A Study on the Expression of mRNAs and lncRNAs in Keloid Fibroblasts Based on GEO Microarray Data Mining

Zheng Liu  
Sun Yat-sen University First Affiliated Hospital

Zequan Li  
Sun Yat-sen University First Affiliated Hospital

Shuqia Xu  
Sun Yat-sen University First Affiliated Hospital

Bing Han  
Chinese Academy of Medical Sciences & Peking Union Medical College

Zhaowei Zhu  
The first affiliated hospital of sun yat-sen university  https://orcid.org/0000-0003-4469-5417

Yu Deng  
Sun Yat-sen University First Affiliated Hospital

Yujing Zhang  
Sun Yat-sen University First Affiliated Hospital

Xiangxia Liu  
liuxiangxia@gmail.com  Sun Yat-sen University First Affiliated Hospital

Yangbin Xu  
Sun Yat-sen University First Affiliated Hospital

Research Article

Keywords: keloid, scar, mRNA, lncRNA, bioinformatic, GEO database

DOI: https://doi.org/10.21203/rs.3.rs-633718/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Objective

The purpose of this study was to find the coding RNA (messenger RNA, mRNA) and long non-coding RNA (lncRNA) expressed in keloid through the analysis of GEO microarray chip of keloid fibroblasts.

Method:

GEO database GSE7890 database was downloaded with selection of keloids and normal scar group data. The data were analyzed by R language combined with online database. The log2FC > 1, p < 0.01 was chosen as screening criteria, and the differentially expressed mRNAs were screened for GO and KEGG function analysis.

Results

155 mRNA expression in the keloid group was significantly different from that in the normal group, including 31 groups with up-regulated mRNA expression and 124 groups with down-regulated mRNA expression. Meanwhile, eight lncRNAs were changed in the keloid group, including 3 up-regulated (Rp11-420a23.1, Rp11-522b15.3 and Rp11-706j10.1) and 5 down-regulated (LINC00511, LINC00327, Hoxb-as3, Rp11-385n17.1 and Rp3-428l16.2). qPCR analysis of DElncRNAs in keloid fibroblasts showed that the expression of all DElncRNAs except for RP11-385N17.1 was increased in the keloid group compared to the control group. Moreover, the differences in LINC00511 and RP11-706J10.1 were statistically significant.

Conclusion

The non-coding RNA information of GEO chip data can be deeply mined through bioinformatics, and the potential epigenomic mechanism affecting keloid formation can be found from the existing database.

1. Introduction:

Keloid is a type of pathological scar, and the pathogenesis involves abnormal hyperplasia of dermal fibrous tissue, which leads to excessive collagen deposition in the process of wound healing\(^1\). Although a large number of studies have investigated the mechanism of keloid occurrence and development and treatments for keloid, the outcomes of keloid treatment remain unsatisfactory\(^2\). Gene microarray is a powerful tool for studying gene expression profiles. However, due to limitations of the production process and algorithm of early gene microarrays, the required data cannot be obtained completely\(^3\). With the rise of bioinformatic analytical methods, many scholars have attempted to mine new transcriptome...
information from old microarray or sequencing data through improved algorithms, providing novel research ideas for clinical and scientific research.\(^3\)–\(^7\)

This study performed a secondary analysis on microarray data related to keloid formation downloaded from the Gene Expression Omnibus (GEO), an open source database of the National Center for Biotechnology Information (NCBI), using a combination of R language and bioinformatic analytical technology in the network database. The aim was to identify the coding RNA (messenger RNA, mRNA) and long non-coding RNA (lncRNA) expressed in keloid and to predict their expression status. First, this study verified the possibility of lncRNA retrieval by conducting a secondary analysis on microarray data with bioinformatic methods. Second, this study identified transcription information related to the pathophysiologic process of keloid from these data, providing a novel research method for better exploring the pathological mechanism of and treatments for keloid.

2. Materials And Methods:

2.1 Acquisition of GEO microarray results

The microarray dataset GSE7890 was obtained from the NCBI GEO database (http://www.ncbi.nlm.nih.gov/geo). This dataset mainly includes a keloid fibroblast group without prednisone treatment (n = 5), a normal scar fibroblast group without prednisone treatment (n = 5), a keloid fibroblast group after prednisone treatment (n = 5), and a normal scar fibroblast group after prednisone treatment (n = 4). Based on the needs of the experiment, we downloaded the raw data for the keloid and the normal scar groups without prednisone treatment (including the Affymetrix CEL file and the probe annotation file) and performed a further analysis. The platform of the microarray was GPL570 [Affymetrix Human Genome U133 Plus 2.0 (Affymetrix, USA)].

2.2 Processing and analysis of the keloid microarray data

After the data were successfully downloaded, we adopted the method reported by Yang et al.\(^3\). Namely, the custom chip description files (CDFs) for GENCODE genes were download from the BRAINARRAY website (http://brainarray.mbnl.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp), and the data were subjected to preprocessing (including background correction and normalization) and log2 transformation using Affymetrix Power Tools (APT). The gene-level probe sets were mapped to the human GENCODE annotation (version 28) using custom Perl scripts. Only the RNAs in the GENCODE database whose probe sets were annotated as "LNCRNA" and "PROTEIN-CODING RNA" were retained, while all other genes were filtered out. The validity of data normalization was confirmed by a straight line in the log2PM box plot. Differentially expressed genes were defined as genes with |logFC|>1 and an adjusted \(p<0.01\), including mRNA and lncRNA. Thus, the differentially expressed mRNA (DEmRNA) and lncRNA (DElncRNA) data were obtained.

2.3 Gene Ontology (GO)/Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis of
DEmRNA

DEmRNAs were subjected to GO and KEGG functional enrichment analyses using the DAVID database (Annotation, Visualization and Integrated Discovery, http://david.ncifcrf.gov/)\(^8\). The false discovery rate (FDR) was used as the inspection index. An FDR < 0.05 was considered statistically significant. GO and KEGG enrichment charts were created.

2.4 Quantitative polymerase chain reaction (qPCR) analysis of changes in DElncRNA expression in keloid and normal skin fibroblasts

Pathological tissues obtained from keloid patients and normal skin obtained from patients with a non-pathological scar constitution during initial surgery were randomly selected (n = 3). The entire process was in line with the ethical standards of the First Affiliated Hospital of Sun Yat-Sen Medical University, and informed consent was obtained from all patients. The detailed treatment method of the specimens is described in previous literature published by our research group\(^1\).

qPCR was performed to examine DElncRNA expression in the two groups of fibroblasts, and the method is described in detail in previous literature\(^1\). Briefly, the two groups of fibroblasts were digested and resuspended, and total RNA was extracted from these fibroblasts using TRIzol. The concentration and purity of the RNAs were examined using a NanoDrop 2000, and the final concentration of the RNAs was adjusted to 200 ng/µl. RNA samples (1 µg) were reverse transcribed using the RevertAid First Strand cDNA Synthesis Kit (Thermo, Massachusetts, USA). Appropriate amounts of cDNAs were then amplified in a fluorescent quantitative PCR instrument (StepOnePlus, Thermo, Massachusetts, USA) using FastStart Universal SYBR Green Master Mix (Roche, Basel, Switzerland). The specific steps were performed in accordance with the manufacturer’s instructions. Each sample was subjected to three experimental repeats. GAPDH expression was used to normalize the expression of mRNAs. Information for the genes tested and their primers is shown in Supplemental Table 1. The relative expression levels of the genes were calculated using the 2\(^{-ΔΔCt}\) method.

The data were analysed using SPSS V. 15 software. Comparisons between the groups were performed using the t test. The significance level was set to 0.05. A \(p\) value less than 0.05 indicated that a difference was statistically significant.

3. Results:

3.1 Analysis of DEmRNAs in the keloid group

The results of data standardization are shown in Fig. 1a. After transformation, the genes could be divided into protein coding RNA (mRNA) and non-coding RNA. Among the mRNAs of the keloid group, a total of
155 mRNAs displayed changes (see Fig. 1B, 1C). Specifically, 31 mRNAs showed upregulated expression. The top 10 upregulated genes were MEST, NTN4, IGFBP5, ESM1, SGMS2, JAG1, ERCC6, KCTD16, HECW2, and NGF. In contrast, expression was downregulated for 124 mRNAs. The top 10 downregulated genes were IFI27, TFPI2, MME, HOXA9, TEMN2, STMN2, S1PR1, HOXC10, HOXC6, and MMP3 (Supplemental Table 2).

3.2 GO and KEGG enrichment analyses of DEmRNAs

DEmRNAs were subjected to GO enrichment analysis (Fig. 2A). In the biological process subclass, DEmRNAs were found to be mainly enriched in the “single organization process” category (Fig. 2B). In the cell component subclass, DEmRNAs were mainly enriched in the “intracellular”, “intracellular parts”, “intracellular organelle”, “organelle”, “membrane-bound organelle”, and “intracellular membrane-bound organelle” categories (Fig. 2C). In the molecular function subclass, DEmRNAs were mainly enriched in the “binding” category (Fig. 2D).

The KEGG enrichment analysis showed that the DEmRNAs were mainly enriched in the p53 signalling pathway, sphingolipid metabolism, protein processing in the endoplasmic reticulum, the tumour necrosis factor (TNF) signalling pathway, the cell cycle, and neuroactive ligand-receptor interaction (Supplemental Table 3).

3.3 Analysis of DElncRNAs in the keloid group

Among the mRNAs of the keloid group, a total of eight lncRNAs exhibited changes (Fig. 3). Specifically, three lncRNAs were upregulated (RP11-420A23.1, RP11-522B15.3, and RP11-706J10.1), while five were downregulated (LINC00511, LINC00327, HOXB-AS3, RP11-385N17.1, and RP3-428L16.2) (Supplemental Table 4).

3.4 qPCR examination of DElncRNA expression in keloid fibroblasts

qPCR analysis of DElncRNAs in keloid fibroblasts showed that the expression of all DElncRNAs except for RP11-385N17.1 was increased in the keloid group compared to the control group. Moreover, the differences in LINC00511 and RP11-706J10.1 were statistically significant (Fig. 4).

4. Discussion:

Keloid is a clinical disease that is difficult to cure. Although keloid is a benign mass, it often grows beyond the boundary of the injury and invades the surrounding normal skin with the appearance of a crab claw, which seriously affects the appearance of the healed skin\(^9\). Keloids are mostly found in populations aged 10–30 years, especially in African, Hispanic, or Asian ethnic groups. Keloids often occur on the chest, earlobes, shoulders, and back\(^9\). Despite the existence of different treatment methods such as radiotherapy, hormone therapy, and surgical resection, keloid has a relatively high recurrence rate\(^10\). Many scholars believe that fibroblasts are one of the main participants in the occurrence and
development of keloid. After skin injury, a complex signal regulatory network is activated to control the proliferation, migration, and secretion of fibroblasts. Therefore, the biological behaviour of fibroblasts is a hot spot in the study of the mechanism of keloid formation.

Gene microarray and sequencing are important means to study gene expression profiles and transcription levels and are widely used in fields such as regenerative medicine and in the study of diseases and tumours. In recent years, the research focus on transcriptomes has gradually shifted from protein-coding genes to the epigenetic field involving non-coding RNAs (ncRNAs). ncRNAs are a class of RNAs that are not directly translated into polypeptides and were once considered ineffective components in the process of gene expression and transcription. With the development of epigenetics and gene and proteomic research methods in recent years, ncRNAs have gradually been discovered to be related to gene expression. ncRNAs not only regulate the process of gene transcription, post-transcriptional modification, and translation but also form a regulatory network of competing endogenous RNA (ceRNA), thereby affecting the biological functions of cells, tissues, and organisms. Based on the length of the RNA molecules, ncRNAs are divided into small non-coding RNAs (sncRNAs), IncRNAs, and circular RNAs (circRNAs). At present, the functions of most IncRNAs remain unclear. Wang et al. investigated the expression and effect of IncRNA-H19 in keloid fibroblasts. H19 was found to regulate the vitality and apoptosis of fibroblasts through the action of the miR-29a/COL1A1 axis. Therefore, we used existing gene information on keloid fibroblasts to identify new IncRNA molecules and enable further exploration of epigenetic regulatory mechanisms in keloid.

A large amount of sample data can be obtained through big network databases such as the GEO database or The Cancer Genome Atlas (TCGA). However, previous microarray data can often be used only for transcriptome or microRNA analysis due to limitations of the number and type of probes. When conducting analyses on non-coding RNAs, a new round of sample collection and sequencing analysis is often required. In addition, some microarrays, such as the Affymetrix Human Genome U133 Plus 2.0, allow the acquisition of certain ncRNA data. In these cases, the probe annotation must be updated, and secondary data mining is required. Based on the method proposed by Yang et al., we combined R language with network databases and selected the *.CEL file of GSE7890 (raw Affymetrix data) for secondary analysis. In the present study, we found that a total of 155 mRNAs exhibited changes in the keloid group compared to the control group. Among these mRNAs, the expression of 31 mRNAs was upregulated, while the expression of 124 mRNAs was downregulated. The GSE7890 dataset has also been analysed by other researchers. Wang et al. reported in 2017 that the expression of 67 genes was changed in the keloid group compared with the normal group (i.e., the expression of 15 genes was upregulated, while the expression of 52 genes was downregulated). However, Zhang et al. reported in 2019 that the expression of 832 genes was changed in the keloid group (including 269 upregulated genes and 563 downregulated gene). We believe that the discrepancy between our results and the results of previous studies is due to differences in algorithms and inclusion criteria. By analysing the KEGG enrichment data, we found significant gene enrichment in the p53, TNF, and cell cycle signalling pathways, which was different from the findings of the above articles. In addition, we identified eight
IncRNAs that were differentially expressed in keloid, including RP11-420A23.1, RP11-522B15.3, RP11-706J10.1, LINC00511, LINC00327, HOXB-AS3, RP11-385N17.1, and RP3-428L16.2. By comparing the expression of DElncRNAs between keloid and normal skin fibroblasts, we found that all DElncRNAs except for RP11-385N17.1 had increased expression in the keloid group compared to the control group. The differences in LINC00511 and RP11-706J10.1 expression were statistically significant. However, only the change in rp11-706j10.1 was consistent with the trend obtained from microarray mining. The expression trend and effect of rp11-706j10.1 still require confirmation with more specimens and animal experiments. Although differences were found between the qPCR and the microarray results, these differences may be related to the choice of threshold during microarray analysis and the small number of clinical samples. Moreover, the differences also indicated the possibility of further IncRNA mining through a combination of secondary analysis of the microarray data in open databases and bioinformatic methods to a certain extent.

Given the constantly updated database information, the old microarray data can be transformed into a powerful tool using bioinformatic methods. More information can be extensively mined from these old microarray data, which expands their use. The data included in the present study were obtained from few sources, and we can try to include the same type of data from multiple sources. These data can be analysed under the premise of satisfying normalization analysis, which would greatly reduce sample consumption, manpower, and time and money investments and quickly yield the required information. However, the present study still had limitations: First, microarrays have a relatively small data capacity, and the amount of data obtained was related to the type and number of probes designed. Therefore, only a small amount of data could be mined, which was not comparable with the amount of data obtained using other methods such as whole transcriptome sequencing. The results obtained can only be used as a reference. Second, due to the uncertainty of simple bioinformatic analysis, the results of secondary mining needed to be combined with more clinical samples and animal experiments, which would allow a better verification of the accuracy of the results and the clarification of the specific biological effects of the predicted molecules.

In summary, the present study achieved in-depth mining of IncRNA information from GEO microarray data using bioinformatic methods and identified the potential epigenetic regulatory mechanisms affecting keloid formation using existing databases.

**Declarations**

**Acknowledgement**: This research was supported by the Natural Science Foundation of Guangdong Province (2020A151501159) and China postdoctoral science foundation (2020M672990).

**Conflict of interest**: The authors have no conflicts of interest to declare.

**Contributions**
ZL and SX downloaded the data and analysed them. ZL and ZL interpreted the experimental data. YD and YZ helped draw the figures. BH was responsible for the statistical analyses. YX and XL designed the research. ZZ edited the manuscript. YX and XL revised the manuscript. All authors read and approved the final manuscript.

**Consent to Participate:**

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.

**Consent to Publish**

The Authors confirm that the work described has not been published before; that it is not under consideration for publication elsewhere; that its publication has been approved by all co-authors, if any; that its publication has been approved by the responsible authorities at the institution where the work is carried out.

**References**

1. Han B, Fan J, Liu L, Tian J, Gan C, Yang Z, Jiao H, Zhang T, Liu Z, Zhang H (2019 Feb) Adipose-derived mesenchymal stem cells treatments for fibroblasts of fibrotic scar via downregulating TGF-β1 and Notch-1 expression enhanced by photobiomodulation therapy. Lasers Med Sci 34(1):1–10
2. Han B, Fan J, Liu L, Tian J, Gan C, Yang Z, Jiao H, Zhang T, Liu Z, Zhang H (2018 Oct) Faciocervical Reconstruction Using a Large Expanded Forehead Island Flap Grafted Using a Microsurgical Technique for Burned Cicatrical Contracture Correction. J Craniofac Surg 29(7):1848–1850
3. Zhang XQ, Wang ZL, Poon MW, Yang JH (2017) Spatial-temporal transcriptional dynamics of long non-coding RNAs in human brain. Hum Mol Genet 26(16):3202–3211
4. Chen R, Zhang Z, Xue Z, Wang L, Fu M, Lu Y, Bai L, Zhang P, Fan Z (2015) Protein-protein interaction network of gene expression in the hydrocortisone-treated keloid. Int J Dermatol 54(5):549–554
5. Wang H, Quan L, Liang J, Shi J, Qiu T, Zhang Y, Wang Y, Hui Q, Zhang Y, Tao K (2017 Dec) Gene expression profiling analysis of keloids with and without hydrocortisone treatment. Exp Ther Med 14(6):5283–5288
6. Zhang L, Qin H, Wu Z, Chen W, Zhang G (2019) Gene expression profiling analysis: the effect of hydrocortisone on keloid fibroblasts by bioinformatics. J Dermatolog Treat 30(2):200–205
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 Jan 29:10.1002/jcp.28172
8. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms
to better extract biology from large gene lists. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W169-75

9. 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, de Rinaldis E, McGrath JA (2018 Dec) Time Series Integrative Analysis of RNA Sequencing and MicroRNA Expression Data Reveals Key Biologic Wound Healing Pathways in Keloid-Prone Individuals. J Invest Dermatol 138(12):2690–2693

10. Ogawa R Keloid and Hypertrophic Scars Are the Result of Chronic Inflammation in the Reticular Dermis. Int J Mol Sci. 2017 Mar 10;18(3):606

11. Tsai CH, Ogawa R (2019 Aug) Keloid research: current status and future directions. Scars Burn Heal 19:5:2059513119868659

12. Liu JH, Tang Q, Liu XX, Qi J, Zeng RX, Zhu ZW, He B, Xu YB (2018 Dec) Analysis of transcriptome sequencing of sciatic nerves in Sprague-Dawley rats of different ages. Neural Regen Res 13(12):2182–2190

13. Wang Z, Feng C, Song K, Qi Z, Huang W, Wang Y. IncRNA-H19/miR-29a axis affected the viability and apoptosis of keloid fibroblasts through acting upon COL1A1 signaling. J Cell Biochem 2020 Jan 13:10.1002/jcb.29649

14. He Z, Yang D, Fan X, Zhang M, Li Y, Gu X, Yang M. The Roles and Mechanisms of IncRNAs in Liver Fibrosis. Int J Mol Sci. 2020 Feb 21;21(4):E1482

Figures
Figure 1

Charts of the normalization analysis of the raw data for the keloid and control groups and a volcano plot and heat map of differentially expressed mRNAs. A) Normalization chart of the raw data; B) Volcano plot of the differentially expressed mRNAs between the two groups; C) Heat map of the differentially expressed mRNAs between the two groups. In the grouping bar, blue-green represents the keloid group, and pink represents the normal control group. In the heat map, blue represents genes with downregulated
expression, while red represents genes with upregulated expression. Darker colour corresponds to a greater difference in gene expression.

Figure 2

GO analysis of the differentially expressed mRNAs between the keloid and control groups and a bubble chart of the top 20 gene-enriched categories in various subclasses. A) GO analysis histogram of the differentially expressed mRNAs, where blue represents the biological process subclass, orange represents the molecular function subclass, and green represents the cellular component subclass; B) GO enrichment map of the biological process subclass; C) GO enrichment map of the cellular component subclass; D) GO enrichment map of the molecular function subclass. In B-D, bubble size represents the number of genes in each subclass. A larger bubble area corresponds to a greater number of genes. The colours represent the numerical magnitude of the FDR value.
Figure 3

Heat map of the differentially expressed lncRNAs between the keloid group and the control group. In the grouping bar, blue-green represents the keloid group, and pink represents the normal control group. In the heat map, blue represents genes with downregulated expression, while red represents genes with upregulated expression. Darker colour corresponds to a greater difference in gene expression.
Figure 4

qPCR examination of DElncRNA expression in the keloid group and the control group. In the figure, keloid indicates the keloid group, and dermal represents the control group. * indicates a statistically significant difference between the two groups (p<0.05).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table.docx