Longitudinal Analysis of Gene Expression Changes During Cervical Carcinogenesis Reveals Potential Therapeutic Targets

Evolutionary Bioinformatics
Volume 16: 1–10

DOI: 10.1177/1176934320920574
SAGE

Lijun Yu, Meiyan Wei and Fengyan Li

Department of Gynecology, First Hospital of Shanxi Medical University, Taiyuan, China.

ABSTRACT: Despite advances in the treatment of cervical cancer (CC), the prognosis of patients with CC remains to be improved. This study aimed to explore candidate gene targets for CC. CC datasets were downloaded from the Gene Expression Omnibus database. Genes with similar expression trends in varying steps of CC development were clustered using Short Time-series Expression Miner (STEM) software. Gene functions were then analyzed using the Gene Ontology (GO) database and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Protein interactions among genes of interest were predicted, followed by drug-target genes and prognosis-associated genes. The expressions of the predicted genes were determined using real-time quantitative polymerase chain reaction (RT-qPCR) and Western blotting. Red and green profiles with upward and downward gene expressions, respectively, were screened using STEM software. Genes with increased expression were significantly enriched in DNA replication, cell-cycle-related biological processes, and the p53 signaling pathway. Based on the predicted results of the Drug-Gene Interaction database, 17 drug-gene interaction pairs, including 3 red profile genes (TOP2A, RRM2, and POLA1) and 16 drugs, were obtained. The Cancer Genome Atlas data analysis showed that high POLA1 expression was significantly correlated with prolonged survival, indicating that POLA1 is protective against CC. RT-qPCR and Western blotting showed that the expressions of TOP2A, RRM2, and POLA1 gradually increased in the multistep process of CC. TOP2A, RRM2, and POLA1 may be targets for the treatment of CC. However, many studies are needed to validate our findings.

KEYWORDS: Cervical cancer, time-series analysis, drug–gene interactions

RECEIVED: July 10, 2019. **ACCEPTED:** March 24, 2020.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Lijun Yu, Department of Gynecology, First Hospital of Shanxi Medical University, No. 85 Jiefang South Road, Taiyuan 030001, Shanxi, China. Email: yulijun100@163.com

Background

Cervical cancer (CC) is the third most common cancer worldwide and the third most malignant tumor after breast and colon cancer.[1,2] The occurrence of CC has recently been increasing, and the latest statistics indicate that CC ranks fourth in the incidence of (570 000 cases) and mortality (311 000 deaths) associated with cancers.[3] The incidence of CC is relatively high in developing countries. Cervical cancer cases and related deaths in China account for 12% of all CC cases and CC-related deaths worldwide.[4] Furthermore, CC is a threat to the health of women in China.

Scientific research has proven that cancer is caused by genetic changes, and researchers have identified several marker genes and potential drug targets for different cancer types. A previous study has shown that tumorigenesis is closely related to the abnormal expression of proteins associated with cell signal regulation.[5] These genetic and molecular events ultimately contribute to tumor initiation and progression. Therefore, the identification of genetic changes associated with CC can provide a conceptual framework for further analysis of this complex disease.

Study on molecular mechanisms can shed light on specific markers that play key roles in the early diagnosis and treatment of CC. The discovery of new biomarkers may facilitate the effective prevention and treatment of CC. Reportedly, there are a large number of cancer-related genes in CC tissues.[6-8] The discovery of these genes has played an important role in the diagnosis and treatment of CC.[9] In addition, abnormal activation of some signaling pathways, such as the MAPK, Wntll, and Notch signaling pathways, which are involved in tumor growth and metastasis, has been identified.[10-13] Studies on major genes involved in the signaling pathway of CC, such as *PAI-1*, *HK-2*, and *BCL-2*, can provide data on molecular targets for anticancer drugs. Critical genes playing important roles in carcinogenesis have attracted extensive attention as targets for anticancer therapy.[14-16] However, the genetic changes underlying the multistep process of CC have not been clearly elucidated. Effective cancer treatment needs to focus on gene changes and the specific characteristics of signal pathways.

Recent developments in gene chip and sequencing technology have increased the amount of high-throughput data. High-throughput sequencing technology, also known as next-generation sequencing (NGS), has revolutionized the complete processes of genome sequencing, transcriptomics, and epigenetics. The Gene Expression Omnibus (GEO) database is a large public database that provides high-throughput data for research on various diseases.[17] Time-series analysis of microarray data facilitates the analysis of dynamic biological processes of altered genes.[18] To further explore the pathogenesis of CC and screen biological targets, gene expression data of 24 normal, 14 cervical intraepithelial neoplasia 1 (CIN1) lesions, 22 CIN2 lesions, 40 CIN3 lesions, and 28 cancer specimens were downloaded for further analysis. Time series, drug-gene

interaction prediction, and survival analyses of key genes were conducted. This study attempted to provide new insights into future research on CC treatment.

Materials and Methods

Data source and data preprocessing of expression profiles

Microarray data (ID: GSE63514) on CC were downloaded from the GEO database[17] based on the platform of the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. The mRNA expression profile data contained 24 normal cervical tissue, 14 CIN1 lesion, 22 CIN2 lesion, and 40 CIN3 lesion samples and 28 cancers specimens. The raw data were normalized and pretreated using oligo package version 1.58.0 (http://bioconductor.org/help/search/index.html?q=oligo). Finally, the standardized matrices of 5 disease states were obtained for further analysis.

Time-series analysis of sequence expression profiles

The time-series analysis for genes was performed using the Short Time-series Expression Miner (STEM) (http://www.cs.cmu.edu/~jernst/stem/) software (version 1.3.11), which is used for clustering, comparing, and visualizing exogenous genes based on gene chip expression data.[18] To screen for the gene clusters significantly correlated with the time series, the count of genes for each cluster was set as >20, the correlation coefficient of gene expression in each cluster was set as >0.7, and the gene annotation source was set as *Homo sapiens*. All significant gene clusters were obtained at $P < .05$, and genes with similar expression trends (profiles with the same color) were assembled for subsequent analysis.

Analysis of gene clusters related to CC progression

The Gene Ontology (GO) database (http://www.geneontology.org/) provides access to gene function annotations that cover molecular and cellular bases.[19] There are 3 GO categories: cell cycle (CC), molecular function (MF), and biological process (BP). Database for Annotation, Visualization and Integrated Discovery (DAVID) is a commonly used enrichment analysis tool, which was established by American scientists to provide comprehensive information on gene lists or biometric data of large-scale proteins.[20]

To analyze the differential expression genes (DEGs) involved in biological functions, the Gene Ontology-Biological Process (GO-BP)[21] functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG)[22] pathway enrichment analyses were conducted using DAVID (Version 6.8; https://david-d.ncifcrf.gov/).[23] Enrichment gene numbers ⩾2 and significance threshold (P) < .05 were considered to indicate significant enrichment.

Protein-protein interaction network and modular analysis of significantly clustered genes

Protein-protein interactions (PPIs) reveal complex regulatory networks of functional proteins at the molecular level. Search Tool for the Retrieval of Interacting Genes (STRING) database collects information on predicted and experimental PPIs in a given cell.[24] DEGs encoding proteins were mapped to the STRING database (Version: 10.0, http://www.string-db.org/) to construct a PPI network. Protein pairs with a combined score of >0.4 were collected.

Genes with similar functions can be clustered together. Cytoscape plug-in MCODE (Version 1.4.2, http://apps.cytoscape.org/apps/MCODE)[25] was used to analyze the clustering module in PPI networks with a threshold score of >5. GO-BP and KEGG pathway enrichment analyses were performed for genes in the significant modules.

Prediction of drug-gene interactions

Existing resources were mined using the Drug-Gene Interaction database (DGIdb) to generate hypotheses on how genes are targeted or prioritized for drug development.[26] To explore drug targets among genes of interest, DGIdb2.0 (http://www.dgidb.org/) was used to predict drugs related to significant cluster genes (Preset Filters: FDA Approved + Antineoplastic). Drug databases were limited to the Food and Drug Administration (FDA) and DrugBank. The drug-gene interaction network was visualized using Cytoscape software (version: 3.2.0; http://www.cytoscape.org/).

Survival analysis of key genes

To explore the key genes associated with CC prognosis, survival analysis was performed. First, the matrix data and clinical information of the key genes involved in the drug-gene network in The Cancer Genome Atlas (TCGA)-cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) were collated. Then, the results of the survival analysis and Cox regression analysis of important node genes were analyzed using the coxph function in the R package (Version: 3.4.2, https://cran.r-project.org/web/packages/survival/index.html). According to the median of the expression value of a given gene, samples were divided into high and low expressions. The correlation coefficient $P < .05$ was set as the significant threshold, and Kaplan-Meier (K-M) survival curve was generated.[27] Clinical information was analyzed based on overall survival (OS) provided by TCGA.

Real time real-time PCR analysis

The CIN1, CIN2, and CIN3 lesions as well as CC tumor samples and adjacent normal tissues (n = 5 per group) were obtained from patients admitted in our hospital. Approval was obtained

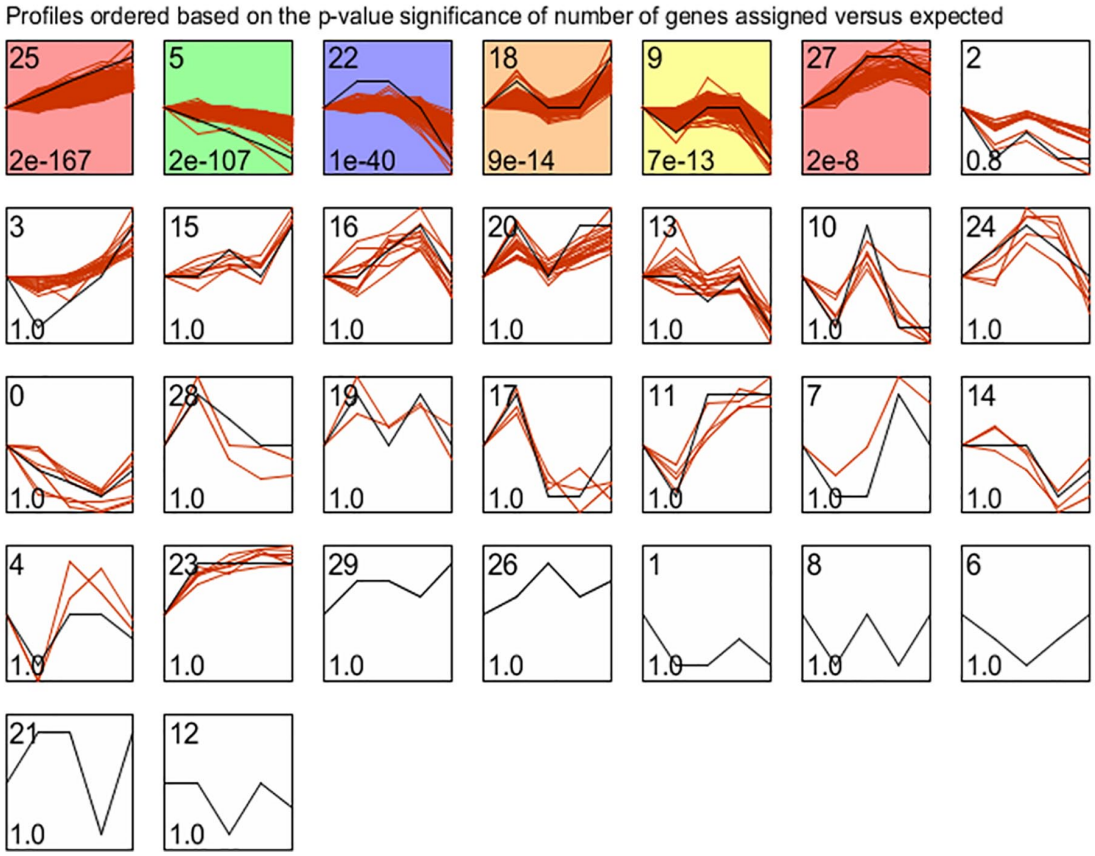

Figure 1. Cluster map of the time-series expression profile analysis.
Time-series analysis of genes was performed using the STEM software. Profiles with colors were significantly clustered ($P < .05$). Red profile genes with increasing expression trends and green profile genes with declining expression trends were collected for further analysis. Block: gene sets of different clusters; number on the upper left: number of clustered gene sets; black line: the overall trend of gene expression.

from the Ethics Committee of the First Hospital of Shanxi Medical University, and all the study procedures were performed according to the ethical standards. Informed consent was obtained from all the patients before performing the study.

Total RNA was isolated using RNAiso plus and TRIzol reagents (TaKaRa, Shiga, Japan). RNA (1 μg) was reverse transcribed to cDNA, and polymerase chain reaction (PCR) was performed with the following conditions: 50°C for 2 minutes, 95°C for 10 minutes, followed by 40 cycles of 95°C for 10 seconds and 60°C for 30 seconds. Melt curve analysis was performed from 60°C to 95°C, with an increment of 0.5°C every 10 seconds. The sequences of the primers used for PCR analysis are listed in Supplemental Table S1.

Western blotting

Proteins were extracted from corresponding normal and tumor tissues using RNA immunoprecipitation analysis lysis buffer (Solarbio, Beijing, China). After the quality of the extracted protein was detected using the bicinchoninic acid assay, the purified protein was separated using 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Then, the protein was transferred onto a polyvinylidene difluoride (PVDF) membrane and blocked with 5% skim milk. Then, the membranes were incubated with primary antibodies of rabbit anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody (Abcam, Ab22555, 1:300), rabbit anti TOP2A polyclonal antibody (Abcam, Ab138683, 1:200), rabbit anti-RRM2 antibody (Abcam, Ab172476, 1:200), rabbit anti-POLA antibody (Abcam, Ab225562, 1:200) at 4°C overnight, followed by incubation with goat anti-rabbit IgG secondary antibodies (Thermo, A16104, 1:3000) for 1 hour at room temperature. The protein expressions of TOP2A, RRM2, and POLA1 were detected relative to that of GAPDH.

Statistical analysis

Data were expressed as mean ± standard deviation. Between-group comparisons were analyzed using Graphpad prism 5 (Graphpad Software, San Diego, CA). Multiple comparison analysis was performed using 1-way analysis of variance (ANOVA), followed by Tukey test. $P < .05$ was considered statistically significant.

Results

Sequence analysis of gene expression profile concentrations

The time-series expression profiles of all the genes were analyzed based on the results of STEM. As shown in Figure 1, color consistency indicates trend consistency. Profiles 25 and 27 had the same color and upward trend and were hence pooled as red profile (up-gene). Profile 5 was significantly clustered and showed a

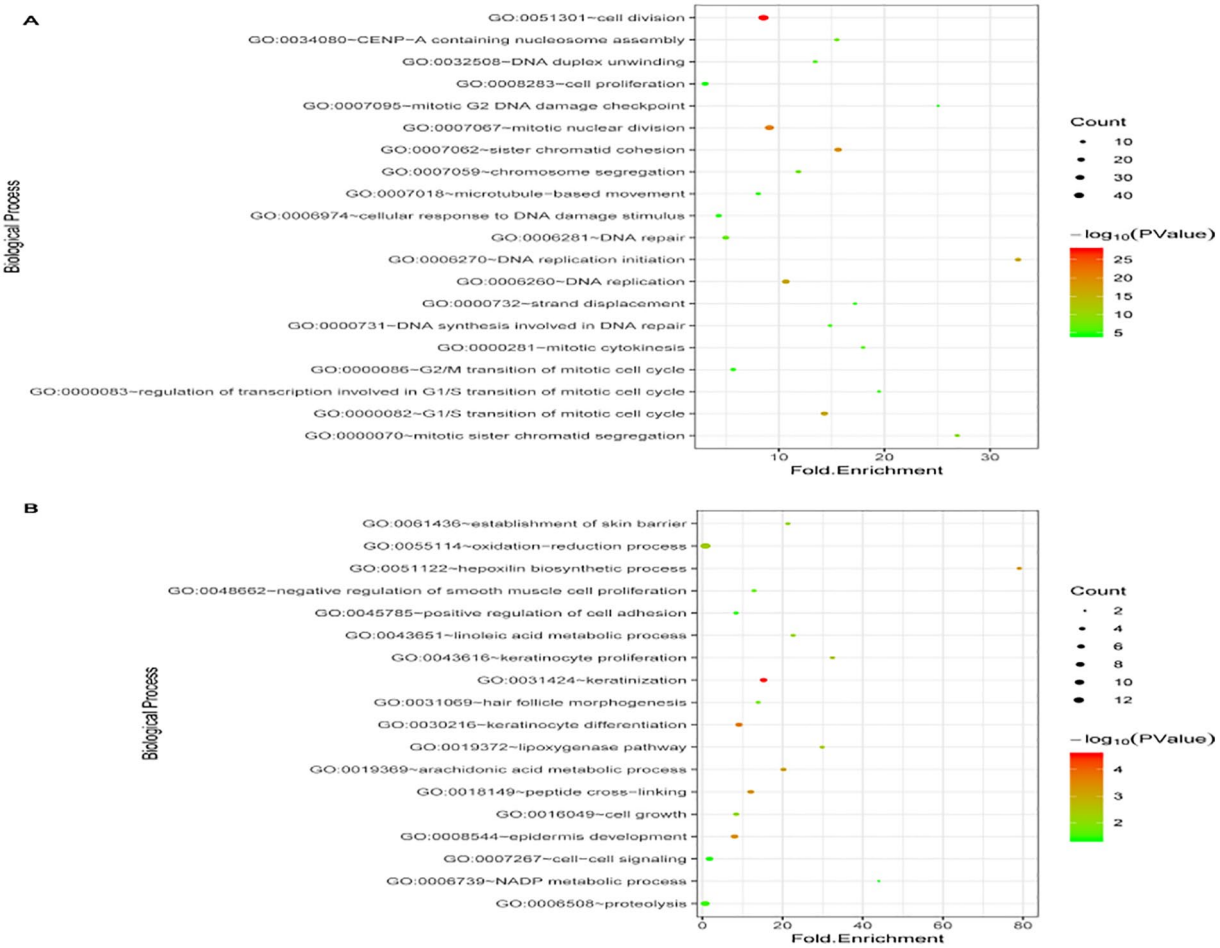

Figure 2. Results of the GO enrichment analysis. GO enrichment analysis of genes in the red and green profiles was performed using DAVID. GO terms with $P < .05$ were considered significant. (A) Enrichment results of the top 20 red profile genes and (B) all enrichment results of the green profile genes. DAVID indicates Database for Annotation, Visualization and Integrated Discovery; GO, gene ontology.

downward trend; thus, its enriched genes were pooled as green profile (down-gene). The trend changes of profiles 22, 18, and 9 were nondirective and not included in further analyses.

Functional analysis of significant clustering genes in CC

GO-BP functional enrichment analyses of stage-related genes were performed using DAVID. The results showed that the red and green profile genes were significantly enriched in 76 and 18 BP, respectively. The enrichment results of the top 20 genes in the red profile and all of the enrichment results of the green profile genes have been shown in Figure 2. The main functions of the red profile genes are cell division, proliferation, migration, and apoptosis as well as DNA damage stimulus and replication (Figure 2A). The main functions of the green profile genes are mainly immune, inflammatory, and extracellular matrix responses (Figure 2B).

KEGG analysis of significant clustering genes

KEGG analysis of significant clustering genes was performed using DAVID. The results showed that the red and green profile genes were significantly enriched in 8 and 3 KEGG pathways, respectively (Figure 3). The results showed that the pathways of the red profile genes mainly included DNA replication, cell cycle, Fanconi anemia pathway, pyrimidine metabolism, oocyte meiosis, mismatch repair, p53 signal transduction pathway, and microRNAs. In contrast, the pathways of the green profile genes were closely related to arachidonic acid metabolism, metabolic pathways, and serotonergic synapse.

PPI network analysis and enrichment analysis of key nodes

As shown in Figure 4A, the PPI network has 195 nodes and 1994 interaction pairs. The topology score was high and could be regarded as the key node of the network. Three subnetwork modules of the PPI network were gathered using MCODE (score > 5 in Cytoscape plug-in). Module-A (score = 39.4) contained 41 nodes and 788 interaction pairs, module-B (score = 10) contained 24 nodes and 43 interaction pairs, and module-C (score = 9.4) contained 11 nodes and 47 interaction pairs. The nodes in significant modules all belonged to the red profile genes. Detailed information of module genes is listed in Table 1. We also performed KEGG pathway and GO-BP enrichment analyses for the module genes. Function analysis

showed that the module-A genes were closely related to cell division (GO-BP), mitotic nuclear division (GO-BP), oocyte meiosis, and the cell cycle pathway. Module-B genes were significantly enriched in DNA replication initiation (GO-BP), G1/S transition of mitotic cell cycle (GO-BP), cell cycle, and DNA replication-related pathways. Module-C genes were significantly involved in GO BPs such as sister chromatid cohesion, mitotic nuclear division, and cell division (Figure 4B). However, no pathway was significantly enriched by module-C genes.

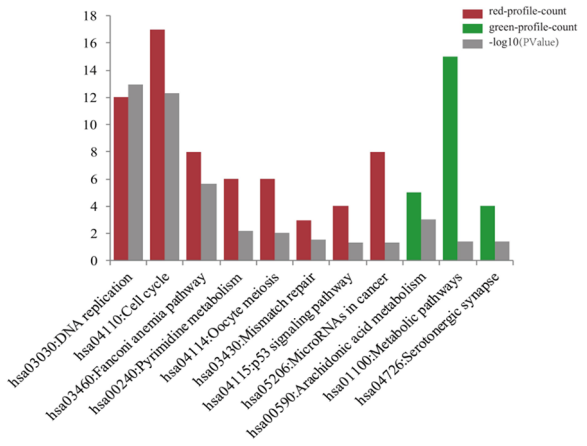

Figure 3. Results of the KEGG pathway enrichment analysis. KEGG pathway analysis was performed for genes clustered in the red and green profiles using DAVID. Significantly enriched pathways were displayed ($P < .05$). Red: red profile; green: green profile; gray: $-\log^{10}$ (*P*-value). DAVID indicates Database for Annotation, Visualization and Integrated Discovery; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Prediction of drug-gene interaction pairs

As shown in Figure 5, we obtained 17 drug-gene interaction pairs based on DGIdb predictions of all module genes. There were 3 red profile genes (TOP2A, RRM2, and POLA1) and 16 drugs. The interactions between these genes and drugs were inhibited.

Survival analysis

As shown in Figure 6, the survival analysis of POLA1 was performed according to TCGA-CESC matrix data and clinical information. The results showed that POLA1 was significantly correlated with CC prognosis ($P = .016$), indicating that POLA1 was a prognostic predictor of CC. Furthermore, high expression of POLA1 was significantly associated with poor survival.

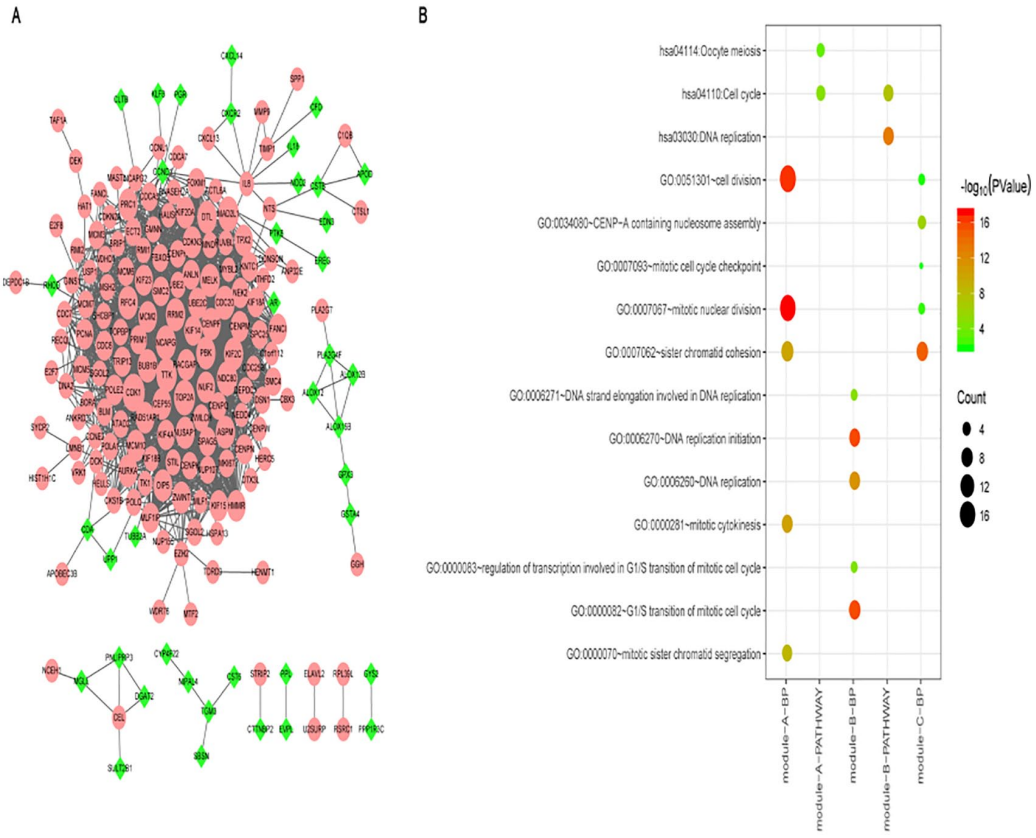

Figure 4. PPI network construction (A) and significant biological functions of module genes (B).
Protein interactions of genes clustered in the red and green profiles were mapped based on the STRING database. A PPI network was constructed with protein pairs with a combined score of >0.4, and 3 significant modules were analyzed using the MCODE plug-in. The PPI network was constructed with 195 nodes and 1994 edges. GO-BP and KEGG pathway enrichment analyses were conducted for genes in modules A, B, and C. The top 5 GO BP terms were listed. Red circle: red profile genes; green prism: green profile genes. BP indicates biological process; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes.

Table 1. Module gene list.

MODULE-A						MODULE-B			MODULE-C		
NODES	PROFILE	DEGREE	NODES	PROFILE	DEGREE	NODES	PROFILE	DEGREE	NODES	PROFILE	DEGREE
CDK1	Red	94	KIF20A	Red	55	PCNA	Red	47	RFC4	Red	64
TOP2A	Red	78	CDKN3	Red	55	CDC6	Red	41	MLF1IP	Red	45
MAD2L1	Red	77	CEP55	Red	54	PRIM1	Red	34	CENPN	Red	41
NDC80	Red	73	HMMR	Red	54	MCM5	Red	34	KNTC1	Red	39
NCAPG	Red	72	UBE2C	Red	53	MCM3	Red	32	NUP107	Red	26
TTK	Red	71	RAD51AP1	Red	53	MCM6	Red	32	CENPQ	Red	26
CDC20	Red	69	TPX2	Red	52	MCM7	Red	30	CENPH	Red	25
BUB1B	Red	64	RACGAP1	Red	51	POLA1	Red	25	CENPK	Red	25
DTL	Red	63	NUF2	Red	51	WDHD1	Red	19	C1orf112	Red	23
MELK	Red	63	KIF15	Red	50	ANKRD32	Red	15	DSN1	Red	23
RRM2	Red	62	FANCI	Red	48				ZWILCH	Red	23
AURKA	Red	62	CENPM	Red	48						
KIF2C	Red	61	KIF23	Red	47						
ASPM	Red	61	KIF4A	Red	44						
PBK	Red	60	FOXM1	Red	43						
CENPF	Red	59	NEK2	Red	43						
PRC1	Red	58	TRIP13	Red	42						
NUSAP1	Red	57	KIF14	Red	42						
OIP5	Red	56	ANLN	Red	40						
ZWINT	Red	56	SPAG5	Red	37						
MCM10	Red	56									

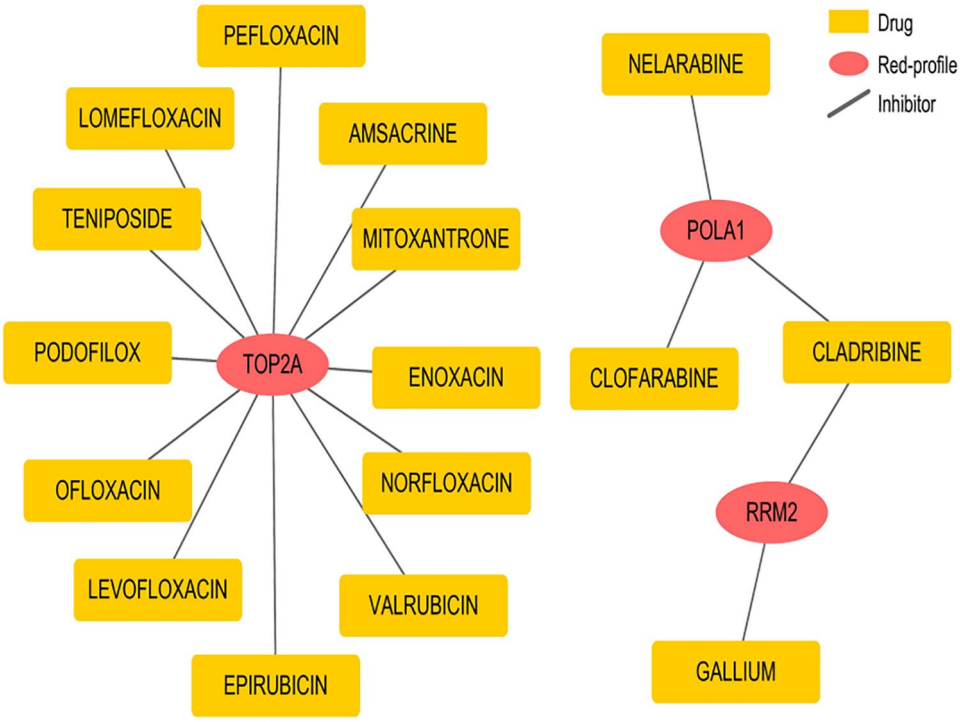

Figure 5. Drug-gene interaction network diagram.
DGIdb2.0 was used to predict drugs related to significant cluster genes, and 17 drug-gene interaction pairs for 3 red profile genes (TOP2A, RRM2, and POLA1) were obtained. Red circles: red profile genes; yellow squares: drugs; yellow triangles: TF.

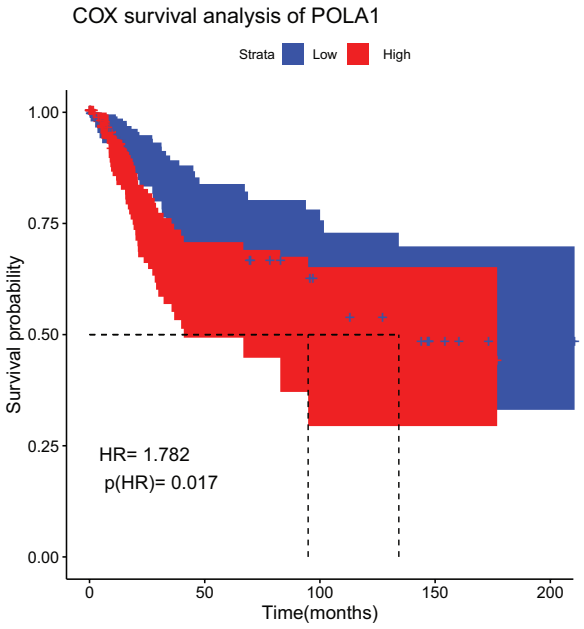

Figure 6. Survival curve of POLA1.
According to the median of POLA1 expression, the samples were divided into high and low expression groups. Kaplan-Meier (K-M) survival curve was generated. Results showed that POLA1 expression was significantly correlated with CC prognosis ($P = .017$). When 50% of the patients survived, the survival time was 134.2 months for patients in the high expression group and 94.9 months for those in the low expression group. Red: highly expressed; black: low expressed.

PCR validation

The expressions of TOP2A, RRM2, and POLA1 were analyzed in CIN1-3 lesions, CC tumor specimens, and normal control tissues, respectively. The results showed that the expressions of TOP2A, RRM2, and POLA1 were gradually upregulated in normal, CIN1-3, and CC tumor specimens, respectively (Figure 7), consistent with the time-series expression profiles analyzed using STEM.

Protein expression of target genes

The protein expressions of TOP2A, RRM2, and POLA1 relative to those of GAPDH were measured using Western blotting. As shown in Figure 8, the protein expressions of TOP2A, RRM2, and POLA1 exhibited an increasing trend in CC. The expressions of TOP2A, RRM2, and POLA1 were significantly upregulated in CC tumor tissues than in normal controls ($P < .01$ for all).

Discussion

CC is a common gynecological cancer affecting largely female patients, and it has the second highest mortality rate in China. At present, surgery and radiotherapy are the most commonly used therapies for patients with CC. However, the prognosis of patients with metastasis who underwent surgery and radiotherapy has not been ideal. In this study, CC-related gene expression data were downloaded and reanalyzed to explore the biomarker genes for CC. We anticipate that our findings will provide new perspectives for the treatment of CC.

Time-series analysis has been widely used in transcriptomics and proteomics.[28,29] In this study, we used time-series analysis to explore genes with similar expression trends during

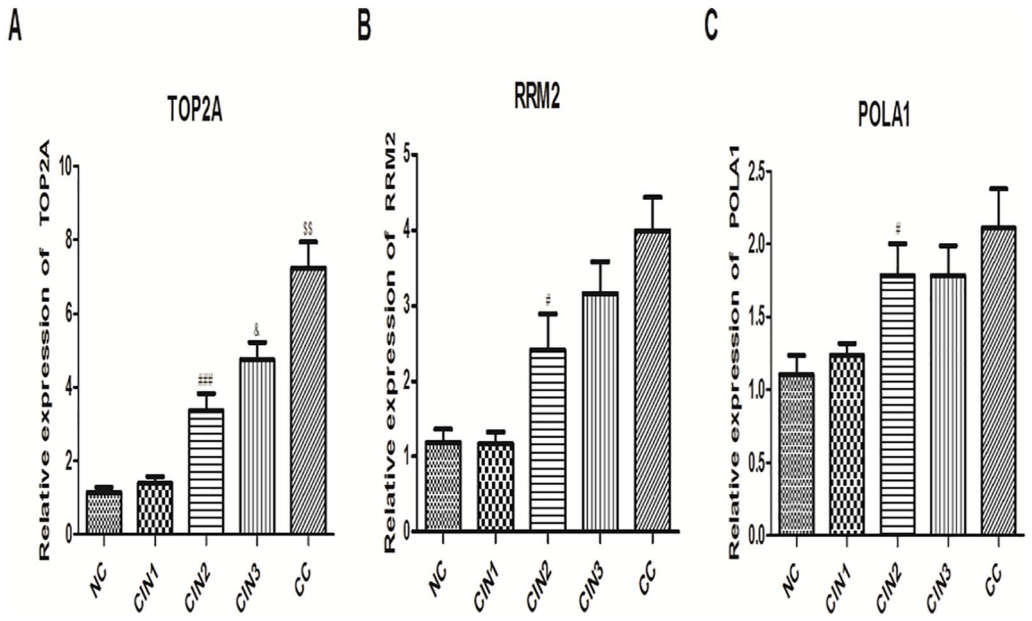

Figure 7. The expression of TOP2A, RRM2, and POLA1 in the multistep process of CC.
The mRNA expressions of TOP2A, RRM2, and POLA1 were evaluated in normal control, CIN1, CIN2, CIN3, and CC tumor samples. The expressions of genes were analyzed by real-time PCR analysis relative to the expression of GAPDH. (A to C) represent the expressions of TOP2A, RRM2, and POLA1, respectively. $^{\#}P < .05$, compared with CIN1; $^{\#\#\#}P < .001$, compared with CIN1 group, $^{\&}P < .05$, compared with CIN2 group, $^{\$\$}P < .01$, compared with CIN3 group.

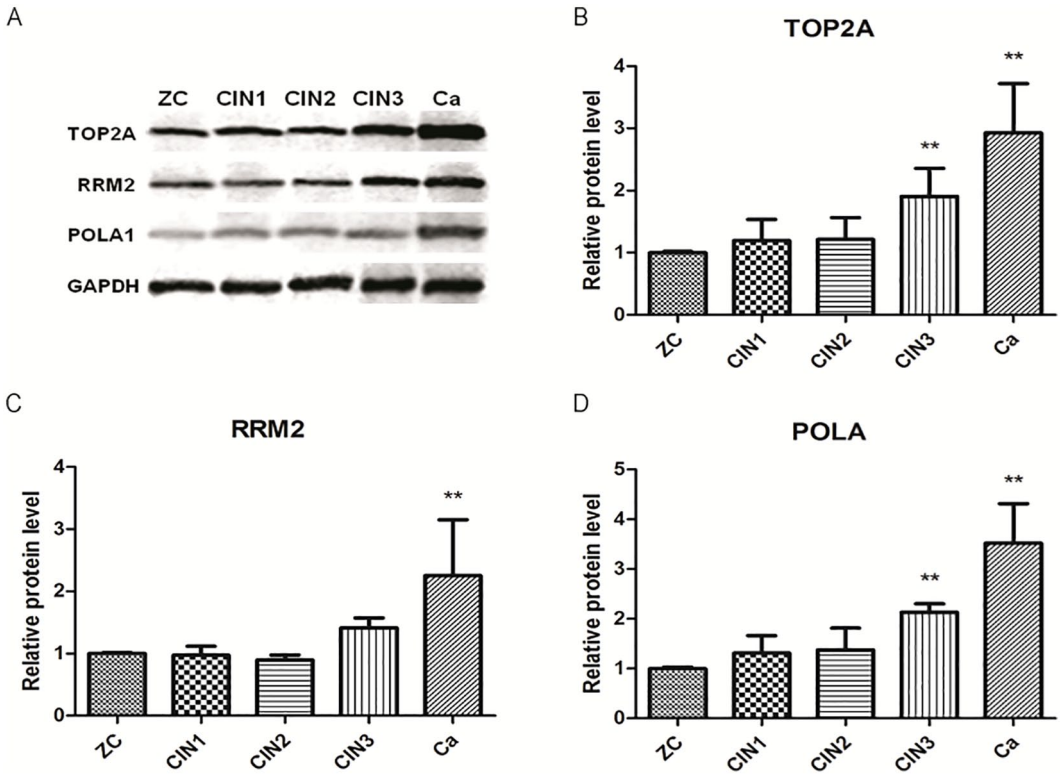

Figure 8. Protein expressions of TOP2A, RRM2, and POLA1 by Western blotting. Protein expression of the gene of interest was detected by Western blotting. The protein levels of TOP2A, RRM2, and POLA1 exhibited increasing expression during cervical carcinogenesis. All 3 proteins were significantly accumulated in CC tumor tissues. CC indicates cervical cancer.
$^{**}P < .01$, compared with normal controls.

different stages of CC development. Two clusters were obtained via the time-series analysis. Genes with increasing expression trends were significantly enriched in DNA replication-, cell division-, and cell proliferation-related GO functions. As reported in a previous study, genes related to DNA replication and cell proliferation were upregulated in CIN1/2 stage lesions and sustained in CIN3 stage tissues,[30] consistent with our findings. Thus, we suspected that changes in DNA replication, cell

division, and cell proliferation were significant events in the multistep process of CC.

Besides, in our study, the p53 signaling pathway was significantly enriched in red profile genes such as RRM2. The p53 signaling pathway has been proposed as a target for cancer therapy[31] and is related to the oncogenesis of squamous cell carcinomas of the head and neck induced by human papillomavirus (HPV) infection.[32,33] Previous evidence has shown that RRM2 is overexpressed and promotes angiogenesis in the development of HPV-associated CC.[34] In this study, RRM2 was also found in the red profile genes, the expression of which changed with the progression of CC, indicating that the genes clustered in the red profile are critical for CC. The discovery of a large number of drug targets, such as ALK, ROS1, c-met, PI3K, mTOR, and HSP90, has brought new hope to patients with cancer.[35-37] Therefore, gene target therapy for CC has become the focus of studies on CC treatment. Identification of biomarkers for diseases can help develop better diagnostic methods and improve clinical efficacy. Microarray data analysis has been widely used in the discovery of biomarkers for diseases.[38] In this study, 3 biomarkers related to drug-gene interactions and one biomarker related to CC prognosis were preliminarily extracted based on the microarray data downloaded from the GEO database. The effects of these biomarkers on the occurrence, development, and correct drug usage of CC were explained by bioinformatics. By constructing a PPI network and predicting gene-drug interactions, 3 key genes that are most closely related to CC were obtained (TOP2A, RRM2, and POLA1).

TOP2A plays a role in cell cycle progression and DNA desynchronization. Cells with TOP2A dysfunction exhibit slower proliferation, G2/M checkpoint arrest, and cell apoptosis. Brase et al[39] indicated that TOP2A RNA level is a good prognostic marker for breast cancer and is also associated with a favorable response to anthracycline-based therapy. In prostate cancer, TOP2A amplification is associated with androgen resistance and reduced survival.[40] Ratnam et al have reported that HPV E6/E7 oncogene expression can induce malignant cell transformation. TOP2A is a biomarker of S-phase cell cycle abnormalities induced by E6/E7 imbalance and activation. In CC, the expression of TOP2A is related to the clinicopathological parameters of patients and is correlated with their prognosis.[41] In this study, we found a drug-gene interaction between TOP2A and 12 drugs in CC tissues.

RRM2 encodes a small subunit of ribonucleotide reductase M2, which contains 2 presumed E2F binding sites in its promoter region. RRM2 protein is a component of ribonucleotide reductase, which is the key enzyme for reducing ribonucleotides.[42] RRM2 provides deoxyribonucleoside 5′-triphosphate (dNTP) for DNA synthesis in the S/G2 phase, which can be activated by ATR/CHK1/E2F3 signals and participates in DNA synthesis and repair.[43] The characteristics of RRM2 in promoting tumor progression are tightly associated with its capability to induce activities of various oncogenes.[44] Rasmussen et al[45] have shown that the high expression of RRM2 in patients with glioma can protect glioblastoma cells from endogenous replication pressure, DNA damage, and apoptosis, which is negatively related to the survival of those patients. Wang et al[34] found that HPVE7 induced the upregulation of RRM2, which then promoted cervical carcinogenesis via ROS-ERK1/2-HIF-1a-VEGF-induced angiogenesis. This study found drug-gene interactions between RRM2 and 3 drugs in CC tissues.

POLA1 is the first gene encoding DNA polymerase that has been isolated and purified. It is widely distributed in organisms and is suggested to be essential for DNA repair in most bacteria. Davidsen et al[46] found that POLA1 plays an important role in DNA repair in *Escherichia coli*. Fraser et al reported that cysteine residues in the POLA1 of *Treponema pallidum* were located in the 2 exonuclease domains, which were unique to *T pallidum* POLA1. Therefore, POLA1 could be used to develop a PCR-based diagnostic test for syphilis.[47] Current evidence shows that POLA1 is the target for retinoid-related molecules with preclinical antitumor activity in inhibiting the proliferation and promoting the apoptosis of cancer cells.[48] Results of the survival analysis showed that the high expression of POLA1 was positively correlated with the poor prognosis of patients with CC, and Western blotting determined the high expression of POLA1 in CC tissues. Thus, POLA1 can be suggested as a target for the treatment of CC.

A limitation of this study is that the data on gene expression profiles during CC development were obtained from only the GEO database, and there could be a lack of functional and experimental validation in cells and clinical samples. A large number of experimental and clinical studies are urgently needed soon.

In summary, changes in the expressions of TOP2A, RRM2, and POLA1 are correlated with CC progression. The 3 genes were predicted to be the targets of 16 drugs. TOP2A, RRM2, and POLA1 may be candidate targets for CC. Our study provides a theoretical basis for further investigating the mechanism of CC and gene target therapy. However, a large number of experimental and clinical studies are warranted in the near future.

Author Contributions

LY designed the research, collected the data, performed the statistical analysis, and wrote the manuscript. MW collected the data and performed the statistical analysis for the manuscript. FL revised the manuscript for important intellectual content.

ORCID iD

Lijun Yu https://orcid.org/0000-0003-2722-6467

Supplemental material

Supplemental material for this article is available online.

REFERENCES

1. Chaichian S, Khateri S, Moradi Y, et al. Trends in cervical cancer incidence in Iran from 2003 to 2009. *Middle East J Cancer*. 2018;9:57-63.
2. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. *Lancet*. 2019;393:169-182.

3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
4. Di J, Rutherford S, Chu C. Review of the cervical cancer burden and population-based cervical cancer screening in China. *Asian Pac J Cancer Prev*. 2015;16: 7401-7407.
5. Goh SH, Hong SH, Hong SH, et al. eIF3m expression influences the regulation of tumorigenesis-related genes in human colon cancer. *Oncogene*. 2011;30: 398-409.
6. Siegel EM, Riggs BM, Delmas AL, Koch A, Hakam A, Brown KD. Quantitative DNA methylation analysis of candidate genes in cervical cancer. *PLoS ONE*. 2015;10:e0122495.
7. Heselmeyer-Haddad K, Sommerfeld K, White NM, et al. Genomic Amplification of the Human Telomerase Gene (TERC) in pap smears predicts the development of cervical cancer. *Am J Pathol*. 2005;166:1229-1238.
8. Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543:378-384.
9. Deng S-P, Zhu L, Huang D-S. Predicting hub genes associated with cervical cancer through gene co-expression networks. *IEEE/ACM Trans Comput Biol Bioinform*. 2015;13:27-35.
10. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer*. 2010;10:550-560.
11. Holden M, Deng S, Wojnowski L, Kulle B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics*. 2008;24:2784-2785.
12. Hua F, Liu S, Zhu L, Ma N, Jiang S, Yang J. Highly expressed long non-coding RNA NNT-AS1 promotes cell proliferation and invasion through Wnt/β-catenin signaling pathway in cervical cancer. *Biomed Pharmacother*. 2017;92: 1128-1134.
13. Xiao W, Chen X, He M. Inhibition of the Jagged/Notch pathway inhibits retinoblastoma cell proliferation via suppressing the PI3K/Akt, Src, p38MAPK and Wnt/β-catenin signaling pathways. *Mol Med Rep*. 2014;10:453-458.
14. Fjeldbo CS, Aarnes EK, Malinen E, Kristensen GB, Lyng H. Identification and validation of reference genes for RT-qPCR studies of hypoxia in squamous cervical cancer patients. *PLoS ONE*. 2016;11:e0156259.
15. Kwee SA, Hernandez B, Chan O, Wong L. Choline kinase alpha and hexokinase-2 protein expression in hepatocellular carcinoma: association with survival. *PLoS ONE*. 2012;7:e46591.
16. Horn LC, Pippig S, Raptis G, et al. Clinical relevance of urokinase-type plasminogen activator and its inhibitor type 1 (PAI-1) in squamous cell carcinoma of the uterine cervix. *Aust N Z J Obstet Gynaecol*. 2002;42:383-386.
17. Barrett T, Edgar R. Gene expression omnibus (GEO): microarray data storage, submission, retrieval, and analysis. *Methods Enzymol*. 2006;411:352-369.
18. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*. 2006;7:191.
19. Consortium GO. The gene ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32:D258-D261.
20. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003;4:P3.
21. Eilbeck K, Lewis SE, Mungall CJ, et al. The sequence ontology: a tool for the unification of genome annotations. *Genome Biol*. 2005;6:R44.
22. Greiner M, Pfeiffer D, Smith R. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med*. 2000;45:23-41.
23. Gralnick JA, Downs DM. The YggX protein of *Salmonella enterica* is involved in Fe(II) trafficking and minimizes the DNA damage caused by hydroxyl radicals: residue CYS-7 is essential for YggX function. *J Biol Chem*. 2003;278: 20708-20715.
24. Damian S, Andrea F, Stefan W, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43:D447-D452.
25. Kellman P, Chung YC, Simonetti OP, Mcveigh ER, Arai AE. Multi-contrast delayed enhancement provides improved contrast between myocardial infarction and blood pool. *J Magn Reson Imaging*. 2010;22:605-613.
26. Wagner AH, Coffman AC, Ainscough BJ, et al. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res*. 2016;44:D1036-D1044.
27. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res*. 2010;1:274-278.
28. Borirak O, Rolfe MD, de Koning LJ, et al. Time-series analysis of the transcriptome and proteome of *Escherichia coli* upon glucose repression. *Biochim Biophys Acta*. 2015;1854:1269-1279.
29. Derlindati E, Dei Cas A, Montanini B, et al. Transcriptomic analysis of human polarized macrophages: more than one role of alternative activation? *PLoS ONE*. 2015;10:e0119751.
30. den Boon JA, Pyeon D, Wang SS, et al. Molecular transitions from papillomavirus infection to cervical precancer and cancer: role of stromal estrogen receptor signaling. *Proc Natl Acad Sci USA*. 2015;112:E3255-E3264.
31. Stegh AH. Targeting the p53 signaling pathway in cancer therapy—the promises, challenges and perils. *Expert Opin Ther Targets*. 2012;16:67-83.
32. Munger K, Baldwin A, Edwards KM, et al. Mechanisms of human papillomavirus-induced oncogenesis. *J Virol*. 2004;78:11451-11460.
33. Chakravarthy A, Henderson S, Thirdborough SM, et al. Human papillomavirus drives tumor development throughout the head and neck: improved prognosis is associated with an immune response largely restricted to the oropharynx. *J Clin Oncol*. 2016;34:4132-4141.
34. Wang N, Zhan T, Ke T, et al. Increased expression of RRM2 by human papillomavirus E7 oncoprotein promotes angiogenesis in cervical cancer. *Br J Cancer*. 2014;110:1034-1044.
35. Ou Tan J, Yen Y, Soo RA. ROS1 as a "druggable" receptor tyrosine kinase: lessons learned from inhibiting the ALK pathway. *Expert Rev Anticancer Ther*. 2012;12:447-456.
36. Friedman JA, Wise SC, Hu M, et al. HSP90 inhibitor SNX5422/2112 targets the dysregulated signal and transcription factor network and malignant phenotype of head and neck squamous cell carcinoma. *Transl Oncol*. 2013;6:429-441.
37. Chen M-H, Chiang K-C, Cheng C-T, et al. Antitumor activity of the combination of an HSP90 inhibitor and a PI3K/mTOR dual inhibitor against cholangiocarcinoma. *Oncotarget*. 2014;5:2372-2389.
38. Cooper CS, Campbell C, Jhavar S. Mechanisms of disease: biomarkers and molecular targets from microarray gene expression studies in prostate cancer. *Nat Clin Pract Urol*. 2007;4:677-687.
39. Brase JC, Schmidt M, Fischbach T, et al. ERBB2 and TOP2A in breast cancer: a comprehensive analysis of gene amplification, RNA levels, and protein expression and their influence on prognosis and prediction. *Clin Cancer Res*. 2010;16:2391-2401.
40. Murphy AJ, Hughes CA, Barrett C, et al. Low-level TOP2A amplification in prostate cancer is associated with HER2 duplication, androgen resistance, and decreased survival. *Cancer Res*. 2007;67:2893-2898.
41. Ratnam S, Fontaine D, Coutlee F, et al. O63 HPV E6/E7 mRNA and MCM2/TOP2a testing in cervical cancer screening: results from a multicentre Canadian study. *Int J Antimicrob Agents*. 2009;34:S22-S23.
42. Stubbe J. Ribonucleotide reductases. *Annu Rev Biochem*. 2006;75:681-706.
43. Gong C, Liu H, Song R, et al. ATR-CHK1-E2F3 signaling transactivates human ribonucleotide reductase small subunit M2 for DNA repair induced by the chemical carcinogen MNNG. *Biochim Biophys Acta*. 2016;1859:612-626.
44. Qi H, Lou M, Chen Y, et al. Non-enzymatic action of RRM1 protein upregulates PTEN leading to inhibition of colorectal cancer metastasis. *Tumour Biol*. 2015;36:4833-4842.
45. Rasmussen RD, Gajjar MK, Tuckova L, et al. BRCA1-regulated RRM2 expression protects glioblastoma cells from endogenous replication stress and promotes tumorigenicity. *Nat Commun*. 2016;7:13398.
46. Davidsen T, Rodland EA, Lagesen K, Seeberg E, Rognes T, Tonjum T. Biased distribution of DNA uptake sequences towards genome maintenance genes. *Nucleic Acids Res*. 2004;32:1050-1058.
47. Fraser CM, Eisen J, Fleischmann RD, Ketchum KA, Peterson S. Comparative genomics and understanding of microbial biology. *Emerg Infect Dis*. 2000;6:505-512.
48. Han T, Goralski M, Capota E, et al. The antitumor toxin CD437 is a direct inhibitor of DNA polymerase α. *Nat Chem Biol*. 2016;12:511-515.