

# Integrated miRNA-mRNA expression profiles revealing key molecules in ovarian cancer based on bioinformatics analysis

**Chao Li**

Foshan Women and Children Hospital Affiliated to Southern Medical University

**Xiaodan Zhu**

Foshan Women and Children Hospital Affiliated to Southern Medical University

**Linghua Zhang**

Foshan Women and Children Hospital Affiliated to Southern Medical University

**Miaoling Ou**

Foshan Women and Children Hospital Affiliated to Southern Medical University

**Zhantong Hong** (✉ [fshongzt@hotmail.com](mailto:fshongzt@hotmail.com))

Foshan Women and Children Hospital Affiliated to Southern Medical University <https://orcid.org/0000-0003-4829-5867>

---

## Research

**Keywords:** Ovarian cancer, miRNA-mRNA, Bioinformatics analysis, network analysis, GEO

**Posted Date:** March 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-17869/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BioMed Research International on October 25th, 2021. See the published version at <https://doi.org/10.1155/2021/6673655>.

# Abstract

**Background:** Ovarian cancer was one of the leading causes of death in gynecological malignancies, of which molecular mechanism hadn't been elucidated clearly yet. Our research aimed to reveal the potential key molecular and biological processes of ovarian cancer by means of bioinformatics.

**Methods:** The microarray sets of miRNA and mRNA expression profiles were downloaded from the GEO database. The target prediction was performed on the differentially expressed miRNAs identified and the overlapped differentially expressed genes (DEGs) were obtained combined with miRNA and mRNA datasets. The regulatory network of miRNA-gene was further constructed by cytoscape software. The overlapped DEGs in the network were analyzed to explore the biological processes involved by enrichment analysis. The molecular protein-protein interaction (PPI) network was used to identify key genes among the DEGs.

**Results:** A total of 167 overlapped DEGs were identified. The miRNA-gene network analysis found that miR-29c-3p, miR-1271-5p, and miR-133b, existed the most extensive targeting relationship with overlapped DEGs, being three key miRNAs of the regulatory network, and played the role of tumor suppressor. The GO enrichment showed that the overlapped DEGs were mainly involved in process named extracellular related organization, embryonic organ development, postsynaptic specialization, collagen trimer and DNA-binding transcription activator et al. The KEGG pathway analysis showed that these DEGs were involved in protein digestion and absorption and relaxin signaling pathway. The PPI network identified 10 key genes, playing the role in promoting tumor.

**Conclusion:** The methodology used and identification of key molecules in our study contributed to understanding the pathogenesis of ovarian cancer and providing new candidate biomarkers for early screening of ovarian cancer.

## Background

Ovarian cancer was one of the leading causes of death in gynecological malignancies. The latest figures showed 295,414 new cases worldwide in 2018(1). The overall 5-year survival rate of ovarian cancer was less than 45%, mainly because the distant metastasis had been occurred when diagnosed. The biomarker such as CA125 currently used in clinical laboratories lacked enough specificity, and ultrasound examination was particularly easy to miss early cases. Therefore, the re-excavation of new biomarkers of ovarian cancer and the re-identification of key molecules of the disease were the key problems that need to be addressed in prevention and control level.

In recent years, the extensive application of expression profiles had accumulated a large amount of omics data, which depended on in-depth interpretation. Retrieving relevant researches in the past three years, some articles on expression profiles combined with bioinformatics to discover key genes of ovarian cancer had been published(2-5). However, most of the research groups selected the same type of microarray profiles, probably causing low accuracy and false positive results. The miRNAs belonged to

non-coding RNAs that bound to complementary sequences in the mRNA by the means of base pairing, which lead to mRNA silencing and regulated downstream gene expression negatively(6). Many studies had found that miRNA disorders could occur in almost all types of tumors, affecting target gene expression(7, 8). Therefore, our study adopted integrated miRNA and mRNA microarray expression profiles for joint analysis. By means of bioinformatics, we constructed regulatory networks to identify key molecules and biological processes of ovarian cancer, so as to provide scientific and accurate theoretical basis for elucidating the mechanism of ovarian cancer onset.

## Materials And Methods

### Data sources

GEO was a public functional genomics database. We searched the microarray expression profiles of ovarian cancer, limiting that tumor tissue compared with normal tissue for the experiment type. GSE119055 and GSE66957 were the two eligible profile datasets. The former was a miRNA profile, which included 3 normal tissues and 6 tumor tissues. The latter was an mRNA profile that included 12 normal tissues and 57 tumor tissues.

### Data processing

#### Differential expression

For the mRNA dataset, the probe ID was converted to the corresponding gene name according to its platform annotation file. The “limma” package of R language was used to analyze the DEGs(9), of which  $\text{adj.P.Val} < 0.05$  and the absolute value of  $\log_2\text{FC} > 1.5$  were defined as the statistically significant expression. The R package “org.Hs.eg.db” was used to convert the gene name to the corresponding gene ID(10), so as to perform the subsequent enrichment analysis. The miRNA dataset was processed by the same method and standard.

#### MiRNA target prediction

The Funrich software 3.1.3 was used to predict the downstream targets of the differentially expressed miRNAs that identified. The predicted gene list was intersected with the differentially expressed genes identified from the mRNA dataset to obtain the overlapped DEGs for subsequent regulatory networks, identification of key genes and functional enrichment analysis and so on.

#### Regulatory network

According to the principle that miRNAs regulated target genes negatively, the combinations of matching miRNA-gene were screened to construct regulatory networks. The construction was visualized with cytoscape 3.7.1 software.

#### Key genes

The string database was used to predict the interaction network between proteins that encoded by the DEGs. Then the cytohubba module in cytoscape software was used to identify key genes, of which source data was from the network file that obtained from the string database. The MCC algorithm of the module was selected to obtain the top key genes.

### **Function enrichment**

The R package “clusterProfiler” was used to perform the GO and KEGG enrichment annotation of the overlapped DEGs(10). GO annotation was divided into three subcategories named molecular function (MF), biological process (BP) and cellular components (CC). KEGG was a comprehensive database that integrated genomic, chemical knowledge and system functional information, which was widely used for enrichment annotation of gene pathways. The pvaluecutoff = 0.05 of the R package’s parameter was defined to filter valuable entries.

## **Results**

### **Differential expression**

A total of 31 differentially expressed miRNAs were identified by the GSE119055 dataset, among which the number of 28 were down-regulated and 3 were up-regulated. A total of 3844 differentially expressed genes were identified by the GSE66957 dataset, among which the number of 1995 were down-regulated and 1849 were up-regulated. The differentially expressed molecules were shown as the figure1.

### **Overlapped DEGs and regulatory networks**

According to the DEGs of mRNA dataset, combined with the target genes of differentially expressed miRNAs predicted by Funrich software, the 167 overlapped DEGs’ network files were obtained (attachment files 1 and 2 shown). The network files were imported into the cytoscape software for visual analysis. The result was shown in figure2A. It was found that miR-29c-3p, miR-1271-5p, and miR-133b were in the hub core of network regulation, of which number of target genes were the most.

### **Identification of key genes**

The results of identification of the key genes in the overlapped DEGs by the string database and cytohubba module were shown in table 1, which were all over-expressing differential genes. The top 10 key genes were screened according to the latest MCC algorithm, namely COL1A1, COL3A1, COL4A1, COL10A1, COL9A1, COL4A5, SST, ADRA2C, ADCY6 and TGFBI. Notably, COL1A1, COL3A1 and COL4A1 had the highest scores and were at the key core of the network (as shown in figure 2B).

Table 1 Function of the top 10 key genes at the hub core of the PPI network

No	Gene	Full name	Gene functions
1	COL1A1	collagen type I alpha 1 chain	Encoding pro-alpha1 chains of type I collagen. Being present in most connective tissues. May be related to lung cancer and esophageal cancer.
2	COL3A1	collagen type III alpha 1 chain	Encoding the pro-alpha1 chains of type III collagen. Genetic mutations are related to aneurysms.
3	COL4A1	collagen type IV alpha 1 chain	Encoding a type IV collagen alpha protein. Abnormal gene expression is associated with brain, kidney and muscle defects.
4	COL10A1	collagen type X alpha 1 chain	Encoding the alpha chain of type X collagen. Abnormal expression may be associated with lung cancer.
5	COL9A1	collagen type IX alpha 1 chain	Encoding one of the three alpha chains of type IX collagen. Existing two transcript variants.
6	COL4A5	collagen type IV alpha 5 chain	Encoding one of the six subunits of type IV collagen. Genetic variation is associated with hereditary nephritis.
7	SST	somatostatin	Belonged to the hormone somatostatin. Affecting the proliferation of normal cells and tumorigenic cells.
8	ADRA2C	adrenoceptor alpha 2C	Members of the G protein-coupled receptor superfamily. It is mainly responsible for mediating neurotransmitter transmission.
9	ADCY6	adenylate cyclase 6	Encoding a member of the adenylyl cyclase family of proteins, which are required for the synthesis of cyclic AMP <sup>2</sup>
10	TGFBI	transforming growth factor beta induced	Encoding an RGD-containing protein. Inhibiting cell adhesion and promotes cell proliferation and migration.

Note: The annotations in this table referred to the clear conclusions in NCBI (<https://www.ncbi.nlm.nih.gov/gene>), and relevant scattered reports weren't included in a single literature.

### Function enrichment annotation

The results of GO and KEGG enrichment analysis using the “clusterProfiler” package were shown as figure3. The GO enrichment showed that the overlapped DEGs were mainly involved in process such as extracellular related organization, embryonic organ development, postsynaptic specialization, collagen trimer and DNA-binding transcription activator et al. The KEGG pathway analysis showed that these DEGs were involved in protein digestion and absorption and relaxin signaling pathway.

## Discussion

Ovarian cancer was one of the common gynecological malignancies. However, the molecular mechanism of related pathogenicity hadn't been fully elucidated. Need to note that BRCA was one of the key genes for ovarian cancer that had been clearly identified. The BRCA mutation frequency of ovarian cancer ranged from 3% to 27%. The BRCA gene test provided precise advice for the prevention, diagnosis and treatment of ovarian cancer. But we still need to identify new key molecules for joint screening for most of the remaining cases.

Based on current researches(8, 11, 12), in addition to genes, the dysregulated expression of non-coding RNAs such as miRNA can also widely mediate various types of malignant tumors. In order to improve the prediction accuracy, our study identified overlapped DEGs based on integrated miRNA and mRNA expression profiles of ovarian cancer. Based on the principle of complementary binding of miRNAs to target mRNAs, negatively regulating genes, we constructed the miRNA-gene regulatory network to identify three key miRNAs as tumor suppressor: miR-29c-3p, miR-1271-5p, and miR-133b. Meanwhile, the 10 key genes were predicted and screened by visualizing the overlapped DEGs in the network with cytoscape software, namely COL1A1, COL3A1, COL4A1, COL10A1, COL9A1, COL4A5, SST, ADRA2C, ADCY6 and TGFBI. We enriched the biological processes involved in overlapped DEGs through R package. It was found that 11.3% (18/160) of the DEGs had the function of DNA transcription activation (as figure3A). The abnormal expression of differential miRNAs and genes were likely to mediate the occurrence and development of ovarian cancer. For the three key miRNAs discovered as tumor suppressor, we searched the experimental literature published in recent years. Hu ZH et al(13) confirmed that overexpression of miR-29c-3p inhibited autophagy via downregulating FOXP1/ATG14 pathway by experiments such as nude mice transplantation models in vitro, suggesting miR-29c-3p as a novel target in ovarian cancer. Lin W et al(14) revealed that miR-1271-5p expression was significantly reduced in ovarian cancer group by qRT-PCR assay and Du HY et al<sup>(15)</sup> performed the overexpression and knockout experiments using miR-1271-5p mimics and inhibitors and demonstrated miR-1271-5p was as a tumor suppressor in breast cancer proliferation and progression via targeting SPIN1 by Dual-luciferase activity assay. Liu XQ et al(16) demonstrated miR-133b inhibited proliferation and invasion of ovarian cancer cells through Akt and Erk1/2 inactivation by targeting epidermal growth factor receptor. These were in line with the key miRNA expression and target prediction results revealed in our study.

## Conclusions

In summary, it was reliable that the study was designed to reveal the key molecules of ovarian cancer by the integrated miRNA-mRNA expression profiles based on bioinformatics analysis. Of course, these key miRNAs and 10 top genes need more in-depth experimental verification in vitro. Nevertheless, the bioinformatics prediction still provided a good method, narrowing the scope of in vitro experiments and saving valuable resources. In the future, we believed that global researchers would be able to reveal key molecules of many complex and diverse tumors at once by a tumor big data strategy dependence on computational biology.

# Declarations

## Author' contributions

The research idea is derived from C Li and ZT Hong, and Chao Li designs the experiments. XD Zhu and LH Zhang analyze the data. C Li write the paper. ML Ou help to revise it. All authors have read and approved the final manuscript.

## Funding

Not applicable.

## Availability of data and materials

The datasets used and/or analyzed during the current study is available from the corresponding author on reasonable request.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

All authors have sufficiently contributed to manuscript preparation, and approve the submission of this manuscript. We thank the data provided by the authors of studies and GEO database.

## Author details

<sup>1</sup> Foshan Fetal Medicine Research Institute, Foshan Women and Children Hospital Affiliated to Southern Medical University, Foshan, Guangdong, 528000, China; <sup>2</sup> Department of Clinical Laboratory, Foshan Women and Children Hospital Affiliated to Southern Medical University, Foshan, Guangdong, 528000, China.

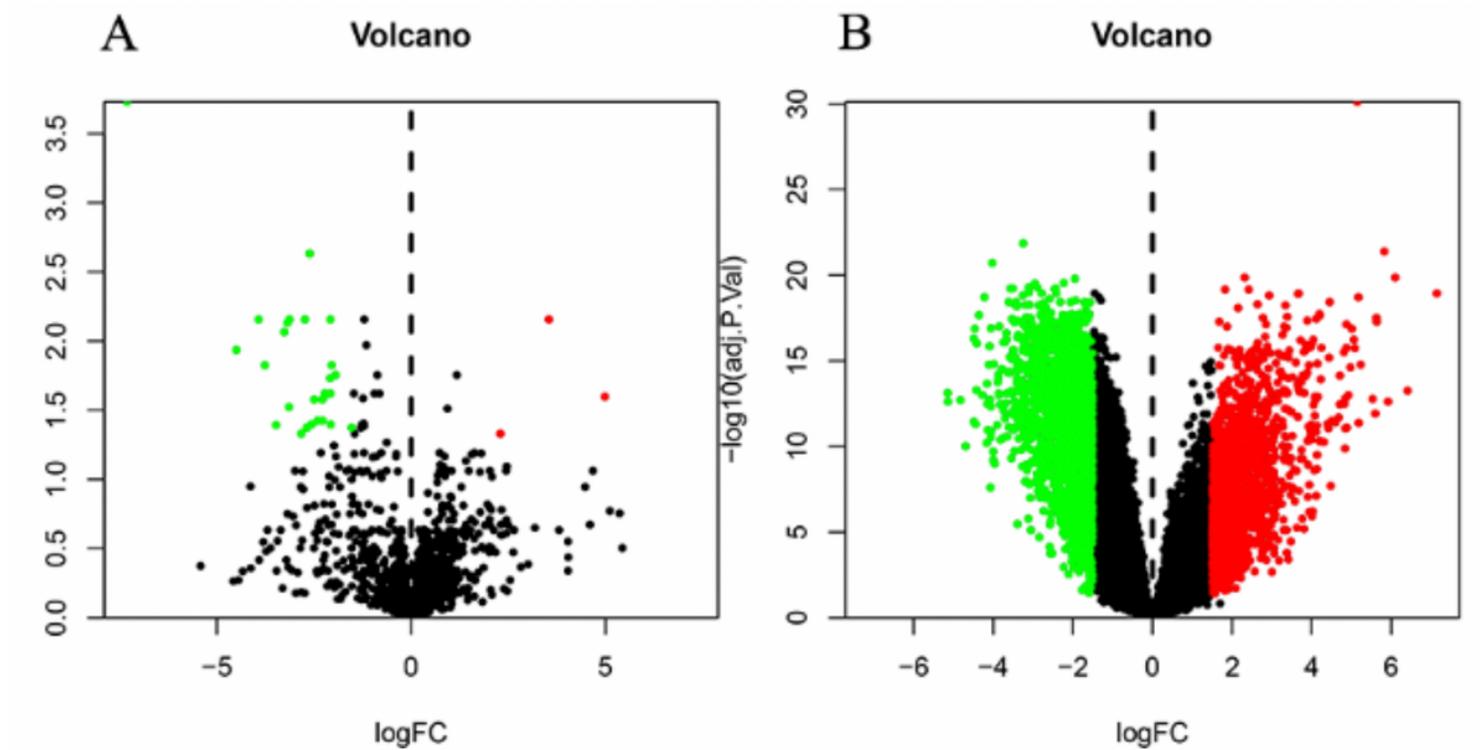
# References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a

- cancer journal for clinicians. 2018;68(6):394-424. DOI: 10.3322/caac.21492
2. Lu B-W, Qian J, Wang J, Qian J-R, Lu Y-Y, Su L-J, et al. Expression of miR-455-5p in epithelial ovarian cancer and functional analysis of its target genes. *Journal of Practical Oncology* 2019, 33 (2): 115-121.
  3. Guo S-M, Li H-M. Using bioinformatics to discover genes HADH and SRC as potential therapeutic target genes for ovarian cancer. *Electronic Journal of Clinical Medical Literature*. 2017.
  4. Yang X, Zhu S, Li L, Zhang L, Xian S, Wang Y, et al. Identification of differentially expressed genes and signaling pathways in ovarian cancer by integrated bioinformatics analysis. *Onco Targets Ther*. 2018;11:1457-74. DOI: 10.2147/OTT.S152238
  5. Feng H, Gu Z-Y, Li Q, Liu Q-H, Yang X-Y, Zhang J-J. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *Journal of ovarian research*. 2019;12(1):35. DOI: 10.1186/s13048-019-0508-2
  6. Lindsay MA, Griffiths-Jones S, Dalmay T. Mechanism of miRNA-mediated repression of mRNA translation. *Essays in biochemistry*. 2013;54:29-38. DOI: 10.1042/bse0540029
  7. Lujambio A, Lowe SW. The microcosmos of cancer. *Nature*. 2012;482(7385):347-55. DOI: 10.1038/nature10888
  8. Rupaimoole R, Calin GA, Lopez-Berestein G, Sood AK. miRNA deregulation in cancer cells and the tumor microenvironment. *Cancer discovery*. 2016;6(3):235-46. DOI: 10.1158/2159-8290.CD-15-0893
  9. Smyth GK, Ritchie M, Thorne N, Wettenhall J. LIMMA: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health. 2005.
  10. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*. 2012;16(5):284-7. DOI: 10.1089/omi.2011.0118
  11. Reddy KB. MicroRNA (miRNA) in cancer. *Cancer cell international*. 2015;15(1):38.
  12. Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics*. 2017;34(3):398-406. DOI: 10.1093/bioinformatics/btx622
  13. Hu Z, Cai M, Zhang Y, Tao L, Guo R. miR-29c-3p inhibits autophagy and cisplatin resistance in ovarian cancer by regulating FDX1/ATG14 pathway. *Cell Cycle*. 2020;19(2):193-206. DOI: 10.1080/15384101.2019.1704537
  14. Wang L, Zhu M-J, Ren A-M, Wu H-F, Han W-M, Tan R-Y, et al. A ten-microRNA signature identified from a genome-wide microRNA expression profiling in human epithelial ovarian cancer. *PLoS One*. 2014;9(5). DOI: 10.1371/journal.pone.0096472
  15. Du H, Liu B. MiR-1271 as a tumor suppressor in breast cancer proliferation and progression via targeting SPIN1. *Eur Rev Med Pharmacol Sci*. 2018;22(9):2697-706. DOI: 10.26355/eurrev\_201805\_14966

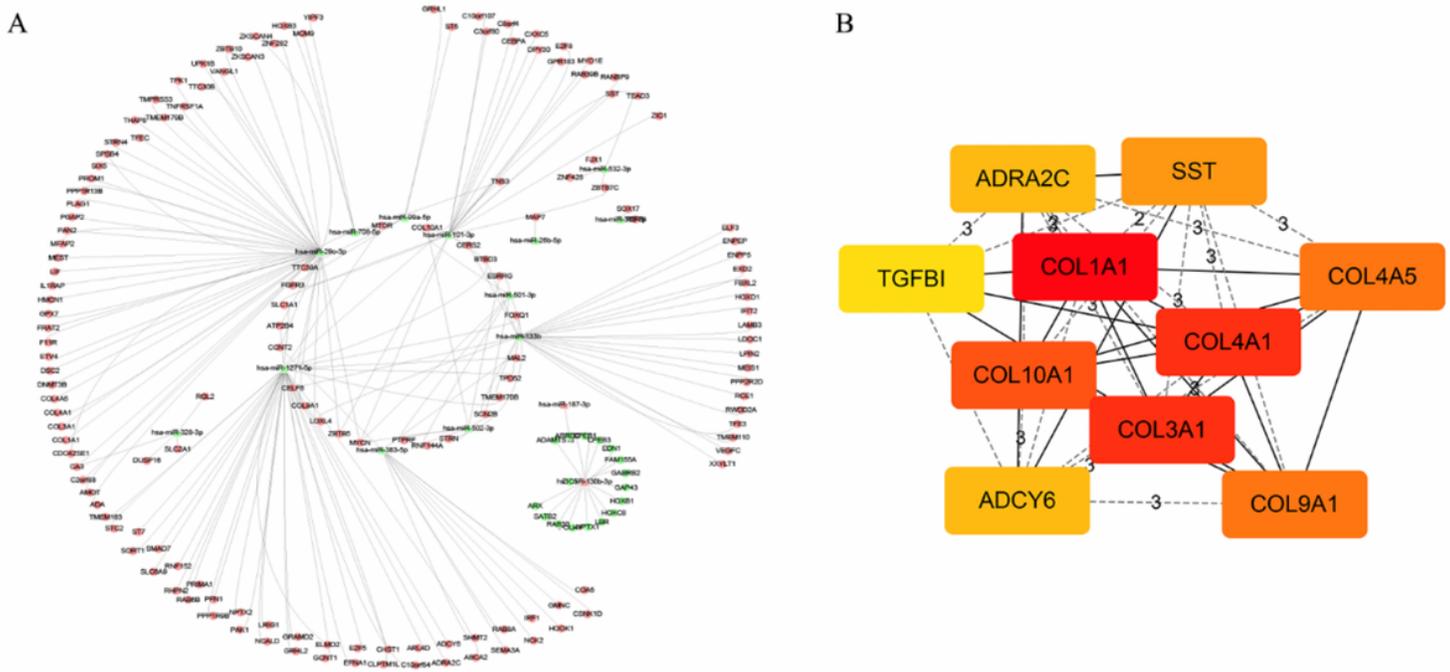
16. Liu X, Li G. MicroRNA-133b inhibits proliferation and invasion of ovarian cancer cells through Akt and Erk1/2 inactivation by targeting epidermal growth factor receptor. *International journal of clinical and experimental pathology*. 2015;8(9):10605.

## Figures



**Figure 1**

Volcanic maps distribution of differential expression. (A): differentially expressed miRNA; (B): differentially expressed gene. Horizontal axis:  $\log_2(\text{FC})$ ; vertical axis:  $-\log_{10}(\text{adj.P.Val})$ . The FC represented the fold change in expression of tumor samples compared to normal samples, and the adj.P.Val represented the calibrated P value. Green color: differential expression was down-regulated; red color: differential expression was up-regulated.



**Figure 2**

Figure2A: Regulation network of miRNA-gene. The ovals represented differential target genes; the triangles represented differential miRNAs, and the lines represented the existence of targeted regulatory relationships. Red color meant up-regulation; green color meant down-regulation. Figure2B: Identification network of key DEGs. The darker the red was, the higher the score was, and the more significant the biometric significance was.

## Supplementary Files

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

- [network.txt](#)
- [type.txt](#)