4939-8879-2\\_24](https://doi.org/10.1007/978-1-4939-8879-2_24) PMID:[30378062](https://pubmed.ncbi.nlm.nih.gov/30378062/)
33. Guan Z. Changes in expression of serum chemokine CXCL13 and IL-6 after hip replacement, and the relationship with lower limb vein thrombus. *Exp Ther Med.* 2020; 19:2113–18.  
<https://doi.org/10.3892/etm.2019.8365> PMID:[32104273](https://pubmed.ncbi.nlm.nih.gov/32104273/)
34. Yao Y, Zhang T, Qi L, Zhou C, Wei J, Feng F, Liu R, Sun C. Integrated analysis of co-expression and ceRNA network identifies five lncRNAs as prognostic markers for breast cancer. *J Cell Mol Med.* 2019; 23:8410–19.  
<https://doi.org/10.1111/jcmm.14721> PMID:[31613058](https://pubmed.ncbi.nlm.nih.gov/31613058/)