



Identification of significant genes signatures and prognostic biomarkers in cervical squamous carcinoma via bioinformatic data

Yunan He^{1,*}, Shunjie Hu^{2,*}, Jiaojiao Zhong³, Anran Cheng⁴ and 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

* These authors contributed equally to this work.

ABSTRACT

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

Methods. Gene set data from the 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.

Submitted 28 July 2020
Accepted 27 October 2020
Published 2 December 2020

Corresponding author
Nianchun Shan,
shannianchun@163.com

Academic editor
Jianye Ge

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.10386

© Copyright
2020 He et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Gynecology and Obstetrics, Oncology, Women's Health, Medical Genetics

Keywords Cervical squamous carcinoma, Bioinformatics, RFC4, Prognostic biomarker

INTRODUCTION

Cervical cancer now ranks fourth in the most prevalent cancers and it is 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) is still the most common type of cervical cancer (Wang *et al.*, 2004; Galic *et al.*, 2012). Currently, a large number of gene mutations have been proved to be related to the pathogenesis 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) (Yang *et al.*, 2015). However, due to the difficulties of early detection and diagnosis, the survival rate of CESC patients still stays weak. 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 treatment and prognosis.

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.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database is a bioinformatics resource for linking genomes to life and the environment (Kanehisa *et al.*, 2017). Based on the 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 samples 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 $2^{-\Delta\Delta C_t}$ 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- μ m-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.

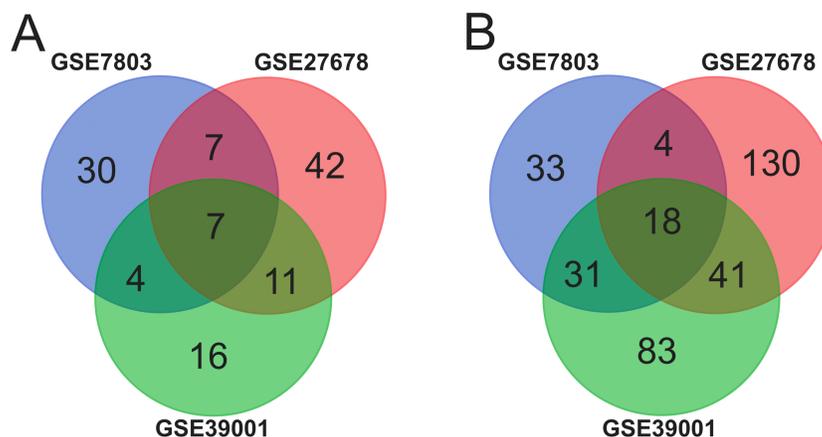


Figure 1 Identification of 25 common DEGs in the three datasets (GSE39001, GSE7803 and GSE27678) through Venn diagrams software. Different color meant different datasets. (A) Seven DEGs were up-regulated in the three datasets ($\log_{FC} > 2$). (B). Eighteen DEGs were down-regulated in three datasets ($\log_{FC} > -2$).

Full-size DOI: 10.7717/peerj.10386/fig-1

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

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 (Fig. 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 (Fig. 2 and Table 2). 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,

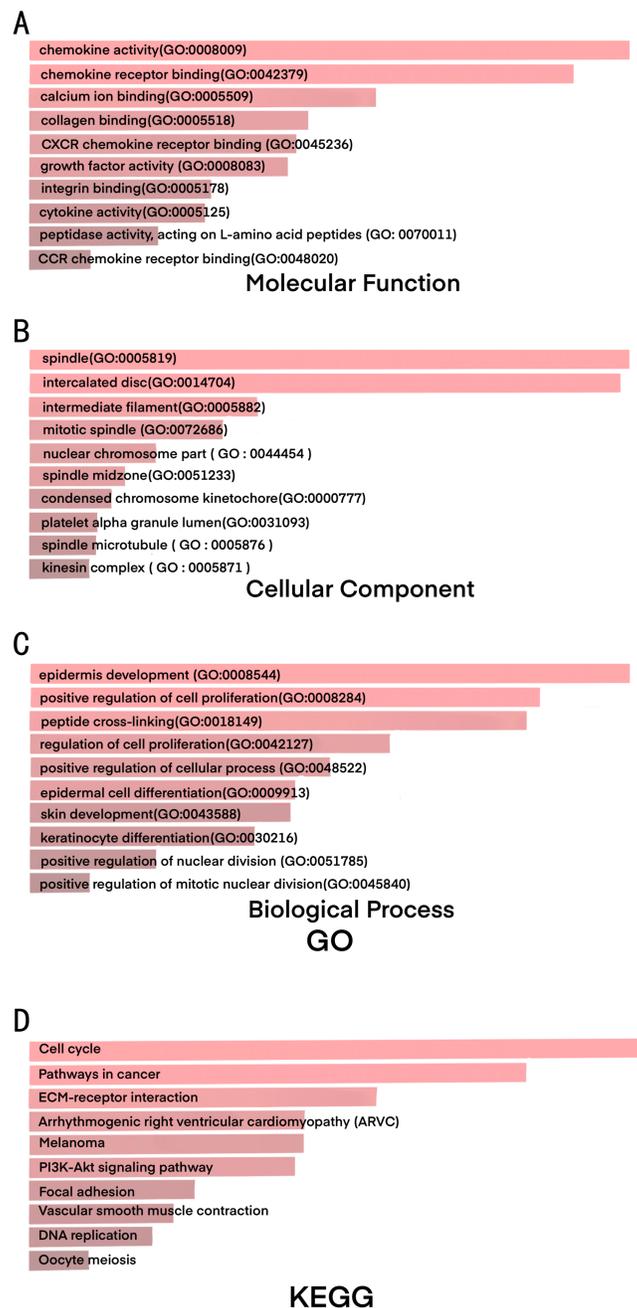


Figure 2 GO and KEGG results shows significant signaling pathways of DEGs. (A) The results of GO analysis for pathways associated with molecular function (MF). (B) The results of GO analysis for pathways associated with cellular component (CC). (C) The results of GO analysis for pathways associated with biological process (BP). (D) The results of KEGG analysis.

Full-size DOI: [10.7717/peerj.10386/fig-2](https://doi.org/10.7717/peerj.10386/fig-2)

chemokine receptor binding, calcium ion binding, collagen binding, CXCR chemokine receptor binding, growth factor activity, integrin binding, cytokine activity, peptidase activity, acting on L-amino acid peptides, CCR chemokine receptor binding; (3) for

Table 2 GO analysis of different expressed genes in CESC.

Expression	Category	Term	Count	%	p-Value	FDR
Up-regulated	GOTERM_MF_DIRECT	GO:0005524~ATP binding	10	21.03	2.84E-4	0.270953
	GOTERM_MF_DIRECT	GO:0003678~DNA helicase activity	2	4.21	0.012333	11.169161
	GOTERM_MF_DIRECT	GO:0003688~DNA replication origin binding	2	4.21	0.018445	16.278756
	GOTERM_MF_DIRECT	GO:0003682~chromatin binding	4	8.41	0.021525	18.752384
	GOTERM_CC_DIRECT	GO:0030496~midbody	6	12.62	2.68E-7	2.65E-4
	GOTERM_CC_DIRECT	GO:0005737~cytoplasm	14	29.44	1.50E-4	0.147574
	GOTERM_CC_DIRECT	GO:0005876~spindle microtubule	3	6.31	0.001486	1.457975
	GOTERM_CC_DIRECT	GO:0005654~nucleoplasm	8	16.82	0.005554	5.351061
	GOTERM_CC_DIRECT	GO:0000784~nuclear chromosome, telomeric region	3	6.31	0.063091	9.946422
	GOTERM_CC_DIRECT	GO:0072687~meiotic spindle	2	4.21	0.014282	13.241928
	GOTERM_CC_DIRECT	GO:0042555~MCM complex	2	4.21	0.014282	13.241928
	GOTERM_CC_DIRECT	GO:0005680~anaphase-promoting complex	2	4.21	0.035339	29.901598
	GOTERM_CC_DIRECT	GO:0005634~nucleus	9	18.93	0.049470	39.406713
	GOTERM_CC_DIRECT	GO:0072686~mitotic spindle	2	4.21	0.054262	42.356169
	GOTERM_CC_DIRECT	GO:0005819~spindle	2	4.21	0.062745	47.263108
	GOTERM_CC_DIRECT	GO:0000776~kinetochore	2	4.21	0.092682	61.726231
	GOTERM_BP_DIRECT	GO:0000910~cytokinesis	3	6.31	9.78E-4	1.154575
	GOTERM_BP_DIRECT	GO:0044772~mitotic cell cycle phase transition	2	4.21	0.007770	8.844187
	GOTERM_BP_DIRECT	GO:0051988~regulation of attachment of spindle microtubules to kinetochore	2	4.21	0.011633	12.969466
	GOTERM_BP_DIRECT	GO:0031145~anaphase-promoting complex-dependent catabolic process	2	4.21	0.013559	14.961799
	GOTERM_BP_DIRECT	GO:0006268~DNA unwinding involved in DNA replication	2	4.21	0.015482	16.908679
	GOTERM_BP_DIRECT	GO:0007095~mitotic G2 DNA damage checkpoint	2	4.21	0.019316	20.670192
	GOTERM_BP_DIRECT	GO:0007076~mitotic chromosome condensation	2	4.21	0.021228	22.486821

(continued on next page)

Table 2 (continued)

Expression	Category	Term	Count	%	p-Value	FDR
	GOTERM_BP_DIRECT	GO:0000070~mitotic sister chromatid segregation	2	4.21	0.034511	34.093744
	GOTERM_BP_DIRECT	GO:0001578~microtubule bundle formation	2	4.21	0.038274	37.079667
	GOTERM_BP_DIRECT	GO:0000281~mitotic cytokinesis	2	4.21	0.040151	38.521683
	GOTERM_BP_DIRECT	GO:0006270~DNA replication initiation	2	4.21	0.040151	38.521683
Down-regulated	GOTERM_MF_DIRECT	GO:0005198~structural molecule activity	9	6.88	5.12E-6	4.557253
	GOTERM_MF_DIRECT	GO:0004252~serine-type endopeptidase activity	7	5.35	1.02E-4	0.004512
	GOTERM_MF_DIRECT	GO:0008201~heparin binding	6	4.59	1.73E-4	0.187501
	GOTERM_MF_DIRECT	GO:0005496~steroid binding	3	2.29	0.003639	3.880030
	GOTERM_MF_DIRECT	GO:0004962~endothelin receptor activity	2	1.53	0.015293	11.655613
	GOTERM_MF_DIRECT	GO:0003707~steroid hormone receptor activity	3	2.29	0.022939	31.330301
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	25	19.12	7.92E-11	7.97E-8
	GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	39	29.82	1.45E-10	4.43E-9
	GOTERM_CC_DIRECT	GO:0005576~extracellular region	12	9.18	2.40E-6	0.002416
	GOTERM_CC_DIRECT	GO:0001533~cornified envelope	5	3.82	4.01E-5	0.040335
	GOTERM_CC_DIRECT	GO:0005578~proteinaceous extracellular matrix	7	5.35	2.35E-4	0.236029
	GOTERM_CC_DIRECT	GO:0045095~keratin filament	4	3.06	0.002188	2.179044
	GOTERM_CC_DIRECT	GO:0042567~insulin-like growth factor ternary complex	2	1.53	0.023333	21.135448
	GOTERM_CC_DIRECT	GO:0031012~extracellular matrix	4	3.06	0.023982	21.661062
	GOTERM_CC_DIRECT	GO:0001527~microfibril	2	1.53	0.034797	29.965502
	GOTERM_CC_DIRECT	GO:0042581~specific granule	2	1.53	0.062878	47.957744
	GOTERM_CC_DIRECT	GO:0016323~basolateral plasma membrane	3	2.29	0.087573	60.215342
	GOTERM_BP_DIRECT	GO:0018149~peptide cross-linking	5	3.82	6.99E-4	0.098265
	GOTERM_BP_DIRECT	GO:0030216~keratinocyte differentiation	5	3.82	3.21E-4	0.451079
	GOTERM_BP_DIRECT	GO:0007565~female pregnancy	3	2.29	0.001635	2.274249

(continued on next page)

Table 2 (continued)

Expression	Category	Term	Count	%	p-Value	FDR
	GOTERM_BP_DIRECT	GO:0008284~positive regulation of cell proliferation	7	5.35	0.002416	3.344514
	GOTERM_BP_DIRECT	GO:0045840~positive regulation of mitotic nuclear division	3	2.29	0.004434	6.057636
	GOTERM_BP_DIRECT	GO:0048146~positive regulation of fibroblast proliferation	3	2.29	0.015351	19.550011
	GOTERM_BP_DIRECT	GO:0006955~immune response	5	3.82	0.016764	21.157726
	GOTERM_BP_DIRECT	GO:0001558~regulation of cell growth	3	2.29	0.018030	22.573586
	GOTERM_BP_DIRECT	GO:0001755~neural crest cell migration	3	2.29	0.0189638	23.602523
	GOTERM_BP_DIRECT	GO:0014826~vein smooth muscle contraction	2	1.53	0.022138	27.006079
	GOTERM_BP_DIRECT	GO:0001775~cell activation	2	1.53	0.033025	37.638886
	GOTERM_BP_DIRECT	GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling	3	2.29	0.034030	38.544070
	GOTERM_BP_DIRECT	GO:0007267~cell-cell signaling	3	2.29	0.042835	45.968556
	GOTERM_BP_DIRECT	GO:0021952~central nervous system projection neuron axonogenesis	2	1.53	0.043793	46.724165
	GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	3	2.29	0.048207	50.079668
	GOTERM_BP_DIRECT	GO:0048484~enteric nervous system development	2	1.53	0.049132	50.758144
	GOTERM_BP_DIRECT	GO:0043568~positive regulation of insulin-like growth factor receptor signaling pathway	2	1.53	0.049132	50.758144
	GOTERM_BP_DIRECT	GO:0005978~glycogen biosynthetic process	2	1.53	0.075392	66.786495
	GOTERM_BP_DIRECT	GO:0006885~regulation of pH	2	1.53	0.075392	66.786495
	GOTERM_BP_DIRECT	GO:0010596~negative regulation of endothelial cell migration	2	1.53	0.075392	66.786595
	GOTERM_BP_DIRECT	GO:0031290~retinal ganglion cell axon guidance	2	1.53	0.080557	69.302524
	GOTERM_BP_DIRECT	GO:0010906~regulation of glucose metabolic process	2	1.53	0.090803	73.777717

(continued on next page)

Table 2 (continued)

Expression	Category	Term	Count	%	p-Value	FDR
	GOTERM_BP_DIRECT	GO:0048662~negative regulation of smooth muscle cell proliferation	2	1.53	0.095883	75.764591
	GOTERM_BP_DIRECT	GO:0048675~axon extension	2	1.53	0.095883	75.764591

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, MCM4, MCM5, MCM2
05200	Pathways in cancer	29	2.77E-5	LAMA2, CKS1B, FGF7, EDNRA, EDNRB, RUNX1T1, PDGFRB, PDGFRA, JUP, CDKN2A, MMP1, ITGA2, PTCH1, FN1, IGF2, MITF, FOS, IGF1, WNT16, GNG11, ESR1, AR, CXCL12, GSTA4, CKS2, BIRC5, FGFR2, GSTM5, FGF10
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, PDGFRA, FGF7, CDKN2A, PDGFD, MITF, IGF1, FGF10
04151	PI3K-Akt signaling pathway	20	3.18E-4	PDGFRB, PDGFRA, TNXB, VWF, LAMA2, ITGA2, IGF2, FN1, IGF1, GNG11, AREG, EREG, GYS2, FGF7, PDGFD, SPP1, ITGA8, FGFR2, FGF10, ITGA9
04510	Focal adhesion	13	9.45E-4	PDGFRB, PDGFRA, TNXB, VWF, LAMA2, ITGA2, FN1, IGF1, MYLK, PDGFD, SPP1, ITGA8, ITGA9
04270	Vascular smooth muscle contraction	10	0.001189	ACTA2, GUCY1A2, PPP1R14A, EDNRA, EDN3, MYH11, MRV11, AVPR1A, ACTG2, MYLK
03030	DNA replication	5	0.001491	PCNA, RFC4, MCM4, MCM5, MCM2
04114	Oocyte meiosis	9	0.002920	CDC20, AR, CCNB1, PTTG1, CDK1, PGR, IGF1, SMC1B, AURKA

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 microtubule and kinesin complex.

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

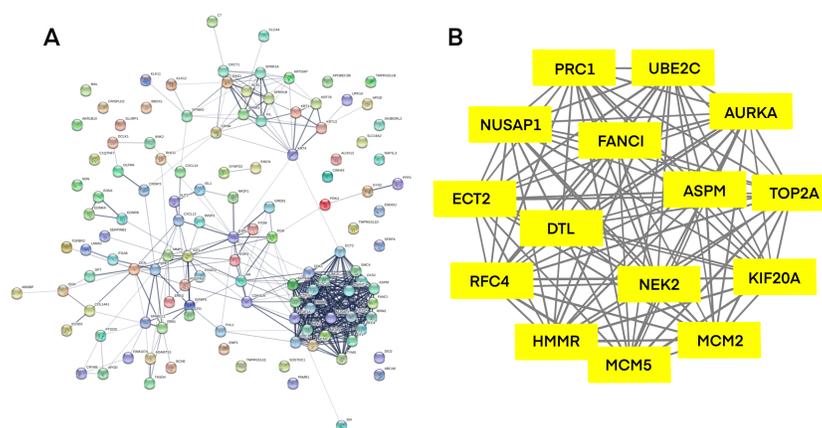


Figure 3 Common DEGs PPI network constructed by STRING online database and Module analysis. (A) Nodes meant proteins; the edges meant the interaction of proteins. (B) Module analysis via Cytoscape MCODE tool (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. Depth = 100).

Full-size [DOI: 10.7717/peerj.10386/fig-3](https://doi.org/10.7717/peerj.10386/fig-3)

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 (Fig. 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 ($P > 0.05$) (Fig. 4 and Table 4). 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 ($P < 0.05$) (Fig. 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 ($P = 0.0197$) (Fig. 6). And the expression of RFC4 in early stage CESC was significantly higher than that in advanced CESC ($P = 0.0314$) (Fig. 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 (Fig. 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 (Fig. 9).

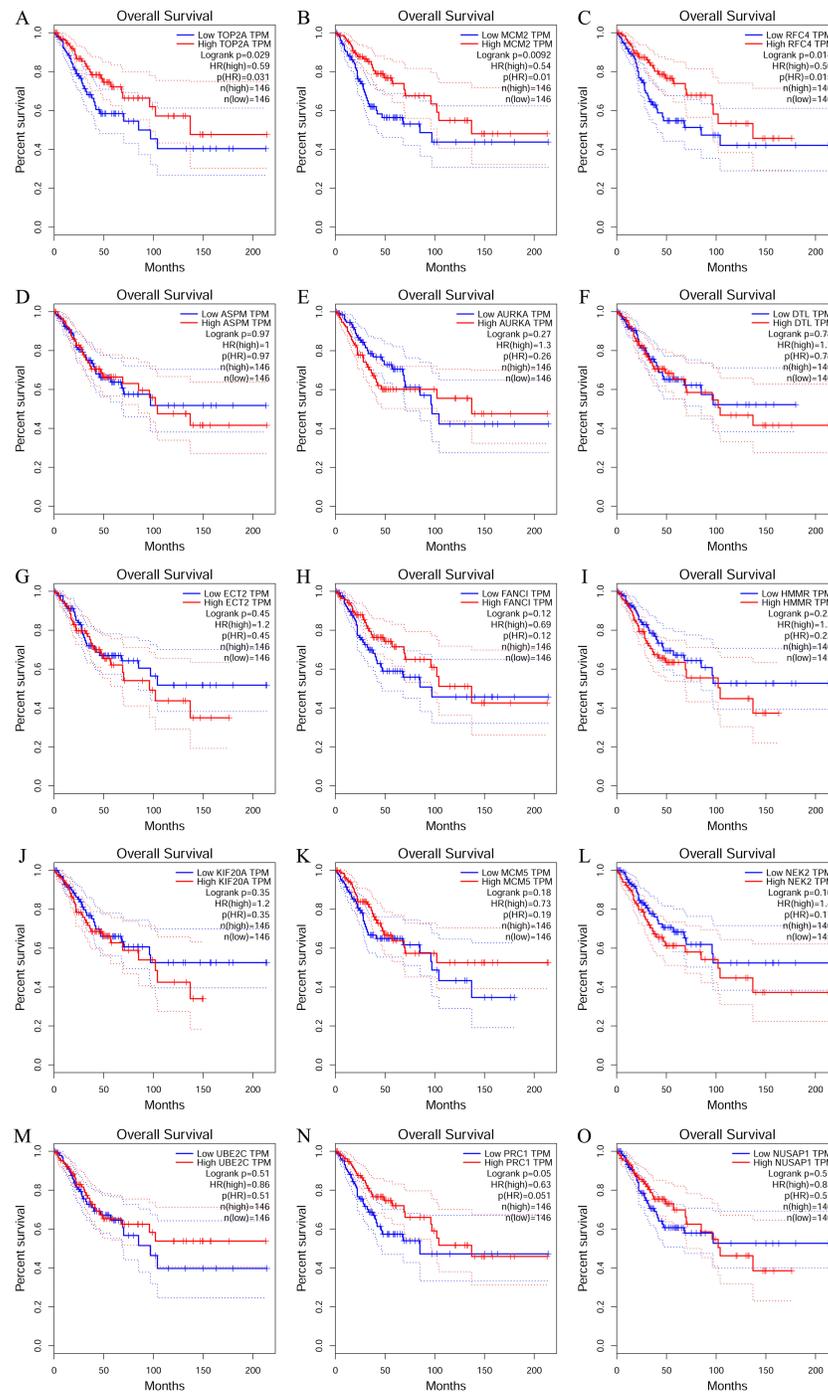


Figure 4 The prognostic information of the 15 core genes. Three (A, B and C) of 15 genes had a significant better survival rate ($P < 0.05$) and twelve genes (D–O) had not significant difference in OS ($P > 0.05$).

Full-size DOI: 10.7717/peerj.10386/fig-4

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

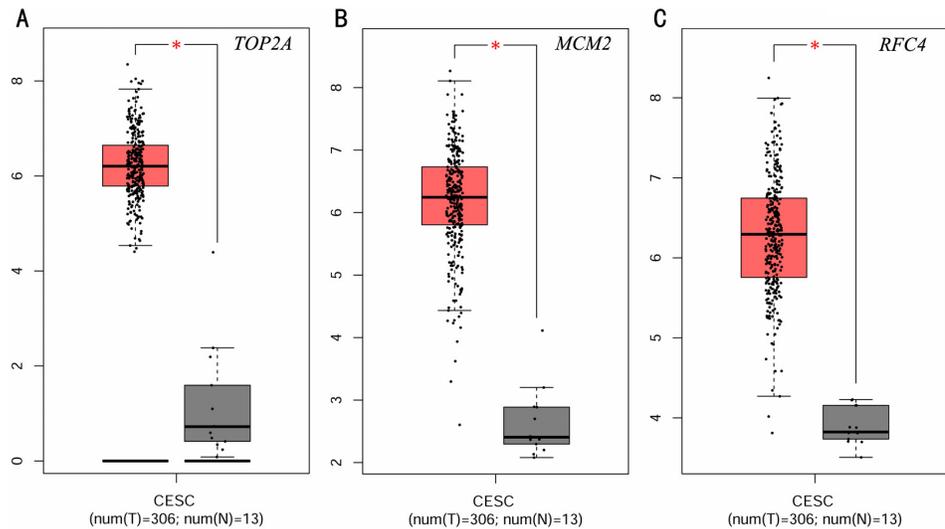


Figure 5 Expression level of three significantly expressed genes in CESC tissues and normal tissues. (A) The expression level of TOP2A in CESC tissues and normal tissues. (B) The expression level of MCM2 in CESC tissues and normal tissues. (C) The expression level of RFC4 in CESC tissues and normal tissues. Red color means tumor tissues and grey means normal tissues.

Full-size DOI: 10.7717/peerj.10386/fig-5

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 seven 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. Research reported that genes from these pathways could be associated with the pathogenesis and progression of cervical cancer. 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).

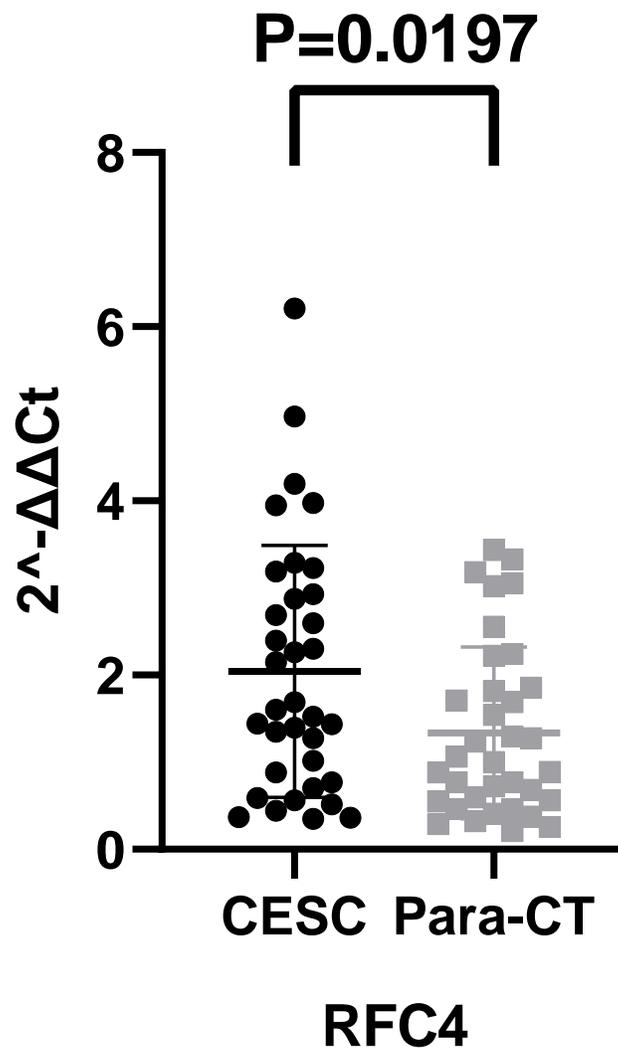


Figure 6 The expression of RFC4 on CESC was significantly different compared with para-cancerous tissues from the result of qRT-PCR.

Full-size  DOI: [10.7717/peerj.10386/fig-6](https://doi.org/10.7717/peerj.10386/fig-6)

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 found three DEGs from the 15 which had a significantly better survival. The results of GEPIA showed that the expression levels of the three 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 five subunits (RFC1-5). Among all five subunits of RFC complex, RFC4 has been reported to play an important role in DNA damage checkpoint and DNA replication

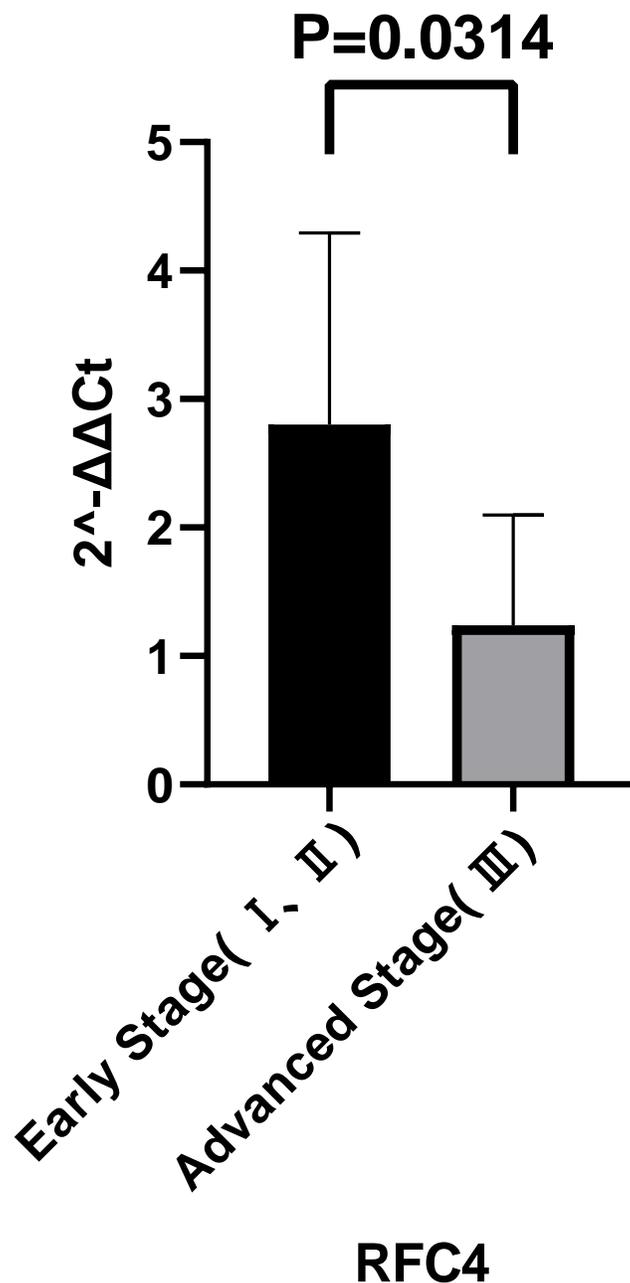


Figure 7 Expression levels of RFC4 in different pathological stages of CESC. The expression of RFC4 in early stage CESC was significantly higher than that in advanced CESC ($P = 0.0314$).

Full-size DOI: 10.7717/peerj.10386/fig-7

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, Choi & Kim, 2009; Erdogan et al., 2009; Barfeld et al., 2014). Research illustrated that up-regulated RFC4

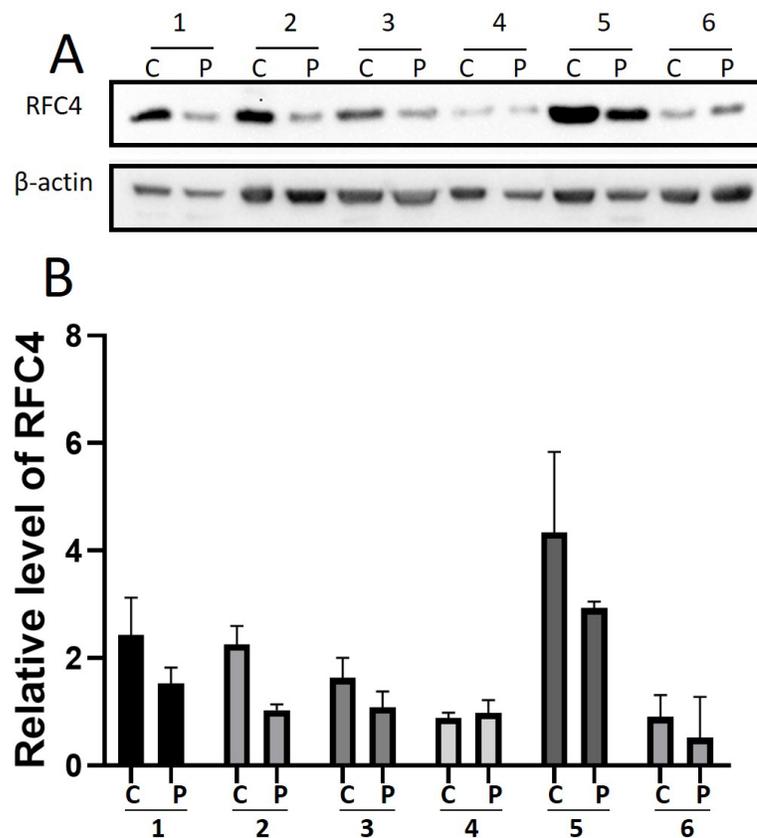


Figure 8 WB analysis of RFC4 protein. C: CESC tissues, P: para-cancerous tissues. (A) Six pairs CESC tissues WB analysis indicated that except that the results of case 4 are not obvious, the others are consistent with the expected results of high expression of RFC4 in tumor tissues. (B) The grayscale analysis of multiple WB bands shows that the WB tests are reliable.

Full-size DOI: 10.7717/peerj.10386/fig-8

expression 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.* (2012) 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. 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 cycle 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

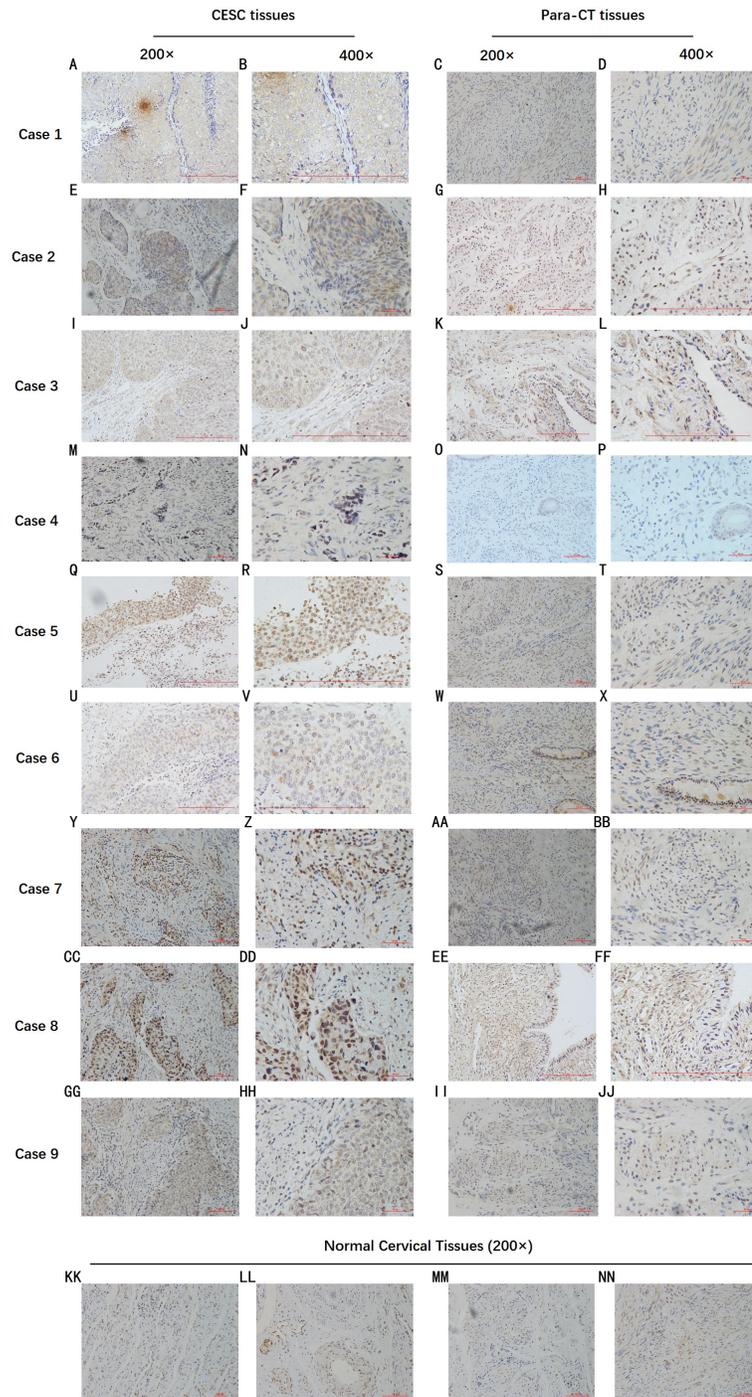


Figure 9 IHC test of CESC. IHC declared that, in general, the RFC4 protein is highly expressed in tumor tissue sections, and is mainly concentrated in the nucleus, while normal cervical tissue and para-cancerous tissues are underexpressed.

Full-size  DOI: [10.7717/peerj.10386/fig-9](https://doi.org/10.7717/peerj.10386/fig-9)

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. A high level expressed RFC4 was validated to exist in CESC tissues using clinical samples. Although further investigation and experiments needs to be conducted, the findings in our study could act as clinical biomarkers which would help us better understand the pathological process and predict the prognostic of CESC.

ACKNOWLEDGEMENTS

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

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yunan He conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Shunjie Hu analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jiaojiao Zhong performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Anran Cheng conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Nianchun Shan conceived and designed the experiments, performed the experiments, prepared figures and/or tables, and approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Medical Ethics Committee of Xiyang Hospital approved this research (No.201912542).

Data Availability

The following information was supplied regarding data availability:

The main data of this study are available at NCBI-GEO: [GSE27678](#), [GSE39001](#), [GSE7803](#). The raw data are available as [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.10386#supplemental-information>.

REFERENCES

- Arai M, Kondoh N, Imazeki N, Hada A, Hatsuse K, Matsubara O, 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 DOI [10.1111/j.1478-3231.2008.01792.x](#).
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25:25–29 DOI [10.1038/75556](#).
- Bader GD, Hogue CW. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2 DOI [10.1186/1471-2105-4-2](#).
- Banwait JK, Bastola DR. 2015. Contribution of bioinformatics prediction in microRNA-based cancer therapeutics. *Advanced Drug Delivery Reviews* 81:94–103 DOI [10.1016/j.addr.2014.10.030](#).
- Barfeld SJ, East P, Zuber V, Mills IG. 2014. Meta-analysis of prostate cancer gene expression data identifies a novel discriminatory signature enriched for glycosylating enzymes. *BMC Medical Genomics* 7:513 DOI [10.1186/s12920-014-0074-9](#).
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* 41:D991–D995 DOI [10.1093/nar/gks1193](#).
- Chaplain MA, Sleeman BD. 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 DOI [10.1007/bf00173886](#).
- Chatterjee S, Do KS, Alam S, Salzberg AC, Milici J, van der Burg SH, Freeman W, Meyers C. 2019. Tissue-Specific Gene Expression during Productive Human Papillomavirus 16 Infection of Cervical, Foreskin, and Tonsil Epithelium. *Journal of Virology* 93:e00915-19 DOI [10.1128/JVI.00915-19](#).

- Crook T, Wrede D, Tidy JA, Mason WP, Evans DJ, Vousden KH. 1992.** Clonal p53 mutation in primary cervical cancer: association with human-papillomavirus-negative tumours. *Lancet* **339**:1070–1073 DOI [10.1016/0140-6736\(92\)90662-m](https://doi.org/10.1016/0140-6736(92)90662-m).
- Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. 2019.** Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research* **18**:623–632 DOI [10.1021/acs.jproteome.8b00702](https://doi.org/10.1021/acs.jproteome.8b00702).
- Ellison V, Stillman B. 2003.** Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5′/recessed DNA. *PLOS Biology* **1**:Article E33 DOI [10.1371/journal.pbio.0000033](https://doi.org/10.1371/journal.pbio.0000033).
- Erdogan E, Klee EW, Thompson EA, Fields AP. 2009.** Meta-analysis of oncogenic protein kinase Ciota signaling in lung adenocarcinoma. *Clinical Cancer Research* **15**:1527–1533 DOI [10.1158/1078-0432.CCR-08-2459](https://doi.org/10.1158/1078-0432.CCR-08-2459).
- Fatima A, Tariq F, Malik M, Qasim M, 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 DOI [10.4048/jbc.2017.20.3.246](https://doi.org/10.4048/jbc.2017.20.3.246).
- Galic V, Herzog TJ, Lewin SN, Neugut AI, Burke WM, Lu YS, Hershman DL, Wright JD. 2012.** Prognostic significance of adenocarcinoma histology in women with cervical cancer. *Gynecologic Oncology* **125**:287–291 DOI [10.1016/j.ygyno.2012.01.012](https://doi.org/10.1016/j.ygyno.2012.01.012).
- Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR, 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 JL, Price S, Hur W, Yang W, Deng X, Butler A, Choi HG, Chang JW, Baselga J, Stamenkovic I, Engelman JA, Sharma SV, Delattre O, Saez-Rodriguez J, Gray NS, Settleman J, Futreal PA, Haber DA, Stratton MR, Ramaswamy S, McDermott U, Benes CH. 2012.** Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**:570–575 DOI [10.1038/nature11005](https://doi.org/10.1038/nature11005).
- Juan C, Hua Q, Ruping Z, Tingting W. 2018.** miRNA-489 as a biomarker in diagnosis and treatment of cervical cancer. *Bratislavske Lekarske Listy* **119**:278–283 DOI [10.4149/BLL_2018_052](https://doi.org/10.4149/BLL_2018_052).
- Jung HM, Choi SJ, Kim JK. 2009.** Expression profiles of SV40-immortalization-associated genes upregulated in various human cancers. *Journal of Cellular Biochemistry* **106**:703–713 DOI [10.1002/jcb.22063](https://doi.org/10.1002/jcb.22063).
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017.** KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**:D353–D361 DOI [10.1093/nar/gkw1092](https://doi.org/10.1093/nar/gkw1092).
- LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL. 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, Zhang Y. 2019.** Nucleolar and spindle associated protein 1 promotes metastasis of cervical carcinoma cells by

- activating Wnt/beta-catenin signaling. *Journal of Experimental & Clinical Cancer Research* **38**:Article 33 DOI [10.1186/s13046-019-1037-y](https://doi.org/10.1186/s13046-019-1037-y).
- Mao Y, Fu Z, Dong L, Zheng Y, Dong J, 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 DOI [10.1089/dna.2018.4533](https://doi.org/10.1089/dna.2018.4533).
- Meuris F, Carthagen L, Jaracz-Ros A, Gaudin F, Cutolo P, Deback C, Xue Y, Thierry F, Doorbar J, Bachelier F. 2016.** The CXCL12/CXCR4 signaling pathway: a new susceptibility factor in human papillomavirus pathogenesis. *PLOS Pathogens* **12**:e1006039 DOI [10.1371/journal.ppat.1006039](https://doi.org/10.1371/journal.ppat.1006039).
- Niu G, Wang D, Pei Y, Sun L. 2017.** Systematic identification of key genes and pathways in the development of invasive cervical cancer. *Gene* **618**:28–41 DOI [10.1016/j.gene.2017.03.018](https://doi.org/10.1016/j.gene.2017.03.018).
- Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. 2010.** A practical guide to understanding Kaplan–Meier curves. *Otolaryngology- Head and Neck Surgery* **143**:331–336 DOI [10.1016/j.otohns.2010.05.007](https://doi.org/10.1016/j.otohns.2010.05.007).
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003.** Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504 DOI [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303).
- Slebos RJ, Yi Y, Ely K, Carter J, Evjen A, Zhang X, Shyr Y, Murphy BM, Cmelak AJ, Burkey BB, Netterville JL, Levy S, Yarbrough WG, Chung CH. 2006.** Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clinical Cancer Research* **12**:701–709 DOI [10.1158/1078-0432.CCR-05-2017](https://doi.org/10.1158/1078-0432.CCR-05-2017).
- Sun L, Jiang R, Li J, Wang B, Ma C, Lv Y, Mu N. 2017.** MicoRNA-425-5p is a potential prognostic biomarker for cervical cancer. *Annals of Clinical Biochemistry* **54**:127–133 DOI [10.1177/0004563216649377](https://doi.org/10.1177/0004563216649377).
- Szymanska Z, Cytowski M, Mitchell E, Macnamara CK, Chaplain M. 2018.** Computational modelling of cancer development and growth: modelling at multiple scales and multiscale modelling. *Bulletin of Mathematical Biology* **80**:1366–1403 DOI [10.1007/s11538-017-0292-3](https://doi.org/10.1007/s11538-017-0292-3).
- Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. 2017.** GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research* **45**:W98–W102 DOI [10.1093/nar/gkx247](https://doi.org/10.1093/nar/gkx247).
- Vu M, Yu J, Awolude OA, Chuang L. 2018.** Cervical cancer worldwide. *Current Problems in Cancer* **42**:457–465 DOI [10.1016/j.currproblcancer.2018.06.003](https://doi.org/10.1016/j.currproblcancer.2018.06.003).
- Wang SS, Sherman ME, Hildesheim A, Lacey JJ, 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 DOI [10.1002/cncr.20064](https://doi.org/10.1002/cncr.20064).

- Yang YK, Xi WY, Xi RX, Li JY, Li Q, Gao YE. 2015.** MicroRNA-494 promotes cervical cancer proliferation through the regulation of PTEN. *Oncology Reports* **33**:2393–2401 DOI [10.3892/or.2015.3821](https://doi.org/10.3892/or.2015.3821).
- Zhang J, Burn C, Young K, Wilson M, Ly K, Budhwani M, Tschirley A, Braithwaite A, Baird M, Hibma M. 2018.** Microparticles produced by human papillomavirus type 16 E7-expressing cells impair antigen presenting cell function and the cytotoxic T cell response. *Scientific Reports* **8**:2373 DOI [10.1038/s41598-018-20779-2](https://doi.org/10.1038/s41598-018-20779-2).
- Zhou Y, Hingorani MM. 2012.** Impact of individual proliferating cell nuclear antigen-DNA contacts on clamp loading and function on DNA. *Journal of Biological Chemistry* **287**:35370–35381 DOI [10.1074/jbc.M112.399071](https://doi.org/10.1074/jbc.M112.399071).