Identification of Significant Genes Signatures and Prognostic Biomarkers in Cervical Squamous Carcinoma via Bioinformatic Data

Yunan He Equal first author, 1, Shunjie Hu Equal first author, 2, Jiaojiao Zhong 3, Anran Cheng 4, Nianchun Shan Corresp. 5

1 Department of Gynecology and Obstetrics, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China
2 Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, China
3 Department of Dermatology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China
4 Department of Gynecology oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong, China
5 Department of Gynecology and Obstetrics, Xiangya Hospital, Central South University, Changsha, Hunan, China

Corresponding Author: Nianchun Shan
Email address: shannianchun@163.com

Background. Cervical squamous cancer (CESC) is an intractable gynecological malignancy because of its high mortality rate and difficulty in early diagnose. Several biomarkers have been found to predict the prognosis of CESC using bioinformatics methods, but they are still lack of clinical effectiveness. Most of the existing bioinformatic studies only focus on the changes of oncogenes but neglect the differences on protein level and molecular biology validation are rarely conducted. Methods. Gene set data from NCBI-GEO database were used in this study to compare the differences of gene and protein levels between normal and cancer tissues through significant pathway selection and core gene signature analysis to screen potential clinical biomarkers of CESC. Subsequently, the molecular and protein levels of clinical samples were verified by quantitative transcription PCR, Western Blot and immunohistochemistry. Results. Three differentially expressed genes (RFC4, MCM2, TOP2A) were found to have a significant survival ($P<0.05$) and highly expressed in CESC tissues. Molecular biological verification using quantitative reverse transcribed PCR, western blotting and immunohistochemistry assays exhibited significant differences in the expression of RFC4 between CESC and para-cancerous tissues ($P<0.05$). Conclusion. This study identified three potential biomarkers (RFC4, MCM2, TOP2A) of CESC which may be useful to clarify the underlying mechanisms of CESC and predict the prognosis of CESC patients.
Identification of Significant Genes Signatures and Prognostic Biomarkers in Cervical Squamous Carcinoma via Bioinformatic Data

Yunan He¹, Shunjie Hu², Jiaojiao Zhong³, Anran Cheng⁴, Nianchun Shan⁵

¹ Department of Gynecology and Obstetrics, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China
² Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, China
³ Department of Dermatology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China
⁴ Department of Gynecology oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong, China
⁵ Department of Gynecology and Obstetrics, Xiangya Hospital, Central South University, Changsha, Hunan, China

Yunan He and Shunjie Hu contributed equally to this work.

Corresponding Author:
Nianchun Shan⁵
No. 87 Xiangya Road, Changsha, Hunan, 410008, China
Email address: shannianchun@163.com
Abstract

Background. Cervical squamous cancer (CESC) is an intractable gynecological malignancy because of its high mortality rate and difficulty in early diagnose. Several biomarkers have been found to predict the prognosis of CESC using bioinformatics methods, but they are still lack of clinical effectiveness. Most of the existing bioinformatic studies only focus on the changes of oncogenes but neglect the differences on protein level and molecular biology validation are rarely conducted.

Methods. Gene set data from NCBI-GEO database were used in this study to compare the differences of gene and protein levels between normal and cancer tissues through significant pathway selection and core gene signature analysis to screen potential clinical biomarkers of CESC. Subsequently, the molecular and protein levels of clinical samples were verified by quantitative transcription PCR, Western Blot and immunohistochemistry.

Results. Three differentially expressed genes (RFC4, MCM2, TOP2A) were found to have a significant survival ($P<0.05$) and highly expressed in CESC tissues. Molecular biological verification using quantitative reverse transcribed PCR, western blotting and immunohistochemistry assays exhibited significant differences in the expression of RFC4 between CESC and para-cancerous tissues ($P<0.05$).

Conclusion. This study identified three potential biomarkers (RFC4, MCM2, TOP2A) of CESC which may be useful to clarify the underlying mechanisms of CESC and predict the prognosis of CESC patients.

Keywords: cervical squamous carcinoma, bioinformatics, RFC4, prognostic biomarker
Introduction

Cervical cancer now ranks fourth in the most prevalent cancers and it's the most common gynecological cancer in developing countries (Vu et al., 2018). Despite the increase in the incidence of cervical adenocarcinoma, cervical squamous carcinoma (CESC) still stays as the commonest pathological type of cervical cancer (Wang et al., 2004; Galic et al., 2012). Currently, hundreds of gene mutations have been proved to be responsible for increase in the incidence of cervical cancer, which can be used as biomarkers for early detection, like DNA mutations occurring on the oncogenes tumor protein 53 (TP53) (Crook et al., 1992), phosphatase and tensin homolog (PTEN) (2014). However, the overall survival of CESC patients still stays weak due to the difficulties in early detection and treatment. Studies also showed that some biological markers can explain the pathogenesis of CESC and predict the consequences of this disease (Mao et al., 2019). Therefore, more reliable biological markers should be explored to comprehensively understanding the pathogenesis of CESC and guide the treatment and prognosis of that.

With the developed bioinformatics and statistical analyses, the potential marker genes can be detected effectively, which shows great strength in the field of discovery and prediction of tumor markers, and plays a guiding role in the treatment and prognosis of the disease (Banwait & Bastola, 2015). Some biomarkers have been found in the field of cervical cancer, such as MicoRNA-425-5p and MicoRNA-489, which have been proposed for prognostic prediction (Sun et al., 2017; Juan et al., 2018).

However, the presented biomarkers for clinical application are far from enough, and in the previous bioinformatics studies, most studies only focus on the changes of oncogenes, which increases the possibility of clinical inefficacy. On the basis of learning the expression of differential genes between cancer tissues and normal tissues, this study analyzed and compared the difference in protein level between cancer tissue and normal tissue, which provides stronger
evidence for the validity of biomarkers found in our bioinformatic research.

Materials & Methods

Information of the microarray data
NCBI-GEO (Gene Expression Omnibus) is known as a free public database of microarray cohort. The gene profiles of GSE27678, GSE39001 and GSE7803 were obtained in this study. The three datasets were on the account of GPL570 platform, GPL201 platform and GPL96 platform, including 14 normal cervical tissues and 28 CESC tissues, 12 normal cervical tissues and 43 CESC tissues, 10 normal cervical tissues and 21 CESC tissues, respectively.

Identification of differentially expressed genes
The differentially expressed genes (DEGs) were analyzed by GEO2R to obtain the number of up-down-regulated genes (Barrett et al., 2013). The genes with $|\log \text{Fold Change}| \geq 2$ and $P < 0.05$ were screened as differentially expressed genes.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses
Gene Ontology (GO) is an international standardized classification system of gene function, which provides a dynamic updating database to describe the attributes of genes and gene products in organisms (Ashburner et al., 2000). The main biological functions of differentially expressed genes could be determined by GO functional significance enrichment analysis. The GO items with $q < 0.05$ were considered to be significantly enriched in DEGs.

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a bioinformatics resource for linking genomes to life and the environment (Goto et al., 1997). Based on KEGG database, the enriched pathway analysis of DEGs was carried out to find out the important pathway.

PPI & module analysis
Cytoscape 3.8.0 is a software that was used for visualization and analyzation of complex network (Shannon et al., 2003). Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is an application that could conduct protein interaction group research, genome research and proteome research (Doncheva et al., 2019). By mapping the information of DEGs to the STRING, we evaluated the protein-protein interaction (PPI) information of DEGs.
Interactions experimentally validated with combined score $> 0.4$ and were selected.

Subsequently, we used another tool embedded in the Cytoscape named Molecular Complex Detection (MCODE) to clustering constructed functional module of PPI network (Bader & Hogue, 2003). The MCODE scores were set to be greater than 10 and nodes number more than 6. Functional and pathway enrichment for DEGs in the modules were also conducted, $P<0.05$ was considered to have significant difference.

Survival analysis of significant genes in CESC and RNA expression of core genes

Kaplan-Meier (K-M) is a widely used method for estimating the survival rate of cancer patients and “Survival” package was applied in the R studio software (Rich et al., 2010). As for the compare of the magnitude of the difference in survival between the 2 groups, a Cox univariate hazard ratio (HR) was calculated. The clinical significance of each genes was also evaluated by performing the survival analysis of single gene in survival-related gene sets. A log-rank test was used to calculate the statistical significance of the survival difference between these 2 groups mentioned above, and the $P$ value set as 0.05 was considered to be significant.

Gene Expression Profiling Interactive Analysis (GEPIA) is visualization tool for gene research (Tang et al., 2017). In this study, GEPIA was applied to analyze RNA expression of selected genes on the basis of thousands of simples from the TCGA database.

Specimen collection

The tissues or cells of CESC patients were collected from Xiangya Hospital of Central South University in order to verify the high expression of RFC4 in tumor tissues for molecular and protein levels. This study was proved by Medical Ethics Committee of Xiangya Hospital (No.201912542). CESC Patients and the kin have signed a consent form, agreeing to use cervical tissue for scientific research.

Molecular biological verification of differences in gene expression

CESC tissues and para-cancerous tissues (para-CT) were selected from CESC patients to conduct the molecular validation of RFC4. The expression levels of RFC in CESC patients with different pathological stages were also compared. The pathological stage of I and II are
regarded as early stage which including 4 I B1 patients, 7 I B2 patients, 3 I B3 patients, 3 II A1 patients and 1 II A2 patient. Stage III are divided into advanced stage and 17 patients in III C1 stage were included. Total RNA was extracted from CESC tissues and para-CT using Trizol Reagent (RNAiso Plus, TaKaRa, 9109) according to the manufacturer's protocols, and reverse transcribed into cDNA using a PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa, RR047A-1). Gene expression levels were assessed by quantitative reverse transcribed PCR (qRT-PCR) with TB Green™ Premix (Tli RNaseH Plus, TaKaRa, RR820A) and specific primers:

RFC4 forward: 5’-GGCAGCTTTAAGACGTACCATGG-3’;
RFC4 reverse: 5’-TCTGACAGAGGCTTGAAGCGGA-3’.

The β-actin expression was used as the normalization control. Relative mRNA levels are analyzed using 2^{-ΔΔCt} method.

Verification of differences in protein expression

We adopted the cancerous tissues and para-CT of CESC patients to analyze the differences in protein expression by Western Blotting (WB) technology. The samples for WB analysis was separated using SDS-PAGE and transferred onto a PVDF membrane (Roche) which was blocked with 5% nonfat milk in Tris-buffered saline and incubated overnight at 4°C with target antibodies against the following proteins: Anti-RFC4 antibody (ab156780, Abcam) and Anti-β-Actin antibody (ab115777, Abcam). After three times washing with PBST (10 min for each time), the membrane was incubated with species-appropriate HRP-conjugated secondary antibodies, the fluorescent signals were detected using SageCapture™ imaging system (SAGECREATION company).

Immunohistochemistry (IHC) assays were also performed to detected protein levels in CESC tissues and para-CT. The tissues were performed into 5-μm-thick tissue sections with formalin fixed and paraffin embedded. Subsequently, there sections were deparaffinized and rehydrated with xylene and graded ethanol respectively, followed by heated in antigen retrieval.
solution (EDTA, PH 9.0) and endogenous peroxidase inactivation with 3% H₂O₂. After blocking, the samples were incubated overnight at 4°C with anti-RFC4 antibody (1:100, ab156780, Abcam). Then the slides were treated with the HRP-conjugated secondary antibody and stained with 3, 3′-diaminobenzidine until brown granules appeared in the membrane, cytoplasm, or nucleus. Finally, the sections were counterstained with hematoxylin at room temperature.

**Results**

**Screening for DEGs**

Ninety-two cancer tissues and 36 normal tissues were selected from the three datasets in total, with the help of GEO2R tools, 211, 134 and 260 DEGs were extracted from GSE39001, GSE7803 and GSE27678. And Venn diagram was made by the Venn diagram software to investigate the commonly DEGs in all the three datasets. The results showed that there were 25 commonly DEGs in total and 18 of them were down-regulated while 7 were up-regulated (Figure 1 and Table 1).

**Significant pathways identified in CESC**

We investigated upregulated and downregulated DEGs to identify the most significantly enriched pathways in each group by GO and KEGG pathway analysis (Figure 2 and Table2). With GO analyzing, the results indicated that 1) for biology processes (BP), the most significantly enriched pathways of the DEGs were epidermis development, positive regulation of cell proliferation, peptide cross-linking, regulation of cell proliferation, positive regulation of cellular process, epidermal cell differentiation, skin development, keratinocyte differentiation, positive regulation of nuclear division, positive regulation of mitotic nuclear division; 2) for molecular function (MF), they were chemokine activity, chemokine receptor binding, calcium ion binding, collagen binding, CXCR chemokine receptor binding, growth factor activity, intergrin binding, cytokine activity, peptidase activity, acting on L-amino acid peptides, CCR chemokine receptor binding; 3) for cell component (CC), DEGs were significantly enriched in spindle, intercalated disc, intermediate filament, mitotic spindle, nuclear chromosome part, spindle midzone, condensed chromosome kinetochore, platelet alpha granule lumen, spindle...
The results of KEGG analysis demonstrated that the most significant signaling pathways of DEGs were cell cycle, pathways in cancer, ECM-receptor interaction, arrhythmogenic right ventricular cardiomyopathy (ARVC), melanoma, PI3K-Akt signaling pathway, focal adhesion, vascular smooth muscle contraction, DNA replication and oocyte meiosis (Table 3).

**Systematic analysis of core genes by PPI network**

PPI network investigated the systematic interaction between the DEGs we got above. Twenty-five DEGs in total were mapped to the DEGs PPI network with 99 nodes and 270 edges. Seven up-regulated DEGs and 18 down-regulated DEGs were included in the PPI network. And then Cytotype MCODE was applied for further analysis of the DEGs in PPI network, and we got a result of 15 particular nodes being identified which were all up-regulated DEGs (Figure 3).

**Analysis of core gene signature in CESC using K-M plotter and GEPIA**

To investigate the survival data of the 15 genes we identified, K-M plotter indicated that three (TOP2A, RFC4, MCM2) of them had a significant survival rate while other 12 genes had not ($P>0.05$) (Figure 4 and Table4). The expression of TOP2A, RFC4, MCM2 in normal tissue and CESC tissue was detected by GEPIA. The results showed that the expression of these three genes in CESC tissue was higher than that in normal tissue ($P<0.05$) (Figure 5).

**RFC4 is validated to be overexpressed in CESC**

By analyzing the data from the NCBI-GEO dataspace for mRNA expression in CESC patients, RFC4 gene was identified as an overexpressed gene in CESC patients. We collected 35 pairs of CESC patients for qPCR, the tissues of 6 pairs CESC patients were used for WB, 9 pairs CESC tissues and 4 normal cervical tissues for IHC. In order to validate our finding, total RNA was extracted from 35 paired CESC tissues and para-CT tissues, and qRT-PCR was conducted to measure the expression level of RFC4 gene. The result showed that the expression level of RFC4 on CESC tissues was significantly high compared with para-CT ($P=0.0197$) (Figure 6). And the expression of RFC4 in early stage CESC was significantly higher than that in advanced CESC ($P=0.0314$) (Figure 7). The same result was invested from WB. The results of WB analysis
indicated that the RFC4 was overexpressed in CESC tissues compared to para-CT tissues (Figure 8). A higher level of RFC4 expression on CESC tissues was observed from the result of IHC, and RFC4 protein was mainly concentrated in the nucleus (Figure 9).

Discussion

In order to identify more effective prognostic biomarkers in CESC, we used different bioinformatics methods to analyze three data sets based on NCBI-GEO database, including 92 CESC tissues and 36 normal tissues. A total of 25 DEGs were selected by GEO2R and Venn software, including 7 up-regulated genes and 18 down-regulated genes. Then GO and KEGG pathway analysis were conducted, and the results of GO and KEGG indicated that the selected DEGs were significantly enriched in various cell pathways. Researches had reported that some genes from these pathways were associated with the pathogenesis and progression of cervical carcinoma. Nucleolar and spindle associated protein 1(NUSAP1) was a gene from spindle associated pathway, and it was reported to promote the metastasis of cervical cancer by activating Wnt/β-catenin signaling (Li et al., 2019). And studies showed that CXCL12/CXCR4 pathways was associated with HPV infection as a co-factor, which means a high risk to the incidence of cervical cancer (Meuris et al., 2016). Genes involved epidermis development were also associated with the high-risk HPV infection (Zhang et al., 2018; Chatterjee et al., 2019).

After that PPI network was constructed using STRING software and MCODE analysis was conducted, and 15 particular DEGs were identified. Furthermore, by K-M plotter analysis we got 3 DEGs from the 15 which had a significantly better survival. The results of GEPIA showed that the expression levels of the 3 selected genes in CESC tissues were higher than that in normal tissues. To further validation, we performed RFC4 relevant molecule biological experiments and the results showed that compared with normal tissues, RFC4 was highly expressed in CESC tissues.

Being short for Replicant Factor C, RFC is a structure specific DNA-binding protein acting as a primer recognition factor for DNA polymerase (Zhou & Hingorani, 2012), which includes 5 subunits (RFC1-5). Among all 5 subunits of RFC complex, RFC4 has been reported to play an
important role in DNA damage checkpoint and DNA replication pathways (Ellison & Stillman, 2003). In 2009, Arai M et al. reported that RFC4 was closely related to the prognosis of liver cancer (Arai et al., 2009). Besides liver cancer, RFC4 has been reported to be associated with several types of cancer, including prostate cancer, colon cancer non-small cell lung cancer and leukemia (LaTulippe et al., 2002; Jung et al., 2009; Erdogan et al., 2009; Barfeld et al., 2014). There were researches illustrated that up-regulated RFC4 expression was found in neck squamous cell carcinoma and it was 3.4-fold higher than that in normal tissues (Slebos et al., 2006). Studies from Garnett et al. showed that RFC4 can be regulated by mutated RB1 in several types of cancers, suggesting that RFC4 could be a potential biomarker associated with the occurrence and prognosis of various cancers (Garnett et al., 2012). Moreover, RFC4 was reported as an independent predictor of overall survival in breast cancer (Fatima et al., 2017; Niu et al., 2017).

In this study we observed RFC4 as a potential independent prognostic biomarker in CESC, and our results suggested that CESC patients with higher expression level of RFC4 may have a better overall survival. A possible reason might be that RFC4 was highly expressed throughout the cell circle process of proliferating cells, and tumor proliferation in situ will become slow with the development of the disease (Szymanska et al., 2018; Chaplain & Sleeman, 1993), which means a decrease in the expression of RFC4. Therefore, highly expressed RFC4 may suggest early stage CESC, which indicates better overall survival.

Several studies have proved that these three genes were associated with numerous types of cancer, but studies of RFC4 in CESC were rarely seen, and very few researches conducted molecule biology validation. Therefore, our study shows that RFC4 is a potential biomarker for the predicting the prognosis of CESC and provides a direction for further study of CESC. What should be noted is that there are some limitations in this study. Clinical samples from one hospital may have either region or race difference. The expression level of RFC4 in different stages of CESC and clinical investigations should be conducted in our future study to validate our results further.
Conclusions

In conclusion, by using bioinformatics analysis we identified three genes (TOP2A, RFC4, MCM2) based on three microarray datasets. These three genes were suggested to have a significant effect on the prognosis of CESC, which could be key factors in the occurrence and progression of CESC. And a high level expressed RFC4 was validated to be existed in CESC tissues using clinical samples. Although further investigation and experiments need conducting, the findings in our study could act as clinical biomarkers which would be helpful for us to better understand the pathological process and predict the prognostic of CESC.

Acknowledgments

The authors appreciate the patients who provided tumor tissue for this study. We also thank to Xiaofan Li for English editing.

References:

2014. Analysis identifies mutations linked to cervical cancer. Cancer Discovery 4:F2. 10.1158/2159-8290.CD-NB2014-010
Arai, M., Kondoh, N., Imazeki, N., Hada, A., Hatsuse, K., Matsubara, O., and Yamamoto, M. 2009. The knockdown of endogenous replication factor C4 decreases the growth and enhances the chemosensitivity of hepatocellular carcinoma cells. LIVER INTERNATIONAL 29:55-62. 10.1111/j.1478-3231.2008.01792.x
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. NATURE GENETICS 25:25-29. 10.1038/75556
Bader, G.D., and Hogue, C.W. 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC BIOINFORMATICS 4:2. 10.1186/1471-2105-4-2
Banwait, J.K., and Bastola, D.R. 2015. Contribution of bioinformatics prediction in microRNA-based cancer therapeutics. Adv Drug Deliv Rev 81:94-103. 10.1016/j.addr.2014.10.030
Barfeld, S.J., East, P., Zuber, V., and Mills, I.G. 2014. Meta-analysis of prostate cancer gene expression data identifies a novel discriminatory signature enriched for glycosylating enzymes. BMC Medical Genomics 7:513. 10.1186/s12920-014-0074-9
Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippey, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., and Soboleva, A. 2013. NCBI GEO: archive for functional genomics data sets--update. NUCLEIC ACIDS RESEARCH 41:D991-D995. 10.1093/nar/gks1193
Chaplain, M.A., and Sleeman, B.D. 1993. Modelling the growth of solid tumours and incorporating a method for their classification using nonlinear elasticity theory. JOURNAL OF MATHEMATICAL BIOLOGY 31:431-473. 10.1007/bf00173886
Chatterjee, S., Do, K.S., Alam, S., Salzberg, A.C., Milici, J., van der Burg, S.H., Freeman, W., and Meyers, C. 2019. Tissue-Specific Gene Expression during Productive Human Papillomavirus 16 Infection of Cervical, Foreskin, and Tonsil Epithelium. *Journal of Virology* 93. 10.1128/JVI.00915-19

Crook, T., Wrede, D., Tidy, J.A., Mason, W.P., Evans, D.J., and Vousden, K.H. 1992. Clonal p53 mutation in primary cervical cancer: association with human-papillomavirus-negative tumours. *Lancet* 339:1070-1073. 10.1016/0140-6736(92)90662-m

Doncheva, N.T., Morris, J.H., Gorodkin, J., and Jensen, L.J. 2019. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research* 18:623-632. 10.1021/acs.jproteome.8b00702

Ellison, V., and Stillman, B. 2003. Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5' recessed DNA. *Plos Biology* 1:E33. 10.1371/journal.pbio.0000033

Erdogan, E., Klee, E.W., Thompson, E.A., and Fields, A.P. 2009. Meta-analysis of oncogenic protein kinase C iota signaling in lung adenocarcinoma. *Clinical Cancer Research* 15:1527-1533. 10.1158/1078-0432.CCR-08-2459

Fatima, A., Tariq, F., Malik, M., Qasim, M., and Haq, F. 2017. Copy Number Profiling of MammaPrint Genes Reveals Association with the Prognosis of Breast Cancer Patients. *Journal of Breast Cancer* 20:246-253. 10.4048/jbc.2017.20.3.246

Galic, V., Herzog, T.J., Lewin, S.N., Neugut, A.I., Burke, W.M., Lu, Y.S., Hershman, D.L., and Wright, J.D. 2012. Prognostic significance of adenocarcinoma histology in women with cervical cancer. *Gynecologic Oncology* 125:287-291. 10.1016/j.ygyno.2012.01.012

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., Liu, Q., Iorio, F., Surdez, D., Chen, L., Milano, R.J., Bignell, G.R., Tam, A.T., Davies, H., Stevenson, J.A., Barthorpe, S., Lutz, S.R., Kogera, F., Lawrence, K., McLaren-Douglas, A., Mitropoulos, X., Mironenko, T., Thi, H., Richardson, L., Zhou, W., Jewitt, F., Zhang, T., O'Brien, P., Boisvert, J.L., Price, S., Hur, W., Yang, W., Deng, X., Butler, A., Choi, H.G., Chang, J.W., Baselga, J., Stamenkovic, I., Engelman, J.A., Sharma, S.V., Delattre, O., Saez-Rodriguez, J., Gray, N.S., Settleman, J., Futreal, P.A., Haber, D.A., Stratton, M.R., Ramaswamy, S., McDermott, U., and Benes, C.H. 2012. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483:570-575. 10.1038/nature11005

Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishio, T., Sato, K., and Kanehisa, M. 1997. Organizing and computing metabolic pathway data in terms of binary relations. *Pac Symp Biocomput*;175-186.

Juan, C., Hua, Q., Ruping, Z., and Tingting, W. 2018. miRNA-489 as a biomarker in diagnosis and treatment of cervical cancer. *Bratisl Lek Listy* 119:278-283. 10.4149/BLL_2018_052

Jung, H.M., Choi, S.J., and Kim, J.K. 2009. Expression profiles of SV40-immortalization-associated genes upregulated in various human cancers. *Journal of Cellular Biochemistry* 106:703-713. 10.1002/jcb.22063

Kahlem, P., and Newfeld, S.J. 2009. Informatics approaches to understanding TGFbeta pathway regulation. *Development* 136:3729-3740. 10.1242/dev.030320

LaTulippe, E., Satagopan, J., Smith, A., Scher, H., Scardino, P., Reuter, V., and Gerald, W.L. 2002. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Research* 62:4499-4506.
Li, H., Zhang, W., Yan, M., Qiu, J., Chen, J., Sun, X., Chen, X., Song, L., and Zhang, Y. 2019. Nucleolar and spindle associated protein 1 promotes metastasis of cervical carcinoma cells by activating Wnt/beta-catenin signaling. *J Exp Clin Cancer Res* 38:33. 10.1186/s13046-019-1037-y

Mao, Y., Fu, Z., Dong, L., Zheng, Y., Dong, J., and Li, X. 2019. Identification of a 26-lncRNAs Risk Model for Predicting Overall Survival of Cervical Squamous Cell Carcinoma Based on Integrated Bioinformatics Analysis. *DNA AND CELL BIOLOGY* 38:322-332. 10.1089/dna.2018.4533

Meuris, F., Carthagena, L., Jaracz-Ros, A., Gaudin, F., Cutolo, P., Deback, C., Xue, Y., Thierry, F., Doorbar, J., and Bachelerie, F. 2016. The CXCL12/CXCR4 Signaling Pathway: A New Susceptibility Factor in Human Papillomavirus Pathogenesis. *PLoS Pathogens* 12:e1006039. 10.1371/journal.ppat.1006039

Niu, G., Wang, D., Pei, Y., and Sun, L. 2017. Systematic identification of key genes and pathways in the development of invasive cervical cancer. *GENE* 618:28-41. 10.1016/j.gene.2017.03.018

Rich, J.T., Neely, J.G., Paniello, R.C., Voelker, C.C., Nussenbaum, B., and Wang, E.W. 2010. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* 143:331-336. 10.1016/j.otohns.2010.05.007

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *GENOME RESEARCH* 13:2498-2504. 10.1101/gr.1239303

Slebos, R.J., Yi, Y., Ely, K., Carter, J., Evjen, A., Zhang, X., Shyr, Y., Murphy, B.M., Cmelak, A.J., Burkey, B.B., Nettervile, J.L., Levy, S., Yarbrough, W.G., and Chung, C.H. 2006. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *CLINICAL CANCER RESEARCH* 12:701-709. 10.1158/1078-0432.CCR-05-0247

Sun, L., Jiang, R., Li, J., Wang, B., Ma, C., Lv, Y., and Mu, N. 2017. MicoRNA-425-5p is a potential prognostic biomarker for cervical cancer. *ANNALS OF CLINICAL BIOCHEMISTRY* 54:127-133. 10.1177/0004563216649377

Szymanska, Z., Cytowski, M., Macnamara, C.K., and Chaplain, M. 2018. Computational Modelling of Cancer Development and Growth: Modelling at Multiple Scales and Multiscale Modelling. *Bull Math Biol* 80:1366-1403. 10.1007/s11538-017-0292-3

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *NUCLEIC ACIDS RESEARCH* 45:W98-W102. 10.1093/nar/gkx247

Vu, M., Yu, J., Awolude, O.A., and Chuang, L. 2018. Cervical cancer worldwide. *Curr Probl Cancer* 42:457-465. 10.1016/j.curpprobcan.2018.06.003

Wang, S.S., Sherman, M.E., Hildesheim, A., Lacey, J.J., and Devesa, S. 2004. Cervical adenocarcinoma and squamous cell carcinoma incidence trends among white women and black women in the United States for 1976-2000. *CANCER* 100:1035-1044. 10.1002/cncr.20064

Zhang, J., Burn, C., Young, K., Wilson, M., Ly, K., Budhwani, M., Tscharmier, A., Braithwaite, A., Baird, M., and Hibma, M. 2018. Microparticles produced by human papillomavirus type 16 E7-expressing cells impair antigen presenting cell function and the cytotoxic T cell response. *Sci Rep* 8:2373. 10.1038/s41598-018-20779-2

Zhou, Y., and Hingorani, M.M. 2012. Impact of individual proliferating cell nuclear antigen-DNA contacts on clamp loading and function on DNA. *JOURNAL OF BIOLOGICAL CHEMISTRY* 287:35370-35381. 10.1074/jbc.M112.399071
Figure 1

Identification of 25 common DEGs in the three datasets (GSE39001, GSE7803 and GSE27678) through Venn diagrams software.

Different color meant different datasets. **A.** Seven DEGs were up-regulated in the three datasets (logFC > 2). **B.** Eighteen DEGs were down-regulated in three datasets (logFC < -2).
Figure 2

GO and KEGG results shows significant signaling pathways of DEGs.
Figure 3

Common DEGs PPI network constructed by STRING online database and Module analysis.

A. Nodes meant proteins; the edges meant the interaction of proteins. B. Module analysis via Cytoscape MCODE tool (degree cutoff =2, node score cutoff =0.2, k-core =2, and max. Depth =100).
Figure 4

The prognostic information of the 15 core genes.

Three (A, B and C) of 15 genes had a significant better survival rate ($P < 0.05$) and twelve genes (D-O) had not significant difference in OS ($P > 0.05$).
Figure 5

Expression level of 3 significantly expressed genes in CESC tissues and normal tissues.

All the 3 genes had significant expression level in CESC specimen compared to normal specimen ($P < 0.05$). Red color means tumor tissues and grey means normal tissues.
Figure 6

The expression of RFC4 on CESC was significantly different compared with para-cancerous tissues from the result of qRT-PCR.
$2^{\Delta \Delta Ct}$

P = 0.0197

CELSC  Para-CT

RFC4
Figure 7

Expression levels of RFC4 in different pathological stages of CESC

The expression of RFC4 in early stage CESC was significantly higher than that in advanced CESC (P=0.0314)
\[ P = 0.0314 \]

![Bar chart showing comparison of \(2^{\Delta \Delta Ct}\) values between Early Stage (I, II) and Advanced Stage (III) for RFC4.](image)
Figure 8

WB analysis of RFC4 protein.

C: CESC tissues, P: para-cancerous tissues. **A.** Six pairs CESC tissues WB analysis indicated that except that the results of case 4 are not obvious, the others are consistent with the expected results of high expression of RFC4 in tumor tissues. **B.** The grayscale analysis of multiple WB bands shows that the WB tests are reliable.
Figure 9

IHC test of CESC.

IHC declared that in general, RFC4 protein is highly expressed in tumor tissue sections, and is mainly concentrated in the nucleus, while normal cervical tissue and para-cancerous tissues are underexpressed.
| Manuscript to be reviewed |
|---------------------------|

| CESC tissues | Para-CT tissues |
|-------------|-----------------|
| ![Image](image1.png) | ![Image](image2.png) |
| ![Image](image3.png) | ![Image](image4.png) |
| ![Image](image5.png) | ![Image](image6.png) |
| ![Image](image7.png) | ![Image](image8.png) |
| ![Image](image9.png) | ![Image](image10.png) |
| ![Image](image11.png) | ![Image](image12.png) |

| Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6 | Case 7 | Case 8 | Case 9 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| ![Image](image13.png) | ![Image](image14.png) | ![Image](image15.png) | ![Image](image16.png) | ![Image](image17.png) | ![Image](image18.png) | ![Image](image19.png) | ![Image](image20.png) | ![Image](image21.png) |
| ![Image](image22.png) | ![Image](image23.png) | ![Image](image24.png) | ![Image](image25.png) | ![Image](image26.png) | ![Image](image27.png) | ![Image](image28.png) | ![Image](image29.png) | ![Image](image30.png) |
| ![Image](image31.png) | ![Image](image32.png) | ![Image](image33.png) | ![Image](image34.png) | ![Image](image35.png) | ![Image](image36.png) | ![Image](image37.png) | ![Image](image38.png) | ![Image](image39.png) |

Normal Cervical Tissues (200x)
Table 1 (on next page)

25 common DEGs identified from the three datasets
Table 1. 25 common DEGs identified from the three datasets

| Expression | Genes Name         |
|------------|--------------------|
| Up-regulated | M2M DTL CDKN2A TOP2A NUSAP1 RFC4 PLOD2 |
| Down-regulated | EMP1 IGF1 ALOX12 EDN3 PTGDS KRT1 FOSB GREB1 ESR1 PAMR1 CXCL12 HPGD AR MAL CRNN CRISP3 CFD NDN |
Table 2 (on next page)

GO analysis of different expressed genes in CESC
Table 2. GO analysis of different expressed genes in CESC

| Expression | Category          | Term                                      | Count | %  | p-Value   | FDR       |
|------------|-------------------|-------------------------------------------|-------|----|-----------|-----------|
| Up-regulated | GOTERM_MF_DIREC | GO:0005524~ATP binding                   | 10    | 21.0 | 2.84E-4 | 0.270953 |
|            |                   | GO:0003678~DNA helicase activity          | 2     | 4.21 | 0.01233 | 11.169161 |
|            |                   | GO:0003688~DNA replication origin binding | 2     | 4.21 | 0.01844 | 16.278756 |
|            |                   | GO:0003682~chromatin binding              | 4     | 8.41 | 0.02152 | 18.752384 |
|            | GOTERM_CC_DIREC   | GO:0030496~midbody                        | 6     | 12.6 | 2.68E-7 | 2.65E-4  |
|            |                   | GO:0005737~cytoplasm                      | 14    | 29.4 | 1.50E-4 | 0.147574 |
|            |                   | GO:0005876~spindle microtubule            | 3     | 6.31 | 0.00148 | 1.457975 |
|            |                   | GO:0005654~nucleoplasm                    | 8     | 16.8 | 0.00555 | 5.351061 |
|            |                   | GO:0000784~nuclear chromosome, telomeric region | 3   | 6.31 | 0.06309 | 9.946422 |
|            |                   | GO:0072687~meiotic spindle                | 2     | 4.21 | 0.01428 | 13.241928 |
|            |                   | GO:0042555~MCM complex                    | 2     | 4.21 | 0.01428 | 13.241928 |
| T | GOTERM_CC_DIREC | GO:0005680~anaphase-promoting complex | 2 | 4.21 | 0.03533 | 29.901598 |
|---|----------------|--------------------------------------|---|------|---------|------------|
| T | GOTERM_CC_DIREC | GO:0005634~nucleus                   | 9 | 18.9 | 0.04947 | 39.406713 |
| T | GOTERM_CC_DIREC | GO:0072686~mitotic spindle           | 2 | 4.21 | 0.05426 | 42.356169 |
| T | GOTERM_CC_DIREC | GO:0005819~spindle                   | 2 | 4.21 | 0.06274 | 47.263108 |
| T | GOTERM_CC_DIREC | GO:000776~kinetochore                | 2 | 4.21 | 0.09268 | 61.726231 |
| T | GOTERM_BP_DIREC | GO:0000910~cytokinesis               | 3 | 6.31 | 9.78E-4 | 1.154575  |
| T | GOTERM_BP_DIREC | GO:0044772~mitotic cell cycle phase transition | 2 | 4.21 | 0.00777 | 8.844187  |
| T | GOTERM_BP_DIREC | GO:0051988~regulation of attachment of spindle microtubules to kinetochore | 2 | 4.21 | 0.01163 | 12.969466 |
| T | GOTERM_BP_DIREC | GO:0031145~anaphase-promoting complex-dependent catabolic process | 2 | 4.21 | 0.01355 | 14.961799 |
| T | GOTERM_BP_DIREC | GO:0006268~DNA unwinding involved in DNA replication | 2 | 4.21 | 0.01548 | 16.908679 |
| T | GOTERM_BP_DIREC | GO:0007095~mitotic G2 DNA damage checkpoint | 2 | 4.21 | 0.01931 | 20.670192 |
| T | GOTERM_BP_DIREC | GO:0007076~mitotic chromosome condensation | 2 | 4.21 | 0.02122 | 22.486821 |
| GOTERM_BP_DIREC | GO:0000070~mitotic sister chromatid segregation | 2 | 4.21 | 0.03451 | 34.093744 |
|----------------|-----------------------------------------------|---|------|---------|------------|
|                | GO:0001578~microtubule bundle formation        | 2 | 4.21 | 0.03827 | 37.079667 |
|                | GO:0000281~mitotic cytokinesis                 | 2 | 4.21 | 0.04015 | 38.521683 |
|                | GO:0006270~DNA replication initiation          | 2 | 4.21 | 0.04015 | 38.521683 |
|                | GO:0005198~structural molecule activity        | 9 | 6.88 | 5.12E-6 | 4.557253 |
|                | GO:0004252~serine-type endopeptidase activity  | 7 | 5.35 | 1.02E-4 | 0.004512 |
|                | GO:0008201~heparin binding                     | 6 | 4.59 | 1.73E-4 | 0.187501 |
|                | GO:0005496~steroid binding                     | 3 | 2.29 | 0.00363 | 3.880030 |
|                | GO:0004962~endothelin receptor activity        | 2 | 1.53 | 0.01529 | 11.655613 |
|                | GO:0003707~steroid hormone receptor activity   | 3 | 2.29 | 0.02293 | 31.330301 |
|                | GO:0005615~extracellular space                 | 25 | 19.1 | 7.92E-11 | 7.97E-8 |
|                | GO:0070062~extracellular exosome               | 39 | 29.8 | 1.45E-10 | 4.43E-9 |
|                | GO:0005576~extracellular region                | 12 | 9.18 | 2.40E-6 | 0.002416 |
| GOTERM_CC_DIREC | GO:0001533~cornified envelope | 5   | 3.82 | 4.01E-5 | 0.040335 |
|                 | GO:0005578~proteinaceous extracellular matrix | 7   | 5.35 | 2.35E-4 | 0.236029 |
|                 | GO:0045095~keratin filament          | 4   | 3.06 | 0.0218  | 2.179044 |
|                 | GO:0042567~insulin-like growth factor ternary complex | 2   | 1.53 | 0.02333 | 21.135448 |
|                 | GO:0031012~extracellular matrix      | 4   | 3.06 | 0.02398 | 21.661062 |
|                 | GO:0001527~microfibril               | 2   | 1.53 | 0.03479 | 29.965502 |
|                 | GO:0042581~specific granule          | 2   | 1.53 | 0.06287 | 47.957744 |
|                 | GO:0016323~basolateral plasma membrane | 3   | 2.29 | 0.08757 | 60.215342 |
| GOTERM_BP_DIREC | GO:0018149~peptide cross-linking     | 5   | 3.82 | 6.99E-4 | 0.098265 |
|                 | GO:0030216~keratinocyte differentiation | 5   | 3.82 | 3.21E-4 | 0.451079 |
|                 | GO:0007565~female pregnancy          | 3   | 2.29 | 0.00163 | 2.274249 |
|                 | GO:0008284~positive regulation of cell proliferation | 7   | 5.35 | 0.00241 | 3.344514 |
|                 | GO:0045840~positive regulation of mitotic nuclear division | 3   | 2.29 | 0.00443 | 6.057636 |
| GOTERM_BP_DIRECT | GO:0048146~positive regulation of fibroblast proliferation | 3 | 2.29 | 0.01535 | 19.550011 |
|------------------|----------------------------------------------------------|---|------|---------|------------|
| GO:0006955~immune response | 5 | 3.82 | 0.01676 | 21.157726 |
| GO:0001558~regulation of cell growth | 3 | 2.29 | 0.01803 | 22.573586 |
| GO:0001755~neural crest cell migration | 3 | 2.29 | 0.01896 | 23.602523 |
| GO:0014826~vein smooth muscle contraction | 2 | 1.53 | 0.02213 | 27.006079 |
| GO:0001775~cell activation | 2 | 1.53 | 0.03302 | 37.638886 |
| GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling | 3 | 2.29 | 0.03403 | 38.544070 |
| GO:007267~cell-cell signaling | 3 | 2.29 | 0.04283 | 45.968556 |
| GO:0021952~central nervous system projection neuron axonogenesis | 2 | 1.53 | 0.04379 | 46.724165 |
| GO:0030198~extracellular matrix organization | 3 | 2.29 | 0.04820 | 50.079668 |
| GO:0048484~enteric nervous system development | 2 | 1.53 | 0.04913 | 50.758144 |
| GO:0043568~positive regulation of insulin-like growth factor receptor signaling pathway | 2 | 1.53 | 0.04913 | 50.758144 |
| GO:0005978~glycogen biosynthetic process | 2 | 1.53 | 0.07539 | 66.786495 |
| Term                                                                 | ID                  | Count | P.Value | Adjust P.Value |
|---------------------------------------------------------------------|---------------------|-------|---------|----------------|
| regulation of pH                                                    | GO:0006885          | 2     | 1.53    | 0.07539        |
| negative regulation of endothelial cell migration                   | GO:0010596          | 2     | 1.53    | 0.07539        |
| retinal ganglion cell axon guidance                                 | GO:0031290          | 2     | 1.53    | 0.08055        |
| regulation of glucose metabolic process                             | GO:0010906          | 2     | 1.53    | 0.09080        |
| negative regulation of smooth muscle cell proliferation             | GO:0048662          | 2     | 1.53    | 0.09588        |
| axon extension                                                       | GO:0048675          | 2     | 1.53    | 0.09588        |
Table 3

KEGG analysis of DEGs in CESC
| Pathway ID | Name                                      | Count | p-Value    | Genes                                                                 |
|-----------|-------------------------------------------|-------|------------|-----------------------------------------------------------------------|
| 04110     | Cell cycle                                | 13    | 7.76E-6    | PCNA, CDKN2A, BUB1B, CDC7, TTK, SMC1B, CDC20, CCNB1, PTTG1, CDK1, CM4, MCM5, MCM2 |
| 05200     | Pathways in cancer                        | 29    | 2.77E-5    | LAMA2, CKS1B, FGF7, EDNRA, EDNRB, RUNX1T1, PDGFRB, PDGFRB, JUP, CDKN2A, MMP1, ITGA2, PTCH1, FN1, IGF2, MITF, FOS, IGF1, WNT16, GNG11, ESR1, AR, CXCL12, GSTA4, CKS2, BIRC5, FGFR2, GSTM5, FGFR2 |
| 04512     | ECM-receptor interaction                  | 9     | 1.36E-4    | TNXB, VWF, LAMA2, ITGA2, ITGA8, SPP1, FN1, HMMR, ITGA9               |
| 05412     | Arrhythmogenic right ventricular cardiomyopathy (ARVC) | 8     | 2.93E-4    | GJA1, LAMA2, JUP, ITGA2, ITGA8, DSG2, DSC2, ITGA9                     |
| 05218     | Melanoma                                  | 8     | 2.93E-4    | PDGFRB, PDGFRB, FGF7, CDKN2A, PDGFD, MITF, IGF1, FGF10               |
| 04151     | PI3K-Akt signaling pathway                | 20    | 3.18E-4    | PDGFRB, PDGFRB, TNXB, VWF, LAMA2, ITGA2, IGF2, FN1, IGF1, GNG11, AREG, EREG, GYS2, FGF7, PDGFD, SPP1, ITGA8, FGFR2, FGFR2, ITGA9 |
| 04510     | Focal adhesion                            | 13    | 9.45E-4    | PDGFRB, PDGFRB, TNXB, VWF, LAMA2, ITGA2, FN1, IGF1, MYLK, PDGFD, SPP1, ITGA8, ITGA9 |
| 04270     | Vascular smooth muscle contraction        | 10    | 0.0011     | ACTA2, GUCA1A2, PPP1R14A, EDNRA, EDN3, MYH11, MRVI1, AVPR1A, ACTG2, MYLK |
Table 3. KEGG analysis of DEGs in CESC

| KEGG ID | Pathway                                      | Count | p-value | Genes                                    |
|---------|----------------------------------------------|-------|---------|------------------------------------------|
| 03030   | DNA replication                              | 5     | 0.0014  | PCNA, RFC4, MCM4, MCM5, MCM2             |
| 04114   | Oocyte meiosis                               | 9     | 0.0029  | CDC20, AR, CCNB1, PTTG1, CDK1, PGR, IGF1, SMC1B, AURKA |

*Manuscript to be reviewed*
**Table 4** (on next page)

The information of prognostic analysis of 15 core DEGs
Table 4. The information of prognostic analysis of 15 core DEGs

| Category                                      | Genes                                      |
|-----------------------------------------------|--------------------------------------------|
| Genes with significant (better) survival (P<0.05) | TOP2A RFC4 MCM2                            |
| Genes without significant survival (P>0.05)    | UBE2C PRC1 NUSAP1 NEK2 MCM5 KIF20A HMMR FANCI |
|                                               | ECT2 DTL AURKA ASPM                        |