

# Identification of Pathways and Genes Associated with Chemoradiotherapy in Locally Advanced Cervical Cancer

**Zhenhua Zhang**

The Affiliated Hospital of Southwest Medical University

**Yao Zhang**

Southwest Medical University

**Yongshun Ma**

Southwest Medical University

**Shixin Xiang**

Southwest Medical University

**Jing Shen**

Southwest Medical University

**Yueshui Zhao**

Southwest Medical University

**Xu Wu**

Southwest Medical University

**Mingxing Li**

Southwest Medical University

**Xiao Yang**

Southwest Medical University

**Fukuan Du**

Southwest Medical University

**Lin Liu**

Southwest Medical University

**Zhangang Xiao** (✉ [zhangangxiao@swmu.edu.cn](mailto:zhangangxiao@swmu.edu.cn))

Southwest Medical University <https://orcid.org/0000-0003-4265-5833>

**Qinglian Wen**

The Affiliated Hospital of Southwest Medical University

---

## Research Article

**Keywords:** locally advanced cervical cancer, radiotherapy and chemotherapy, responders and non-responders, differentially expressed gene, single-cell sequence

**Posted Date:** December 13th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1106713/v1>

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

[Read Full License](#)

---

# Abstract

The standard treatment approach for locally advanced cervical cancer (LACC) is concurrent chemoradiotherapy (CCRT). However, resistance to radiotherapy and chemotherapy often leads to failure of LACC therapy. Thus, the key pathways and genes associated with CCRT in LACC should be identified urgently. The Weighted Gene Co-Expression Network Analysis (WGCNA) was used for the identification of highly correlated gene modules. A protein-protein interaction (PPI) network was constructed using STRING and hub genes were selected. Furthermore, single-cell transcriptome sequencing was used to elucidate the cell type composition of the cervix sample and analyze the expression levels of key genes in cells. We identified 580 differentially expressed genes (DEGs) in LACC, which were mainly involved in Human papillomavirus infection, focal adhesion, and ECM-receptor interaction signaling pathways. Ten hub genes (COL1A1, COL6A1, COL6A2, LAMA4, COL6A3, LAMC1, HSPG2, ITGA9, CTGF, PDGFRB) were screened for further study. We showed that COL1A1, COL6A1, COL6A2 were highly expressed after radiotherapy or chemoradiotherapy. By analyzing the single-cell sequence, we found that the main cell types in cervical tissue include Fibroblasts, Smooth muscle cells, Tissue stem cells, Endothelial cells, Progenitor cells, Epithelial cells, T cells, Basal cells, Macrophages, and Mast cells. COL1A1, COL6A1, COL6A2, COL6A3, CTGF, and PDGFRB were highly expressed in Progenitor cells. Human papillomavirus infection, focal adhesion, and ECM-receptor interaction signaling pathway are related to the failure of CCRT for LACC, which warrants further research to improve CCRT sensitivity in LACC targeting on these candidate genes or pathways.

## Introduction

Cervical cancer is one of the leading causes of cancer death among women worldwide. The treatment of cervical cancer remains a major therapeutic challenge for researchers and physicians [1]. Moreover, morbidity and mortality in developing countries are significantly higher compared with developed countries [2]. There are two types of cervical cancer, including squamous cell carcinoma and adenocarcinoma, of which squamous cell carcinoma accounts for the largest proportion [3, 4]. One report from a developing country has shown that over 80% of new cervical cancer cases are found at advanced stages [5]. At present, the treatment of LACC mainly includes operation, chemotherapy, radiotherapy, or combined treatment [6-8]. CCRT based on cisplatin is the primary treatment for patients with LACC (patients with stage IB2-IVA), CCRT was reported to improve overall survival and progression-free survival in LACC [9]. Cisplatin is used to treat a variety of cancers, it interferes with the DNA repair mechanism and causes DNA damage, induces apoptosis of cancer cells [10]. In the past work, we have found that treatment resistance for some patients with LACC [11-13]. In spite of the improvement in survival of cervical cancer after concurrent chemoradiotherapy, the outcomes of patients with LACC have been unsatisfactory. For patients who do not respond to CCRT, other effective treatment options need to be adopted.

It has been found that 27-gene profile can better predict the clinical outcome of LACC patients receiving CCRT as soon as possible after diagnosis, including ZNF238, SAP30, C10orf137, UHRF1, SUZ12, HMGN4,

RBBP4, PPP1CB, SLFN11, FLJ39378, ENDOGL1, RECQL, TRPC1, TRIO, DNAH6, GNL3L, SLC36A2, SRP9, RPE, LDOC1L, PUS7L, CCDC89, LOC644921, PLEKHG1, FAM111B, RPRD2 and ETAA1 [14]. In addition, Other study have found that DNA replication, recombination and repair are one of the reasons for the resistance of radiotherapy and chemotherapy in LACC [15]. However, the reasons for the resistance of CCRT for LACC have not been fully explained, and there are still few reports on this aspect. Therefore, we need to look for genes and signaling pathways associated with the treatment of LACC.

In the present study, we explored the interaction network of DEGs between responders and non-responders along with interrelated signaling pathways in LACC by analyzing the expression profile of gene expression microarray data (GSE56303-GPL10191, GSE56303-GPL16025, and GSE56363) using bioinformatics tools. The purpose of this study is to find more CCRT sensitivity-related key genes and pathways, which might serve as targets to increase CCRT sensitivity. Moreover, the expression of key genes in cervical tissue cells was analyzed. We attempt to elucidate the causes of failure of radiotherapy and chemotherapy for LACC.

## Results

### Identification of the DEGs between the response group and the non-response group of CCRT therapy in LACC

In order to study the difference of gene expression between the CR and NR groups for LACC, the limma package was used to estimate the fold changes and standard errors by fitting a linear model for each gene for the assessment of differential expression. As shown in Fig. 1a, we set the cut-off criteria as  $p$ -value  $< 0.05$  and  $|\log_2FC| > 0.5$  to screen the DEGs. There were 1358 significant differential genes in GSE56303-GPL10191, including 860 up-regulated genes ( $p$ -value  $< 0.05$  and  $\log_2$  fold change  $> 0.5$ ) and 498 down-regulated genes ( $p$ -value  $< 0.05$  and  $\log_2$  fold change  $< -0.5$ ). There are 6048 differential genes in GSE56303-GPL16025, among which 3204 are up-regulated genes and 2844 are down-regulated. There are 4214 differentially expressed genes in GSE56363, including 2133 up-regulated genes and 2081 down-regulated genes. In summary, 580 up-regulated genes and 153 down-regulated genes represent the differentially regulated genes in two or more datasets (Fig. 1b). We used the cumulative hypergeometric distribution function to estimate if the up-regulated genes were statistically significant. The results show that the adjusted  $P$  values of the overlapping genes in GSE56306-GPL10191 and GSE56306-GPL16025 is  $1.515894E-73$ , the adjusted  $P$  values of the overlapping genes in GSE56363 and GSE56306-GPL10191 is  $4.30E-09$ , the adjusted  $P$  values of the overlapping genes in GSE56363 and GSE56306-GPL10191 is  $7.384116E-146$ . In addition, the cumulative hypergeometric distribution function was used to estimate if the down-regulated genes were statistically significant. The results show that the adjusted  $P$  values of the overlapping genes in GSE56306-GPL10191 and GSE56306-GPL16025 is  $2.024604E-30$ , the adjusted  $P$  values of the overlapping genes in GSE56306-GPL10191 and GSE56363 is 1, the adjusted  $P$  values of the overlapping genes in GSE56306-GPL16025 and GSE56363 is 0. Therefore, the up-regulated genes were statistically significant, 580 up-regulated genes were selected for further analysis.

## DEGs were enriched to key functions and pathways in LACC

In order to understand the functional roles of DEGs, GO and KEGG pathway enrichment analyses were performed. The GO analysis of the 580 up-regulated DEGs was mainly enriched for the BP terms cell adhesion and extracellular matrix organization. GO analysis in the category CC showed that the DEGs were mainly accumulated in the extracellular matrix, extracellular exosome, membrane. The MF of the DEGs was mainly related to protein binding and calcium ion binding (Fig. 2a). Additionally, Up-regulated DEGs were significantly associated with Human papillomavirus infection (Fig S1), Focal adhesion, Lysosome, Hippo signaling pathway, Protein processing in endoplasmic reticulum, and ECM-receptor interaction signaling pathways (Fig. 2b). We showed the correlation between genes and pathways (Fig. 2c, Table S2). Furthermore, we analyzed the correlation between pathways and pathways, and we found that Human papillomavirus infection, Focal adhesion, and ECM-receptor interaction signaling pathways were correlated (Fig. 2d). In addition, we performed GO and KEGG enrichment analysis for each cohort (GSE56303-GPL10191, GSE56303-GPL16025, and GSE56363). The GO analysis of GSE56303-GPL10191 were mainly enriched in transcription factor activity, RNA polymerase II proximal promoter sequence-specific DNA binding, transport vesicle, and positive regulation of cell adhesion. The GO analysis of GSE56303-GPL16025 was mainly enriched in ATPase activity, chromosomal region, and DNA replication. The GO analysis of GSE56363 was mainly enriched in cell adhesion molecule binding, extracellular matrix, and extracellular matrix organization (Fig S2). Further, from the KEGG pathway enrichment analysis, we found that these 3 cohorts were mainly enriched in the pathway of the Human papillomavirus infection (Fig S3).

## Identification of genes associated with clinical features based on WGCNA in LACC

In order to understand the relationship between DEGs and clinical features, we combine the data into one dataset, genes with p-value < 0.05 and  $\log_2$  fold change >0.5 were involved to perform the WGCNA analysis by using WGCNA package in R software. We have chosen the soft threshold power 6 to define the adjacency matrix based on the criterion of approximate scale-free topology (Fig S4). The network and the 4 identified modules are depicted in Fig. 3a, b. Fig. 3c provides an alternate visualization of the module structure via a multi-dimensional scaling plot (standard R function cmdscale). Furthermore, we performed cluster analysis, Fig. 3d depicts the eigengene network using a dendrogram (hierarchical cluster tree). In general, 4 clusters were grouped into two clusters, the blue and grey modules are highly correlated. Combined with Fig. 3e, the blue modules were related to the therapeutic effect. The blue module has a negative correlation with CR (Correlation =-0.2 and p value=0.04) and a positive correlation with NR (Correlation =0.2 and p value=0.04). To better understand the relationships that exist between disease associated genes. We constructed PPI networks from genes enriched in six pathways (Table S2) and genes obtained by WGCNA analysis (Table S3). The analysis was performed using the STRING database and Cytoscape. PPI network contains 52 nodes and 158 edges (Fig. 3f). A total of 20 hub genes were screened out, the ranking and scores of genes were listed in Table S4, these 20 hub genes were chosen for further analysis.

## The key genes were significantly down-regulated and positively correlated in LACC

In order to explore the expression patterns of the hub genes in tumors and normal tissues, the GEPIA 2 was used for detection. Compared with healthy people, the expression levels of COL1A1, COL6A1, COL6A2, LAMA4, COL6A3, LAMC1, HSPG2, ITGA9, CTGF, and PDGFRB in cervical cancer patients were significantly decreased. In addition, the expression levels of PDIA6, RPN1, and RPN2 were higher in cancer patients than in normal subjects (Fig. 4a). Furthermore, COL1A1, COL6A1, COL6A2, LAMA4, COL6A3, HSPG2, ITGA9, CTGF, and PDGFRB had significant positive correlations among each other (Fig. 4b).

## Ten Key Genes Dysregulated Upon Chemoradiotherapy

Through ROC curve analysis, we found that COL1A1, COL6A1, COL6A2, LAMA4, COL6A3, LAMC1, HSPG2, ITGA9, CTGF, and PDGFRB had better diagnostic performance in distinguishing responders from non-responders in datasets of GSE56303-GPL16025 and GSE56363. The area under the curve values of these genes was all greater than 0.7 in GSE56303-GPL16025 and GSE56363 (Fig. 5a). More importantly, we compared the changes in the expression levels of these 10 key genes after radiotherapy or chemoradiotherapy, and we found that the expression levels of COL1A1, COL6A1, COL6A2 were significantly increased in cervical cancer patients after radiotherapy or chemoradiotherapy (Fig. 5b).

## Six key genes were highly expressed in the Progenitor cells of cervical tissue

In order to understand the gene expression at the single-cell level, we performed the single-cell analysis of a cervical sample. The sample contained 10,000 cells, the filtering parameters were set to  $nFeature\_RNA > 200$  &  $nFeature\_RNA < 2000$  &  $Percent.mt < 40$  &  $nCount\_RNA < 4000$ , and 9955 cells were included in the analysis (Fig S5a). We selected 20 principal components (PCs) with an estimated p-value  $< 0.05$  for subsequent analysis (Fig S5b). Afterward, the uniform manifold approximation and projection (UMAP) algorithm was applied, and cells in the cervix were successfully classified into 12 separate clusters (Fig. 6a). We showed the expression levels of the top 5 marker genes in each cluster with a heat map (Fig S5c). According to the expression patterns of the marker genes, these clusters were annotated by singleR, CellMarker, and PanglaoDB (Fig. 6b), the cell marker genes used for cluster annotation are shown in Table S5. Cluster 0, 3, and 4, containing 4126 cells, were annotated as Fibroblasts; clusters 1 and 2, containing 3215, were annotated as Smooth muscle cells; cluster 5, containing 657 cells, was annotated as Tissue stem cells; cluster 6, containing 521 cells, was annotated as Endothelial cells; cluster 7, containing 485 cells, was annotated as Progenitor cells; cluster 8, containing 389 cells, was annotated as Epithelial cells; cluster 9, containing 217 cells, was annotated as T cells; cluster 10, containing 182 cells, was annotated as Basal cells; cluster 11, containing 116 cells, was annotated as Macrophages; cluster 12, containing 47 cells, was annotated as Mast cells. We analyzed the expression of 10 key genes in cervical sample cells, and the results showed that COL1A1, COL6A1, COL6A2, COL6A3, CTGF, and PDGFRB were highly expressed in progenitor cells (cluster 7) (Fig. 6c).

## Discussion

Cervical cancer remains a major health problem for women in developing countries. Cervical cancer is deadly because most patients are already advanced at the time of diagnosis [16, 17]. About half of patients with LACC will relapse or metastasize within the first two years after treatment, up to 40% of patients have no response to conventional treatment [18]. The treatment of LACC is a complex biological process. In recent years, although many new treatment methods have emerged, CCRT is still the main treatment method [19, 20]. The reasons for the ineffectiveness of CCRT in the treatment of LACC are not fully understood, and studies at the signaling pathways may help to explain this phenomenon. In this study, we have integrated the data of responders and non-responders to radiotherapy and chemotherapy for LACC, 10 key genes (COL1A1, COL6A1, COL6A2, COL6A3, LAMA4, LAMC1, HSPG2, ITGA9, CTGF, and PDGFRB) were identified by analyzing LACC patients with CCRT sensitivity, the expression levels of these genes in cervical cancer were decreased and correlated positively with each other. We found that COL1A1, COL6A1, COL6A2 were up-regulated after chemoradiotherapy, which may be related to the stress response. In addition, we also discussed the cell types of cervical samples and the expression levels of these 10 genes in the cells indicated that COL1A1, COL6A1, COL6A2, COL6A3, CTGF, and PDGFRB were highly expressed in Progenitor cells. In general, we have studied the expression of key genes in cervical tissues, the expression changes in cancer compared with healthy people, and the expression changes after radiotherapy and chemotherapy (Fig. 7). These results may be the basis for developing new targeted strategies to improve the prognosis of cervical cancer.

In the previous studies, COL1A1 was confirmed to exert radio-resistance effects in nasopharyngeal carcinoma cells, and knockdown of COL1A1 inhibits nasopharyngeal carcinoma cells sensitivity to irradiation [21]. Other studies have found that COL1A1 plays a crucial role in cervical cells anti-apoptosis induced by radiation [22]. Collagen type VI (COL6; encoded by COL6A1, COL6A2, and COL6A3) has been shown to stimulate proliferation, suppress apoptosis [23]. Overexpression of LAMA4 enhanced cisplatin resistance of Gastric cancer cells, while LAMA4 knockdown led to opposite results [24]. CCN2 (also known as CTGF), overexpression of CCN2 increased the resistance of human osteosarcoma to cisplatin-mediated cell apoptosis [25].

In this study, failure of chemoradiotherapy for cervical cancer was found to be associated with Human papillomavirus infection, focal adhesion, and ECM-receptor interaction signaling pathways. In addition, COL1A1, COL6A1, COL6A2, LAMA4, COL6A3, LAMC1, HSPG2, ITGA9, CTGF, and PDGFRB were mainly enriched in the Human papillomavirus infection, focal adhesion, and ECM-receptor interaction signaling pathways. We speculate that these genes affect the results of chemoradiotherapy in LACC by regulating human papillomavirus infection, focal adhesion, and ECM receptor interaction signaling pathways.

Cervical cancer is the second most common cancer in women worldwide, and persistent infection with one of about 15 genotypes of carcinogenic human papillomavirus (HPV) causes almost all cases, HPV has been identified as a major factor that leads to cervical cancer [26]. Among the myriad of microenvironmental factors impacting on cancer cell resistance, cell adhesion to the extracellular matrix

(ECM) has recently been identified as a key determinant [27]. The extracellular matrix (ECM) drives metastasis via its interaction with the integrin signaling pathway, contributes to tumor progression, and confers therapy resistance by providing a physical barrier around the tumor [28]. Previous studies have shown that cellular responsiveness to external stress (such as radiation and chemotherapy) is affected by the differential expression of genes and proteins involved in the regulation of the ECM, cell adhesion, and gap junction signaling [29].

In summary, our analysis found that chemoradiotherapy resistance in LACC is associated with human papillomavirus infection, focal adhesion, and ECM receptor interaction signaling pathways. And we identified key genes that might be associated with chemoradiotherapy failure, which might be useful for the accurate diagnosis and treatment of LACC.

we still have a long way to go to gain a clear understanding of the mechanisms of chemoradiotherapy failure in CCRT. In order to apply the findings to the treatment of patients, we need to do more work: (i) large-sample studies are needed to validate the findings; (ii) in order to analyze the effect of target gene expression on the efficacy of radiotherapy and chemotherapy for advanced cervical cancer, more experiments should be encouraged to further verify our results; (iii) It is necessary to study the cell types and gene expression of responders and non-responders using single-cell techniques.

## Materials And Methods

### Data acquisition and processing

Microarray gene expression profiles were obtained from a previous study published in the GEO database [30] (GEO Accession number GSE56303 [14]; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56303> and GEO Accession number GSE56363 [15]; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56363>). The keyword of the search is "locally advanced cervical cancer". The restricted research type is expression profiling by array, and the restricted species is Homo sapiens. The selected data meets the following criteria: (1) The data set must be the expression mRNA chip data of the whole cervical cancer genome; (2) The data must be a controlled study of the effectiveness and inefficacy of CCRT for LACC; (3) The data set case (responders) - control (non-responders) group must include or exceed 3 Samples; (4) Information on treatment results for each sample must be provided. Data sets meeting the above criteria will be included in this study. GSE56303 included two platforms: GPL10191 and GPL16025. The platform for GSE6363 was GPL4133 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number version). The GSE56303 contains gene expression profiles of 63 responders (complete response) and 22 non-responders (partial response and stable disease) with pathological diagnosis of FIGO staged IB2 up to IIIB. The GSE56363 contains 21 patients with LACC (FIGO stage IIB-IIIB): 12 responders and 9 non-responders (**Table S1**). The raw probe-level data (.pair file) of GSE56303-GPL10191 and GSE56303-GPL16025 were downloaded from GEO. The RMA [31] in the package of R was used for background correction, quantile normalization and log<sub>2</sub> transformation, respectively. We downloaded the raw data of GSE56363. Then, log<sub>2</sub>

transformation was performed on all gene expression data, and the quantile normalization of median signals was carried out by `normalizeBetweenArrays` function.

### **Identification of DEGs in responders and non-responders**

Limma R package [32] was used to identify the DEGs between responders and non-responders through R software (version 3.5.2; <https://cran.r-project.org/>). Fold-change (FC) values were calculated and the DEGs were further selected based on the following cutoff criteria: p-value <0.05 and the absolute value of  $\log_2$  FC (fold change) >0.5. The intersecting DEGs of responders and non-responders were used for further analysis.

### **Statistical meaning of the given overlap of two sets**

The cumulative hypergeometric distribution function was used to estimate if the given overlap between two sets is statistical meaning. We used the `Phyper` function in R software for calculation, and the code is as follows:

```
Phyper (k-1, M, N-M, n, lower.tail=F)
```

N is the total number of gene probes, M is the number of DEGs in one dataset, n is the number of DEGs in another dataset, and K is a numerical representation of shared genes. The P value was corrected for multiple testing (Benjamin-Hochberg).

### **Gene ontology and pathways enrichment analysis**

To identify the enriched function of 580 up-regulated DEGs, we used the Database for Annotation, Visualization and Integrated Discovery online tool [33] to perform Gene Ontology (GO) analyses. GO analysis including three independent categories: biological process (BP), cellular component (CC), and molecular function (MF). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs was conducted using the `clusterProfiler` package [34]. Adjusted p-value < 0.05 as the threshold level for statistical significance.

### **Identification of genes associated with clinical characteristics**

In this study, DEGs were involved to perform the WGCNA analysis by the R package WGCNA [35]. WGCNA uses the topological overlap measure (TOM) as a proximity measure to cluster genes into network modules that combine the adjacency of two genes and the connection strengths with which these two genes interact with other neighbor genes. A soft thresholding parameter was employed to construct a weighted network. Here, we constructed co-expression modules through WGCNA related to clinical traits, and obtained genes associated with clinical traits through WGCNA.

### **PPI network construction and hub genes identification**

We used STRING [36] to search the potential interactions between known proteins and predicted proteins. It is the largest database for protein interaction information at present. Through this database, we construct a PPI network, which was visualized using Cytoscape (version 3.7.1) [37]. The 20 hub genes were selected from the PPI network using the cytoHubba Cytoscape plugin, and maximal clique centrality (MCC) was identified [38].

### **Expression analysis of 20 hub genes**

The Gene Expression Profiling Interactive Analysis (GEPIA 2, <http://gepia2.cancer-pku.cn/>) webserver has been a valuable and highly cited resource for gene expression analysis based on tumor and normal samples. To compare the difference in gene expression between cervical cancer patients and healthy people, we used GEPIA 2 [39] for analysis. The Cancer Genome Atlas (TCGA, <https://cancergenome.nih.gov/>) database is a publicly funded project aimed at classifying and discovering major genomic changes in cancer to create a comprehensive cancer genome map. In this study, Gene expression data of cervical cancer patients were obtained from the TCGA database and co-expression analysis was performed. The correlation of mRNA expression was analyzed by the Pearson test.  $P < 0.05$  was considered to indicate statistically significant differences.

### **Further analysis of 10 key genes in cancer.**

Receiver operating characteristic (ROC) curve analysis was performed to assess the diagnostic performance in characterizing responders versus non-responders and determine the area under the ROC curve. To compare the difference in gene expression before and after chemoradiotherapy, we searched the public database GEO database (GSE3578 [40]; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE3578>). GSE3578 was used to verify the expression changes of 10 key genes before and after chemoradiotherapy. Paired T-test was used and  $P < 0.05$  was statistically significant.

### **Single-cell sequence analysis of a cervical tissue**

We performed a single-cell transcriptome sequencing analysis of a cervical sample (GSM3980130 [41]; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3980130>). The Seurat package [42] in R 4.0.0 was used for quality control, statistical analysis, and exploration of the scRNA-seq data. The sample contained 10,000 cells, and a total of 9955 cells were included in this analysis after quality control. Then, the gene expression of the remaining 9955 cells was normalized using a linear regression model. PCA was performed to identify significantly available dimensions with a  $P$ -value  $< 0.05$  [43]. Then, the uniform manifold approximation and projection (UMAP) [44] algorithm was applied for dimensionality reduction with 20 PCs and for performing cluster classification analysis. The differential expression analysis among all genes within cell clusters was performed to identify the marker genes of each cluster. Afterward, different cell clusters were determined and annotated by the singleR [45] package according to the marker genes and were then manually verified and corrected with the CellMarker [46] database and PanglaoDB [47]. Here, we elucidated the cell type composition of the cervix sample by scRNA-seq sequence analysis and analyze the expression levels of 10 key genes in cells.

# Declarations

## Funding

This work was supported by National Natural Science Foundation of China (grant number: 81672444, 81972643), Sichuan Science and Technology Project (grant number: 2018JY0079), and Luxian - Southwest Medical University Strategic Cooperation Achievement Transfer and Transformation Cultivation Project (grant number: 2019LXXNYKD-07).

## Competing Interests

The authors have no conflict of interest.

## Author Contributions

ZX, JL, and WQ supervised the project. ZZ, YZ, and YM collected the data and analyzed data. SX, JS, YZ, XW, ML, and XY wrote the manuscript and commented on the manuscript. FD, and LL provided scientific expertise.

## Data Availability

The datasets analyzed for this study are available in the TCGA (<https://cancergenome.nih.gov/>) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>) databases.

# References

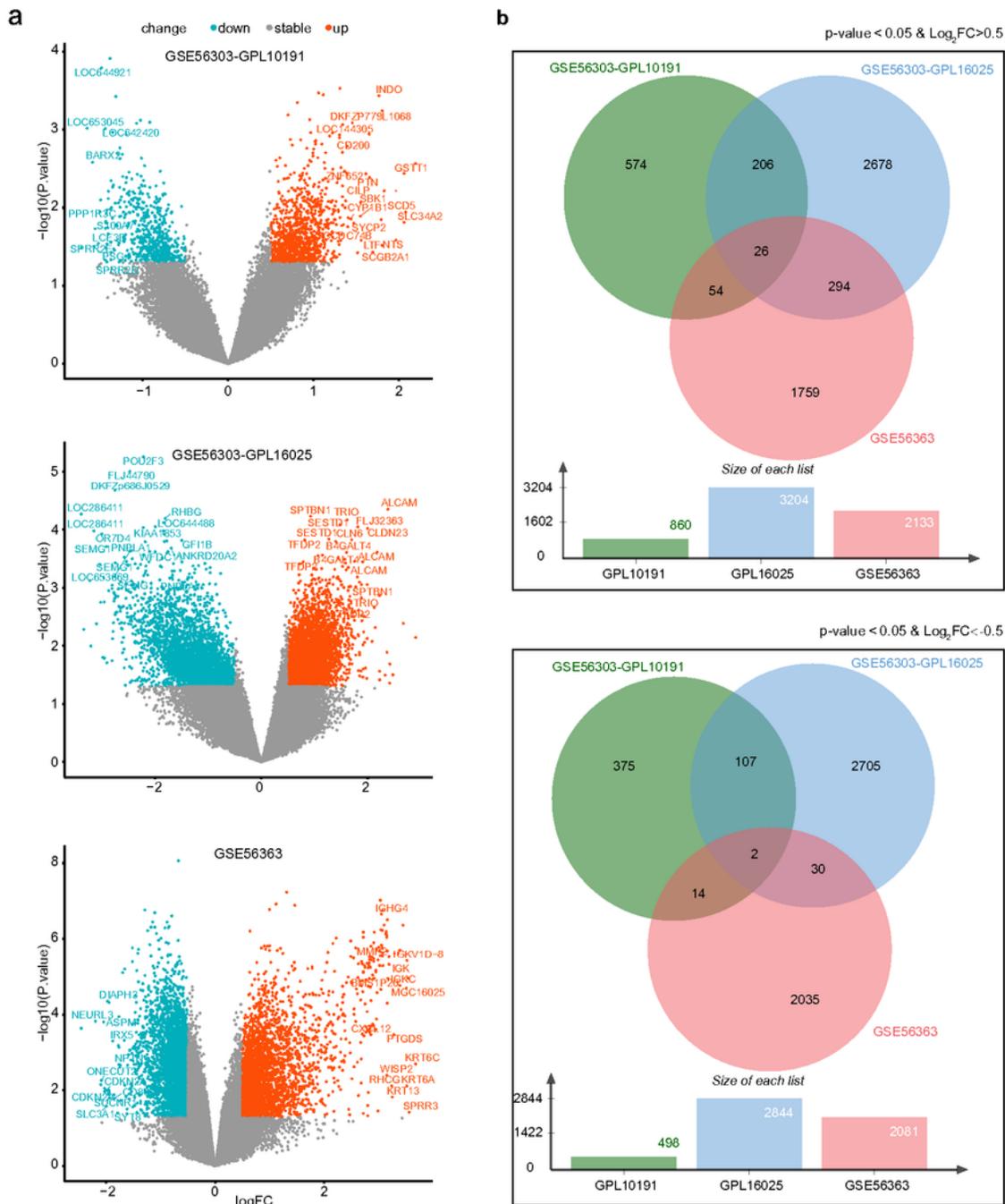
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*. 2015;65(2):87-108.
3. Lonnberg S, Hansen BT, Haldorsen T, Campbell S, Schee K, Nygard M. Cervical cancer prevented by screening: Long-term incidence trends by morphology in Norway. *International journal of cancer*. 2015;137(7):1758-64.
4. Williams NL, Werner TL, Jarboe EA, Gaffney DK. Adenocarcinoma of the cervix: should we treat it differently? *Current oncology reports*. 2015;17(4):17.
5. Kokka F, Bryant A, Brockbank E, Powell M, Oram D. Hysterectomy with radiotherapy or chemotherapy or both for women with locally advanced cervical cancer. *The Cochrane database of systematic reviews*. 2015(4):CD010260.
6. Yang PM, Chou CJ, Tseng SH, Hung CF. Bioinformatics and in vitro experimental analyses identify the selective therapeutic potential of interferon gamma and apigenin against cervical squamous cell carcinoma and adenocarcinoma. *Oncotarget*. 2017;8(28):46145-62.

7. Waggoner SE. Cervical cancer. *Lancet*. 2003;361(9376):2217-25.
8. Marquina G, Manzano A, Casado A. Targeted Agents in Cervical Cancer: Beyond Bevacizumab. *Current oncology reports*. 2018;20(5):40.
9. Pearcey R, Miao Q, Kong W, Zhang-Salomons J, Mackillop WJ. Impact of adoption of chemoradiotherapy on the outcome of cervical cancer in Ontario: results of a population-based cohort study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2007;25(17):2383-8.
10. Dasari S, Tchounwou PB. Cisplatin in cancer therapy: molecular mechanisms of action. *European journal of pharmacology*. 2014;740:364-78.
11. Kumar L, Harish P, Malik PS, Khurana S. Chemotherapy and targeted therapy in the management of cervical cancer. *Current problems in cancer*. 2018;42(2):120-8.
12. Lippert TH, Ruoff HJ, Volm M. Intrinsic and acquired drug resistance in malignant tumors. The main reason for therapeutic failure. *Arzneimittel-Forschung*. 2008;58(6):261-4.
13. Zhu H, Luo H, Zhang W, Shen Z, Hu X, Zhu X. Molecular mechanisms of cisplatin resistance in cervical cancer. *Drug design, development and therapy*. 2016;10:1885-95.
14. Fernandez-Retana J, Lasa-Gonsebatt F, Lopez-Urrutia E, Coronel-Martinez J, Cantu De Leon D, Jacobo-Herrera N, et al. Transcript profiling distinguishes complete treatment responders with locally advanced cervical cancer. *Translational oncology*. 2015;8(2):77-84.
15. Balacescu O, Balacescu L, Tudoran O, Todor N, Rus M, Buiga R, et al. Gene expression profiling reveals activation of the FA/BRCA pathway in advanced squamous cervical cancer with intrinsic resistance and therapy failure. *BMC cancer*. 2014;14:246.
16. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: a cancer journal for clinicians*. 2011;61(2):69-90.
17. Parkin DM, Bray F, Ferlay J, Pisani P. Estimating the world cancer burden: Globocan 2000. *International journal of cancer*. 2001;94(2):153-6.
18. Scatchard K, Forrest JL, Flubacher M, Cornes P, Williams C. Chemotherapy for metastatic and recurrent cervical cancer. *The Cochrane database of systematic reviews*. 2012;10:CD006469.
19. Green J, Kirwan J, Tierney J, Vale C, Symonds P, Fresco L, et al. Concomitant chemotherapy and radiation therapy for cancer of the uterine cervix. *The Cochrane database of systematic reviews*. 2005(3):CD002225.
20. Shrivastava S, Mahantshetty U, Engineer R, Chopra S, Hawaldar R, Hande V, et al. Cisplatin Chemoradiotherapy vs Radiotherapy in FIGO Stage IIIB Squamous Cell Carcinoma of the Uterine Cervix: A Randomized Clinical Trial. *JAMA oncology*. 2018;4(4):506-13.
21. Guo Y, Zhai J, Zhang J, Ni C, Zhou H. Improved Radiotherapy Sensitivity of Nasopharyngeal Carcinoma Cells by miR-29-3p Targeting COL1A1 3'-UTR. *Medical science monitor : international medical journal of experimental and clinical research*. 2019;25:3161-9.

22. Liu S, Liao G, Li G. Regulatory effects of COL1A1 on apoptosis induced by radiation in cervical cancer cells. *Cancer cell international*. 2017;17:73.
23. Howell SJ, Doane KJ. Type VI collagen increases cell survival and prevents anti-beta 1 integrin-mediated apoptosis. *Experimental cell research*. 1998;241(1):230-41.
24. Peng L, Li Y, Wei S, Li X, Dang Y, Zhang W, et al. LAMA4 activated by Androgen receptor induces the cisplatin resistance in gastric cancer. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2020;124:109667.
25. Tsai HC, Huang CY, Su HL, Tang CH. CCN2 enhances resistance to cisplatin-mediated cell apoptosis in human osteosarcoma. *PloS one*. 2014;9(3):e90159.
26. Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. *Lancet*. 2013;382(9895):889-99.
27. Eke I, Cordes N. Focal adhesion signaling and therapy resistance in cancer. *Seminars in cancer biology*. 2015;31:65-75.
28. Kesh K, Gupta VK, Durden B, Garrido V, Mateo-Victoriano B, Lavania SP, et al. Therapy Resistance, Cancer Stem Cells and ECM in Cancer: The Matrix Reloaded. *Cancers*. 2020;12(10).
29. Bai C, Yang M, Fan Z, Li S, Gao T, Fang Z. Associations of chemo- and radio-resistant phenotypes with the gap junction, adhesion and extracellular matrix in a three-dimensional culture model of soft sarcoma. *Journal of experimental & clinical cancer research : CR*. 2015;34:58.
30. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*. 2002;30(1):207-10.
31. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249-64.
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015;43(7):e47.
33. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009;4(1):44-57.
34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012;16(5):284-7.
35. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008;9:559.
36. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*. 2019;47(D1):D607-D13.
37. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of proteome research*. 2019;18(2):623-32.

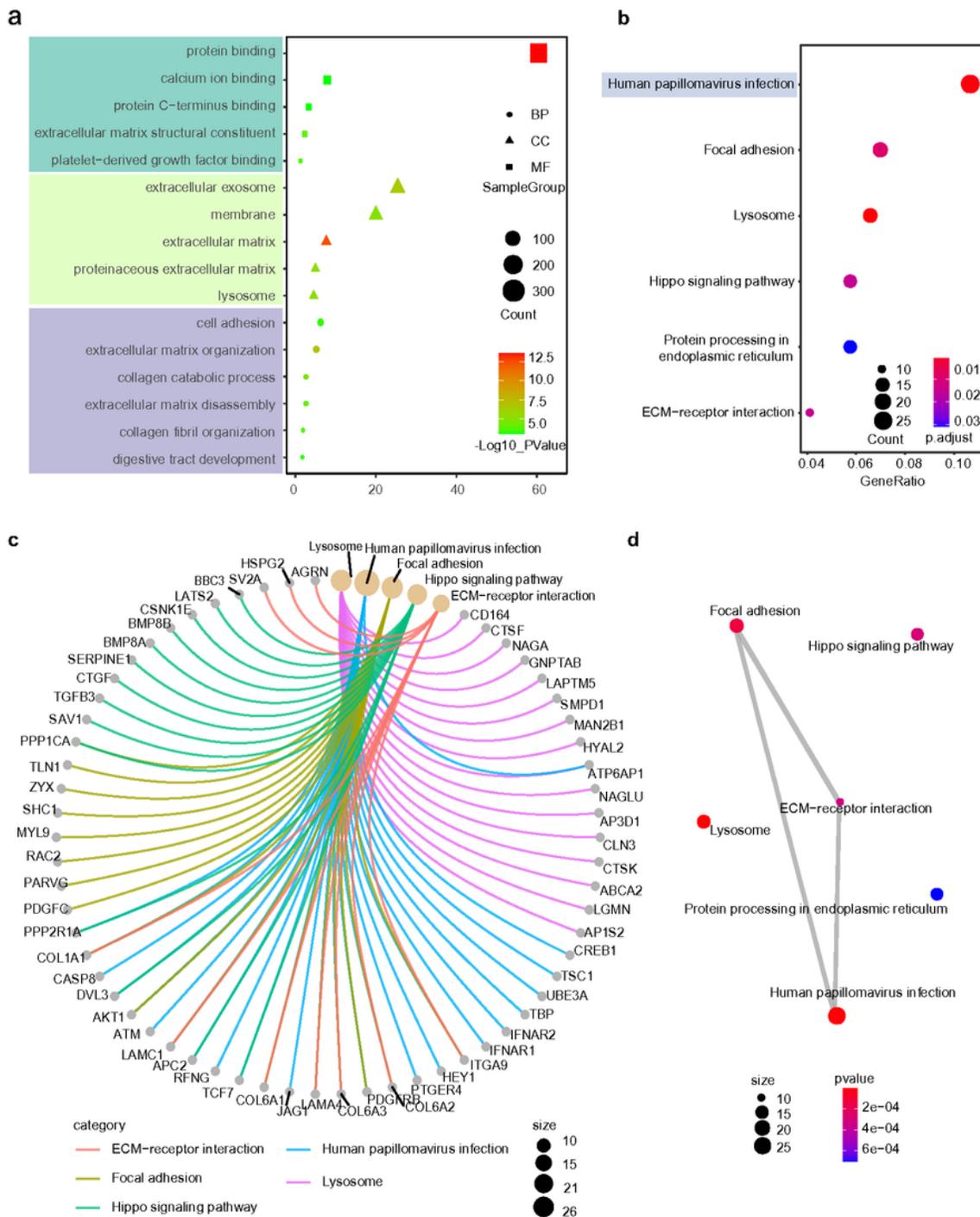
38. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. 2003;4:2.
39. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research*. 2019;47(W1):W556-W60.
40. Iwakawa M, Ohno T, Imadome K, Nakawatari M, Ishikawa K, Sakai M, et al. The radiation-induced cell-death signaling pathway is activated by concurrent use of cisplatin in sequential biopsy specimens from patients with cervical cancer. *Cancer biology & therapy*. 2007;6(6):905-11.
41. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, et al. Construction of a human cell landscape at single-cell level. *Nature*. 2020;581(7808):303-9.
42. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology*. 2018;36(5):411-20.
43. Lall S, Sinha D, Bandyopadhyay S, Sengupta D. Structure-Aware Principal Component Analysis for Single-Cell RNA-seq Data. *Journal of computational biology : a journal of computational molecular cell biology*. 2018.
44. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology*. 2018.
45. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology*. 2019;20(2):163-72.
46. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic acids research*. 2019;47(D1):D721-D8.
47. Franzen O, Gan LM, Bjorkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database : the journal of biological databases and curation*. 2019;2019.

## Figures



**Figure 1**

Identification of DEGs in LACC microarray datasets. (a) Volcano plots were used to show the overall distribution of DEGs between responders and non-responders, red dots represent up-regulated genes, light blue dots represent down-regulated genes. (b) Venn diagram of the DEGs in the 3 datasets. A total of 580 up-regulated DEGs and 153 down-regulated DEGs were screened out.



**Figure 2**

GO and KEGG signaling pathways analysis of the DEGs in LACC. (a) The BP/MF/CC of GO analysis of 580 up-regulated DEGs. The x-axis indicates the rich factor. The y-axis indicates the name of the GO term. The dot size means the gene number. The dot color indicates the  $-\log_{10}(p\text{-value})$ . (b) Scatter plot of KEGG pathway enrichment statistics. where each point represents the enrichment level, the color corresponds to the  $-\log_{10}(\text{adjust } p\text{-value})$ , and the size corresponds to the number of genes enriched for the given

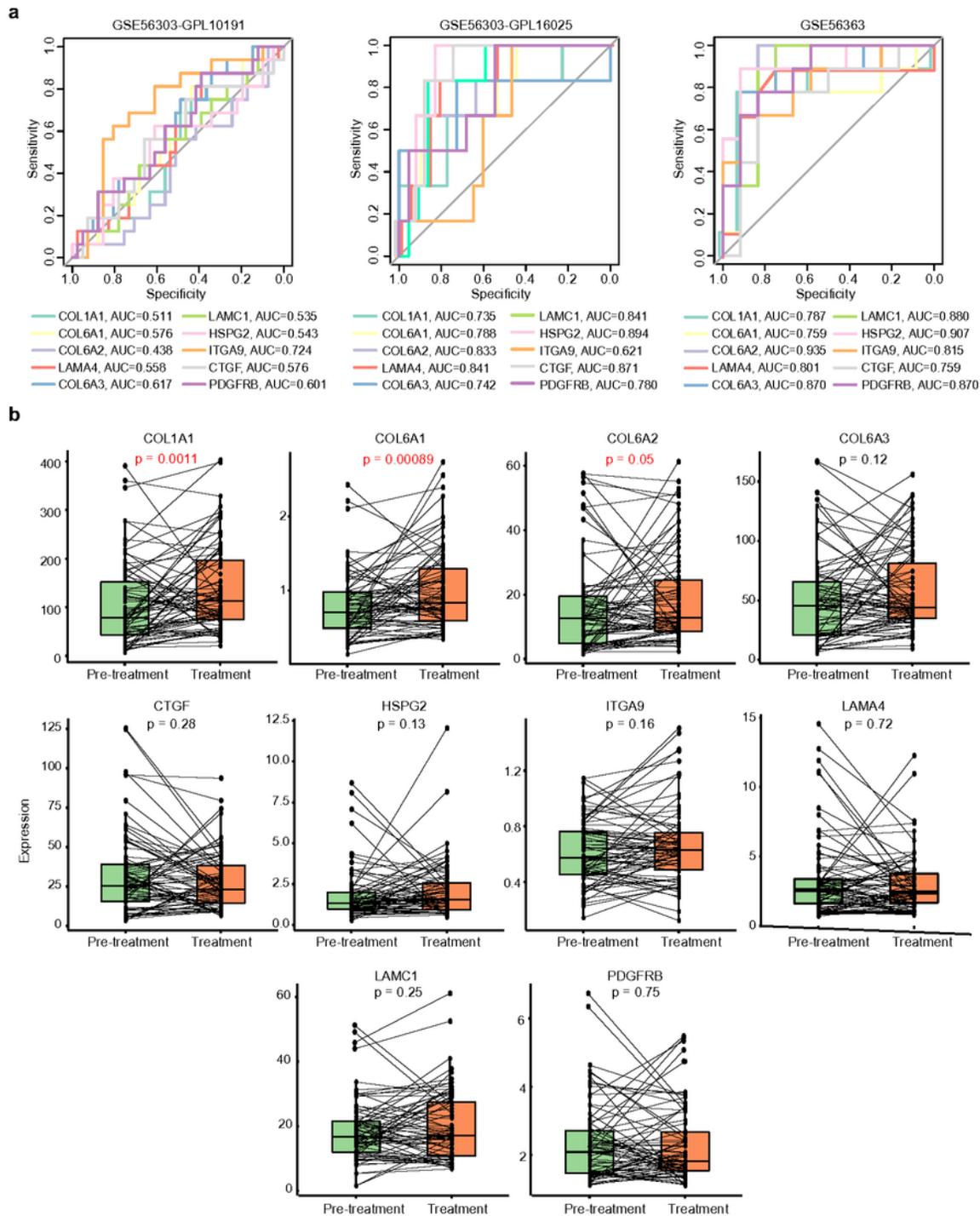
pathway. (c) cnetplot of KEGG signal pathway shown the “pathway-gene” network. (d) emapplot of KEGG signal pathway shown the “pathway-pathway” network. (BP: Biological process, MF: Molecular function, CC: Cellular component)

### Figure 3

Identification of genes associated with clinical features and hub genes identification. (a) Gene dendrogram obtained by average linkage hierarchical clustering. A total of 4 modules were identified. (b) Heatmap plot of topological overlap in the gene network. Each row and column corresponds to a gene, light color denotes low topological overlap, and progressively darker denotes higher topological overlap. (c) Results of multidimensional scaling. (d) Relationships among modules was summarized by a hierarchical clustering dendrogram of their eigengenes. (e) Relationships of consensus module eigengenes and clinical features. Numbers in the table report the correlations of the corresponding module eigengenes and treatment effect, with the p-values printed in parentheses. (f) There were a total of 52 DEGs in the PPI network complex. The nodes meant proteins; the edges meant the interaction of proteins. Hub genes screened by MCC algorithms from cytoHubba, the darker the color, the bigger the score.

### Figure 4

Expression analysis of 20 hub genes. (a) Comparison of expression levels between cervical cancer patients and normal controls. Red represents the cancer group, and green represents the normal group. (b) The heat map shows the co-expression relationship of hub genes, red represents negative correlation, blue represents positive correlation, and the larger the circle is, the smaller the p-value is.



**Figure 5**

Further analysis of 10 key genes in cancer. (a) ROC curve showed the accuracy of gene expression in predicting the results of chemoradiotherapy. Different colors indicate different genes. (AUC, the area under the curve; ROC, receiver operating characteristic curve) (b) Changes of expression levels of 10 key genes before and after chemoradiotherapy. Green indicates the level of gene expression before

radiotherapy or chemoradiotherapy; Red indicates the level of gene expression after radiotherapy or chemoradiotherapy. (Paired T-test,  $P < 0.05$  was statistically significant)

## Figure 6

The cell types of cervical tissue and the expression level of 10 key genes in cervical tissue cells. (a) Identification of 12 cell clusters based on single-cell RNA sequence data. The UMAP algorithm was applied for dimensionality reduction with the 20 PCs. (b) All 12 clusters of cells in the cervix were annotated by singleR, CellMarker, and PanglaoDB according to the composition of the marker genes. (c) Violin diagrams and UMAP diagrams showed the expression level of genes in 12 clusters.

## Figure 7

Flow diagram of the research process.

## Supplementary Files

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

- [SupplementaryFigures.pdf](#)
- [SupplementaryTables.pdf](#)