Gene expression analysis and identification of key pathways in cervical cancer using RNA-Seq data

Parnian Kaveh[1,] - Dorna Dehghani[2]*
[1] Department of Molecular Cell Biology, Royan Research Institute, Jihad Daneshgahi, Isfahan, Iran
[2] Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran

ABSTRACT

Cervical cancer is one of the most prevalent cancers in women globally and typically predominates in countries that lack sufficient resources. Identifying distinct molecular pathways of the cancer will enable researchers to better understand the pathophysiology of cervical cancer as well as develop targeted therapies. The main goal of the project was to identify differentially expressed genes and any underlying pathways associated with cervical cancer via gene expression analysis.
The genomic expression of cervical cancer RNA-Seq data was acquired from the GEO data and then analyzed. After performing normalization and filtering (logFC>2, p<.05), differentially expressed genes (DEGs) were found. Pathway enrichment analysis was then performed using the KEGG and Gene Ontology databases.
A total of 4667 DEGs were identified. KEGG analysis identified significantly enriched pathways, including ECM-receptor interaction, PI3K-Akt signaling pathway, focal adhesion, cell cycle, and calcium signaling pathway (p.adjust < 0.05). GO analysis found enriched processes such as extracellular matrix organization, muscle processes, and cell adhesion.
Based on the results of this study, changes in ECM, PI3K-Akt signaling, and malignant cell adhesion pathways are significant in the pathology of cervical cancer and may be potential targets for future targeted therapy.

Keywords: Cervical cancer, Gene expression analysis, Biological pathways, Enrichment analysis, ECM, PI3K-Akt

1. *INTRODUCTION*

Cervical cancer is a leading cause of cancer death in women, particularly in developing countries. It is reported that approximately 50% of women worldwide are screened for cervical cancer. The disease is associated with high rates of human papillomavirus (HPV) infection, the leading cause of cervical cancer[1]. Despite significant advances in diagnosis and treatment, the mortality rate of women with cervical cancer remains high in many countries[2].
Recent research has shown that the pathology of cervical cancer is associated with complex molecular alterations, including dysregulation of cell cycle and intracellular signaling and interactions with the extracellular matrix[3]. Gene expression and its related studies have provided new insights to more fully understand the underlying mechanisms of the disease, while also providing targets that reveal the biological pathway associated with cancer progression[4]. The aim of the present study was to investigate the patterns of gene expression changes in patients with cervical cancer and identify key biological pathways associated with disease progression using RNA-Seq data and functional enrichment analyses. This approach may help to discover novel biomarkers and target more effective therapies in the future.

2. Materials and Methods

2.1. Gene Expression Data Collection from TCGA

In this study, gene expression data for cervical cancer (CESC) were obtained from the TCGA (The Cancer Genome Atlas) database. Using the TCGAbiolinks package in the R programming language, data related to "Transcriptome Profiling" and the "Gene Expression Quantification" data type, which were processed with the "STAR - Counts" method, were collected for primary tumor and normal tissue samples. These data were prepared in a suitable format for analysis with the GDCprepare function.

2.2. Differential Expression Analysis

To identify differential genes between tumor and normal samples, the DESeq2 package was used. First, the data were filtered and genes with counts less than 10 were removed. Then, a statistical model was built based on the conditions (Tumor vs Normal) and DESeq analysis was performed. Genes with an adjusted p-value less than 0.05 and a logarithmic fold change (log2FC) greater than 2 were considered as significantly differentially expressed genes (DEGs).

2.3. Volcano plot and gene name mapping

To display the results of the differential analysis, a Volcano plot was used using the EnhancedVolcano package. For greater readability, the Ensembl identifiers of the genes were converted to the official gene name (Gene Symbol). Finally, significant genes were extracted and the output file containing the differential genes was saved.

2.4. Functional Enrichment Analysis

2.4.1. Pathway Analysis (GO Enrichment)

To investigate the biological functions of the differentially expressed genes, Gene Ontology (GO) enrichment analysis was used. This analysis was performed with the clusterProfiler package and the org.. Hs.eg.db database. Only pathways related to Biological Processes (BP) were considered. The results included the significant pathways stored, and dotplot and cnetplot plots were drawn to show the relationship between genes and pathways.

2.4.2. KEGG Pathway Analysis

Next, genes were mapped to Entrez IDs, and KEGG analysis was used to identify molecular pathways. The results were displayed in a readable format (with gene names), and the cnetplot plot showed the significant pathways and the genes involved.

2.5. Predicting Effective Drugs with enrichR

To predict potential targeted drugs, the enrichR package and various drug databases including DrugMatrix, DSigDB, and Drug Perturbations from GEO, were used. The results included possible drugs associated with the observed gene expression patterns. For each database, the results were saved as CSV and Excel files, and a bar plot of the top 10 pathways or drugs was drawn based on the adjusted p-value.

2.6. Tools and software used

All analyses were performed using R software (version 4.4.2) and the following packages: Bioconductor Packages: TCGAbiolinks, DESeq2, SummarizedExperiment, clusterProfiler, DOSE, EnhancedVolcano, ComplexHeatmap, org.Hs.eg.db, enrichplot
CRAN Packages: ggplot2, enrichR, openxlsx

3. Results

In this study, by applying the intensity of gene expression change filters (logFC > 2) and statistical significance level (p-value < 0.05), 4667 differentially expressed genes were identified and used for enrichment analyses of biological pathways related to cervical cancer. The results of KEGG pathway analysis showed that major changes in the pathways regulating cell structure, signaling, and cell cycle play an important role in the pathogenesis of cervical cancer.

The "Cytoskeleton in muscle cells" (hsa04820) pathway with a Fold Enrichment of about 2.6 represents important structural changes in cellular architecture that can play a key role in the migration and invasion of cancer cells. Also, the activation of the "Extracellular matrix interaction with receptors" (hsa04512) pathway, which was accompanied by significant changes in related genes, indicates that changes in the surrounding environment of cancer cells facilitate the processes of migration, invasion, and metastasis.

The calcium signaling pathway (hsa04020) was also observed as a critical pathway in intracellular signaling, with significant activity in cervical cancer samples; this could regulate cancer progression pathways by controlling proliferation and cell death processes. The "local adhesion" pathway (hsa04510) also increases the expression of related genes, increasing the possibility of changes in the ability of cancer cells to attach to adjacent tissues and migrate.

Activation of the cell cycle (hsa04110) and PI3K-Akt signaling (hsa04151) pathways, which have known roles in increasing cell division and preventing apoptosis, represent important regulatory changes in the growth and survival of cervical cancer cells. Also, the activation of the "dilated cardiomyopathy" (hsa05414) pathway may indicate structural changes or secondary cellular stresses in surrounding tissues or concomitant disease contexts that require further investigation

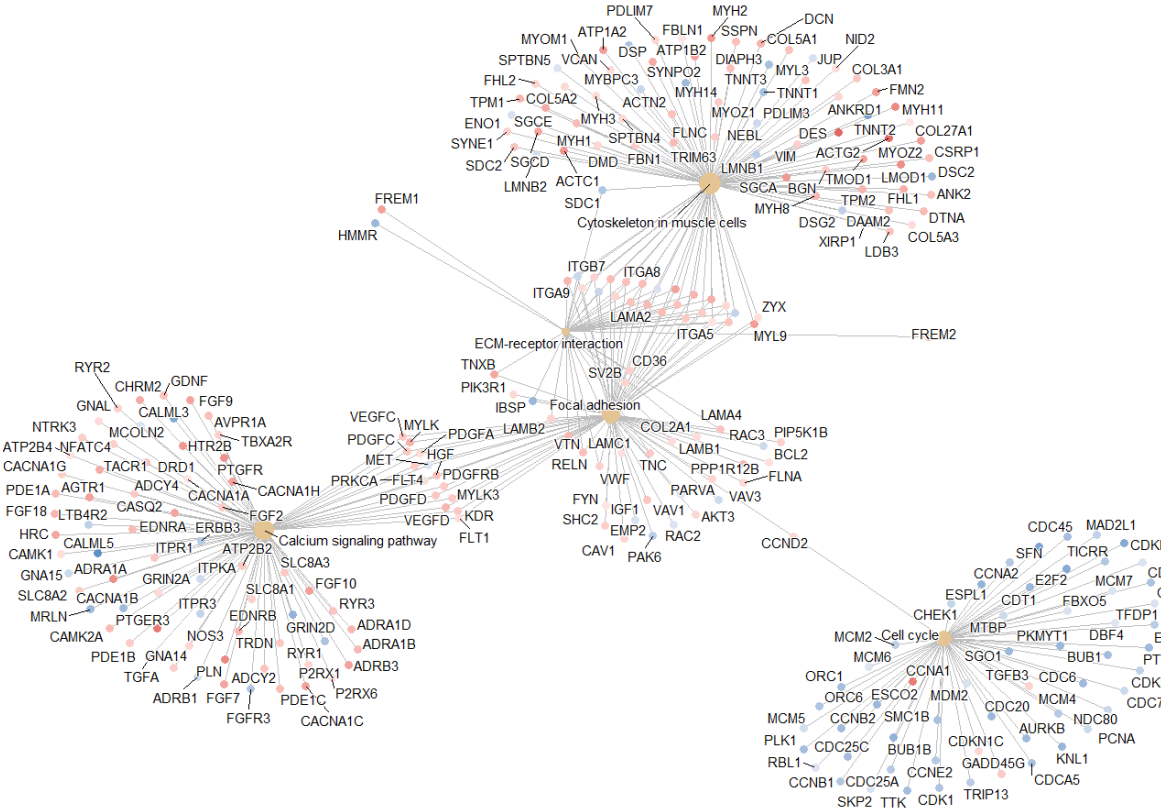

Figure 1:shows gene pathways as a network, where the nodes are genes.

.

Gene Ontology (GO) enrichment analysis also showed a complete alignment with the KEGG data; in particular, pathways related to extracellular matrix organization and cellular structures that play an important role in cervical cancer invasion and progression were highly enriched. Processes related to muscle function and cell adhesion also emphasized the importance of changes in cell interactions and the migratory ability of cancer cells(Figure 2).

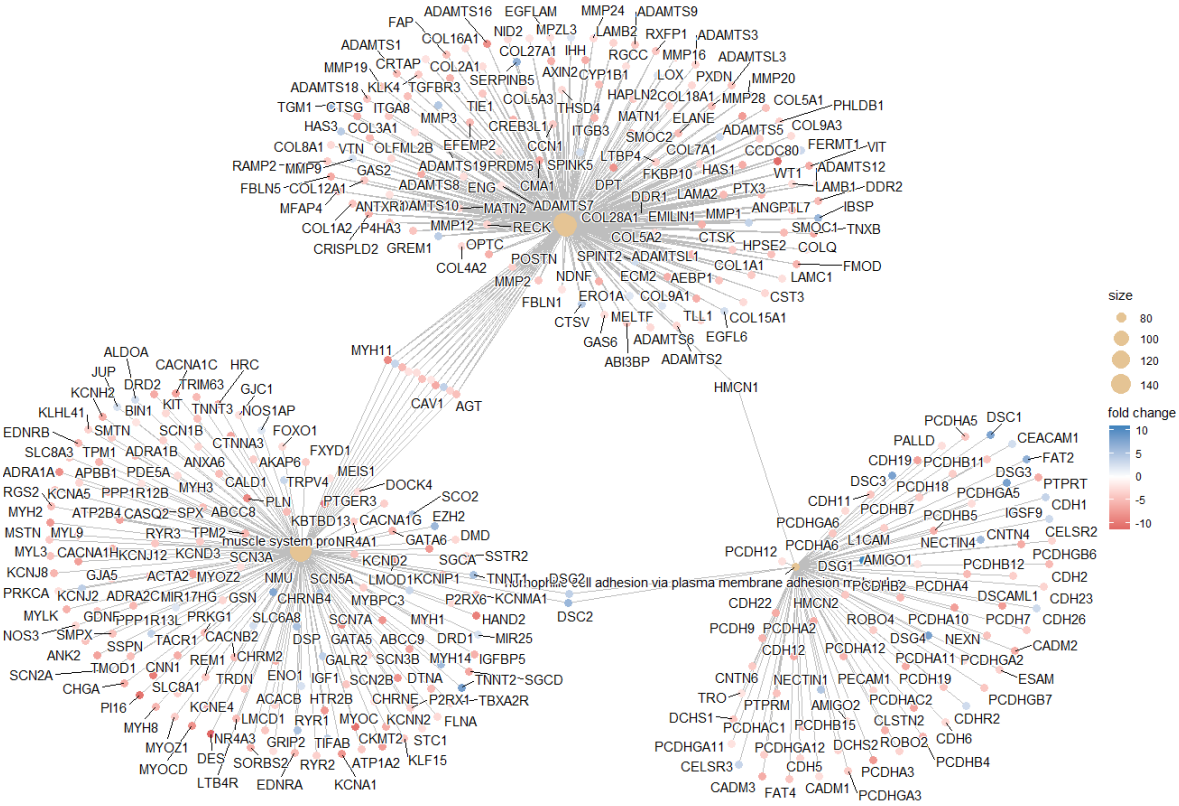

Figure 2 shows the interaction network between genes and biological processes, with orange nodes representing key pathways.

Overall, the results of this study indicate that concurrent disturbances in structural pathways, calcium signaling, cell cycle, and extracellular matrix interactions are key axes in the pathogenesis and progression of cervical cancer. These findings may provide important guidance for the identification of novel biomarkers and the development of targeted therapies in cervical cancer.

4. Discussion

In the current study, we utilized gene expression analysis to investigate significant biological pathways together with key genes involved in cervical cancer which may be relevant in establishing targeted therapeutics. The results demonstrated that pathways involving Extracellular Matrix (ECM) and cellular receptor interactions were significantly upregulated in the cervical cancer specimens. ECM interactions with receptors ultimately lead to the extent of cell migration, invasion and tumor progression, which has been documented in previous studies.

Furthermore, both the PI3K-Akt and cell cycle signaling pathways, which appeared as enriched pathways in this study, have been described as key identified pathways regulating cancer cell growth and survival. When activated these pathways can lead to increased cell proliferation (mitotic division) and apoptosis resistance which can contribute to cancer progression and development of resistance to treatments.

In addition, GO enrichment results showed that cell adhesion processes and cytoskeletal organization were also significantly affected. These findings indicate structural and functional changes in cancer cells that could play an important role in cancer metastasis and spread[4].

In summary, the findings related to these pathways and genes may aid in the identification of potential diagnostic biomarkers and novel targeted treatments for cervical cancer. Future studies should be aimed at understanding the precise role of the genes and pathways identified in this study using functional assays.

5. Conclusion

The results of this study and key biological investigations such as PI3K-Akt signaling, cell cycle, extracellular matrix and cell adhesion processes indicate changes in gene expression in the cervix during cancer. Identification of these changes is helpful in the treatment and prevention of disease progression. In this study, molecular mechanisms were investigated, and understanding these mechanisms is helpful in the recognition and treatment of this disease. It is hoped that by understanding these molecular mechanisms in detail, we will find a way to treat this disease.

6. Reference

[1] E. T. Fontham *et al.*, "Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society," *CA: a cancer journal for clinicians,* vol. 70, no. 5, pp. 321-346, 2020.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians,* vol. 68, no. 6, pp. 394-424, 2018.

[3] S. Das *et al.*, "Molecular mechanisms augmenting resistance to current therapies in clinics among cervical cancer patients," *Medical Oncology,* vol. 40, no. 5, p. 149, 2023.

[4] Y. Yi, Y. Fang, K. Wu, Y. Liu, and W. Zhang, "Comprehensive gene and pathway analysis of cervical cancer progression," *Oncology letters,* vol. 19, no. 4, pp. 3316-3332, 2020.