Transcriptome sequencing reveals pathways and genes dysregulated in HPV infected cervical cancer

Qianxin Chen¹, Jiajia Hu¹, Xingwang Sun², Wenbo Long²

¹Basic Medical College, Southwest Medical University, 646099 Sichuan, China
²Pathology Department of the First Affiliated Hospital, Southwest Medical University, 646099 Sichuan, China

Summary

Objective: Cervical cancer is the fourth leading cause of cancer mortality in women worldwide. Cervical cancer is predominately caused by chronic infection with high-risk human papillomavirus (HR-HPV) genotypes. While several oncogenes and tumor suppressors are implicated in the initiation and progression of cervical cancer, the complicated genetic network regulating cancer is largely unknown. We aimed to study these oncogenes and tumor suppressors. Materials and methods: We extracted mRNA from paraffin-embedded tissues derived from cervical cancers infected with HPV-16, HPV-58, HPV-52, HPV-33 and HPV-35 and normal controls. Transcriptome sequencing was undertaken to analyze the differentially expressed genes (DEGs) between cancer and normal tissues. Results: Transcriptomic analysis screened 10,025 DEGs between cancerous and normal tissues (5,419 upregulated and 4606 downregulated). In KEGG analysis, most of the annotated DEGs were enriched in four sub-categories involved in MAPK, mTOR, PI3K-Akt, and Ras signaling pathways. And most of the key genes in these pathways were dysregulated at the mRNA level, including FGFR3, PDK2, PDK3, Akt1, Ras, MAPK1, MAPK3 and mTOR. The GO classification, KEGG enrichment and DEGs profiles of HPV-58 infected samples aligned closer to those of HPV-16 infected but were different from those of tissues infected with the other HPVs. Conclusions: These results suggest that the genesis of cervical cancer is associated with gene expression changes in specific cancer related signaling pathways. Thus, developing biomarkers and targets from these pathways may aide the diagnosis and targeted treatment of cervical cancers.

Key words: Cervical cancer; Human papillomavirus; Transcriptome sequencing; Differentially expressed genes; Signaling pathway.

Introduction

Though cervical cancer is preventable, there are an estimated 530 thousand newly diagnosed cases annually worldwide [1]. Cervical cancer is predominately caused by chronic infection with high-risk human papillomavirus (HR-HPV) genotypes. Although many well-established screening tests and intervention methods are efficacious in reducing the incidence of cervical cancer, many cancers are still diagnosed and are hard to treat. Currently, the palliative platinum-based chemotherapy (CT) is the standard first-line treatment for metastatic/recurrent cervical cancer, but the prognosis is poor [2]. Meanwhile, target therapy has been proved to be an effective method to increase the survival rate of cancer patients [3]. Furthering our understanding of the cancer genome and identifying novel therapeutic targets could facilitate the treatment of advanced cervical cancer.

With the development of next-generation sequencing (NGS), RNA sequencing (RNA-Seq) has become a powerful and cost-effective tool for analyzing cancer transcriptomes and for screening for differentially expressed genes (DEGs), alternative splicing, gene mutations, IncRNA, and novel transcripts with high accuracy and efficiency [4]. Compared to the original microarray-based method, RNA-seq offers less background noise and richer data for analysis. RNA-Seq has thus been used to screen the expression profiles of many cancers including breast cancer, lung cancer, and et al. [5-7]. Nevertheless, the transcriptome landscape of squamous cell cervical cancer from paraffin-embedded tissues has yet to be elucidated.

The malignant transformation of cervical epithelial cells is caused by HR-HPV proteins interacting with the host genome. Differences in the regional prevalence of HR-HPV genotypes generally account for the geographical variance in incidence [8]. Our group has previously shown that, despite being the second-most prevalent HPV genotype worldwide, HPV-18 is uncommon within the Sichuan region of China. The present study therefore aimed to identify DEGs that can be used as molecular markers for diagnosing or prognosing malignant lesions detectable in paraffin embedded tissue. We herein undertook transcriptomic analysis of cervical cancer tissues infected by different genotypes of HR-HPV infection and normal tissues.

Materials and Methods

Clinical samples collection

Cervical squamous cancer samples and normal cervical samples were collected from female patients at the First Affiliated Hospital of Southwest Medical University during 2017-2018. Normal samples describe those who are negative for intraepithelial lesion or malignancy, or other severe lesions. Exclusion criteria encompassed previous operation for cervical disease (including the loop electro-
surgical excision procedure (LEEP), cold-knife conization, hysterectomy, laser therapy, chemotherapy/radiation treatment, pregnancy, and malignant diseases outside the cervix. All samples were screened for HPV infection and genotyping by qRT-PCR using a Slan-96P Real-time PCR System (Hongshi Medical Technology Co. Ltd. Shanghai, China). HPV-16 was the most prevalent genotype in cervical squamous cancer in this area and HPV-58 the second, while HPV-33, HPV-35 and HPV-52 were not very prevalent in cervical squamous cancer in southwestern China [9]. We selected paraffin-embedded samples with these HR-HPV infections to perform the following experiments. The study set (Supplemental Table S1) consisted of 30, 30, 8, 8, 4 cervical cancer samples infected with HPV-16, HPV-58, HPV-33, HPV-35 and HPV-52 single HR HPVs, respectively, and 30 normal cervical epithelium (NCE) samples without HPV infection were collected for RNA extraction.

**RNA extraction**

Total RNA was extracted from cancer and normal tissues red using a total RNA extraction kit for paraffin-embedded tissue sections (TinGen biochemical technology (Beijing) Co., Ltd., Item: DP439) according to the manufacturer’s protocol. The RNA quality was determined using an AmoyDx SMA4000 Bioanalyzer. Only RNA extracts with a 28S:18S ratio > 1.8 and an OD range 1.95 to 2.05 were used for transcriptome analysis.

**RNA mixed pool preparation**

To increase the amount of RNA and reduce intersequence variance, RNA extracts were mixed into several pools. For sample groups infected with HPV-16/HPV-58, and NCE group, every ten RNA extracts from the total of thirty samples were mixed into a pool. Thus, a total of three pools were created for any of the three groups. For samples from patients infected with HPV-33/HPV-35/HPV-52, RNA extracts were mixed to create a pool with a ratio of 4:4:2 (HPV-33/HPV-35/HPV-52, respectively), and a total of two pools were obtained. These pools were submitted for transcriptome sequencing. Before library construction, all RNA materials were treated with DNase I, and ribosomal RNA (rRNA) was removed by using an Illumina Truseq Standed Total RNA LT- (with Ribo-Zero TM hu-man/Mouse/Rat) kit according to the manufacturer’s guide-lines.

**Transcriptome sequencing and data analysis**

The NEB NextUltr RNA Library Prep Kit for Illumina was employed to generate sequencing libraries according to the manufacturer’s recommendations. Fragmentation was carried out using divalent cations under elevated temperature. The first-strand cDNA was synthesized using random hexamer primers and M-MLV Reverse Transcriptase. The Second-strand cDNA synthesis was then performed using DNA Polymerase I and RNase H. NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridization, after adenylation of the 3’ ends of DNA fragments. The library fragments were purified with the AMPure XP system. PCR (polymerase chain reaction) was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Then, the products were purified, and library quality was assessed on an Agilent Bioanalyzer 2100 system. The clustering of the indexed samples was performed on a Cbot Cluster Generation System according to the manufacturer’s instructions. Finally, the library preparations were sequenced on an Illumina X10 (Illumina, Inc., San Diego, CA, USA).

After sequencing, cutadapter was used to process raw data. Unqualified sequences were excluded. We then excluded I) adapter reads II); reads containing more than 5% N (where N is undetermined base information; and III) low-quality reads (base number of Q < 10 accounts for more than 20% of the whole read). We thereafter analyzed the original sequencing quantity, effective sequencing quantity, Q20, Q30, GC content, and comprehensive evaluation.

**Function annotation of differentially expressed genes (DEGs)**

The function of DEGs, which were two-fold dysregulation in cases relative to controls, was annotated from the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Query UniGenes were matched with the subject sequences in the multiple databases using BLAST. Significant DEGs were enriched for each term in the GO database; the number of genes in each term was calculated and GO items that were significantly enriched with significant DEGs compared with the whole genome background was identified using hypergeometric test. The classification of DEGs into the functional pathways was conducted by KEGG analysis. The KEGG automatic annotation server was used for KEGG Orthology (KO) and KEGG pathway annotation (www.genome.jp/kegg).

**Quantitative reverse transcription (qRT)-PCR analysis**

The mRNAs were reverse-transcribed using a M-MLV Reverse Transcriptase Kit according to the manufacturer’s instructions (M1701 Promega, USA). SYBR Green-based qPCR was then performed on an ABI (Applied Biosystems, USA) ViiA 7 Real-Time PCR system. Beta-actin was used as the endogenous control and all reactions were performed in triplicate. Relative gene expression was calculated using the comparative cycle threshold \(2^{-\Delta\Delta CT} \) method. PCR cycling conditions consisted of 2 min at 95 °C, followed by 45 cycles of 15 s of denaturation at 95 °C and 1 min of annealing and extension at 60 °C. Statistical analysis for qRT-PCR results was conducted by t-test and \( p < 0.05 \) was considered significant.

**Results**

**Transcriptome sequencing**

A total 164.42 GB clean data were acquired after removing reads with adapters, unknown nucleotides and low-quality reads, with an average of 14.95 GB (ranging
Figure 1. — Clustering and scattering of 10025 differentially expressed genes. (A) Clustering of 10025 genes with differential expression between invasive cervical carcinomas (ICC) and normal cervical epithelium (NCE) using a twofold threshold. The cluster is color coded using red for up-regulation, green for downregulation, and black for median expression. For ICC samples infected with HPV-16 and HPV-58, and NCE samples, every ten RNA from each were mixed to make a pool. For ICC samples infected with HPV-33, HPV-35 and HPV-52, RNA was pooled in a ratio of 2 : 2 : 1, and a total of two pools were made. (other HPVs pool). The value in this map was the \( \log_{2} \frac{\text{ICCpool}}{\text{NCEpool}} \) for each unigene. (B) Venn diagram analysis illustrating shared differentially expressed genes (DEGs) among different HPV infected groups.

13.09 to 17.20 GB). Average of the Q30 percentage was 93.45%, ranging from 90.72 to 97.61%. Among the ten sequenced samples, 70.45% (ranging 66.67 to 77.83%) gene transcripts were sequenced with > 10-times coverage and 56.47% (ranging 44.60 to 69.46%) with > 30-times coverage. From these samples, over 80 thousand mRNAs were identified as being expressed in at least one sequenced sample. With the selected criteria \( |\log_{2}FC| > 1, p < 0.05 \), a total of 10,025 DEGs were identified; 5,419 of which were upregulated and 4,606 were downregulated (Supplemental Table S2, Figure 1).

**GO enrichment**

In order to compare the differences among HPV-16, HPV-58 and the other HPV-infected samples, DEGs from the three pools were analyzed by GO enrichment. A total of 8,835, 8,838, and 7,580 UniGenes were sequenced from HPV-16, HPV-58 and the other HPV-infected (HPV-33, HPV-35 and HPV-52 infected) samples, respectively.

The biological process, molecular function, and cellular component for these genes were annotated using the GO databases (Figure 2). In the biological process category, 5,898 (65.56%), 5,831 (65.98%) and 4,873 DEGs (64.31%) from the three types of samples, respectively, were distributed in cellular processes. In the molecular function category, genes distributed in the protein binding sub-category accounted for the highest proportion of total annotated DEGs (HPV-16 55.46%, HPV-58 55.50%, and the other HPVs 56.75%). In the cellular component category, the genes involved in the cell or cellular part sub-categories accounted for the greatest proportion of total DEGs (HPV-16 91.66%, HPV-58 91.75%, the other HPVs 87.61%).

**KEGG pathway enrichment**

We consulted the KEGG database to analyze the functions and characteristics of 3,690, 3,690, and 3,038 unique DEGs sequenced in HPV-16, HPV-58, and the other HPV-infected samples, respectively. These were classified into
Transcriptome sequencing reveals pathways and genes dysregulated in HPV infected cervical cancer

Figure 2. — GO classification. Annotation statistics of differentially expressed genes in the secondary node of GO. The horizontal axis shows the secondary nodes of three categories in GO. The vertical axis shows the number of annotated genes.

112 pathways ($p < 0.05$). A summary of 62 KEGG pathways ($p < 0.01$) is presented (Figure 3). Among these were the pathways dysregulated during cancer, which accounted for the largest number of DEGs of all human disease categories. The other notable DEGs-enriched pathways in the same category were associated with HPV and Epstein-Barr (EB) virus infections. In the environmental information process category, UniGenes in four related sub-categories accounted for most of the annotated DEGs in all samples. Notably, these sub-categories included the MAPK, mTOR, PI3K-Akt, and Ras signaling pathways. In cellular processes category, actin cytoskeleton and focal adhesion were the most DEGs enriched pathways.

**qRT-PCR verification**

To confirm the results of transcriptome sequencing, four well-known dysregulated genes in cervical cancer were chosen to analyze, including CDKN1A, CDKN2A, KRT4, and KRT17. The mRNA levels of CDKN1A, CDKN2A and KRT17 were increased in the RNAseq analysis, but that of KRT4 decreased. The qRT-PCR validation confirmed that CDKN1A, CDKN2A and KRT17 were upregulated and KRT4 was downregulated in cancer samples relative to normal cervical tissues (Figure 4). In the KEGG analysis, four cancer-related pathways including the MAPK, mTOR, PI3K-Akt, and Ras signaling pathways were enriched with DEGs. Thus, we used qRT-PCR to analyze the mRNA levels of key genes in these pathways (Figure 4). The qRT-PCR results indicated most of the key genes in PI3K-Akt-mTOR signaling and Ras-Raf-MAPK signaling ($FGFR3$, $PKD2$, $PDK3$, $Akt$, $mTOR$, $NRas$, $HRas$, $Raf1$, $MAPK1$, $MAPK3$) were upregulated in cancer samples relative to normal tissues.

**Discussion**

The initiation and progression of HPV-induced cervical cancer is a complicated process comprising of various dynamic changes in the genome. These genetic variations contribute to the formation of malignant cells which present favorable genotypes through genetic mutations, epigenetic modifications and differential gene expression [10, 11]. Thus, genetic mutations and downstream dysregulation in genes are hallmarks of high mutability in cancer tissues. We used NGS to study the variations of gene expression in paraffin-embedded cervical tissues. Though the processes of fixation, processing, embedding and storage on samples by paraffin could reduce the quality and length of RNA, RNA-seq remains an effective method for both prospective and retrospective analyses of archived paraffin-embedded samples [12]. We have demonstrated through deep transcriptome sequencing of paraffin-embedded samples screened that 10,025 DEGs between cancerous and normal tissues. These DEGs could provide quantitative estimates for known genes species with the potential to discover low frequency dysregulated genes [13].

Pathways related to tumorigenesis and cancer progression accounted for the largest number of DEGs in the human disease sub-category. The second in the same sub-category was HPV infection, which is aligns with our understanding that chronic infection with HR-HPV is the cause of cervical cancer [14]. It is known HR-HPV infection induces
genetic variations and gene expression changes. For example, CDKN1A (p21), CDKN2A (p16) and KRT17 expression were increased in squamous cervical cancer and high expressed in part of tumorous tissues [15-17], while KRT4 expression was reduced [17]. Our present study confirmed these results by transcriptome sequencing and qRT-PCR analysis.

By annotating the function of the DEGs in the KEGG database, our findings suggest that four pathways were enriched with most of the DEGs, including MAPK signaling pathway, mTOR signaling pathway, PI3K-Akt signaling pathway and Ras signaling pathway. PI3K/Akt and mTOR pathway regulate many biological processes and correlated with cell proliferation and metastasis in cervical cancer [18]. Many key nodes in PI3K/Akt signaling are proto-oncogenes or tumor suppressors. Dysregulation of these genes is associated with the increased or decreased occurrence of tumor [19]. PI3K activity was implicated in cell transformation and tumor progression in diverse cancers, but the mRNA levels of key subunits PIK3CA and PIK3R2 in this study were not obviously changed, different from its antagonizing factor PTEN which mRNA level was significantly increased. That was consistent with previous reports that PI3K/PTEN expression was frequently deregulated in a majority of malignancies [20]. PDK and Akt are downstream factors of PI3K. PDK can phosphorylate and activate Akt, and the activated Akt can modulate the function of many substrates to regulate cell apoptosis and survival [20-22]. The data from the present study screened upregulated mRNA levels of PDK2, PDK3 and Akt1 that may contribute to the activation of the PI3K/Akt pathway.

Ras-Raf-MAPK signaling regulates various cellular functions and is closely associated with tumorigenesis. Indeed, Ras is the most frequently mutated oncogene in human cancer for which there exists three types: K-Ras, H-Ras and N-Ras [23]. Dysregulation of K-Ras is reported in
Transcriptome sequencing reveals pathways and genes dysregulated in HPV infected cervical cancer

Figure 4. — Transcriptome sequencing confirmation of differentially expressed genes by qRT-PCR. (A-D) qRT-PCR were performed for four genes that have been reported to be differentially expressed between cervical cancer and normal cervical epithelium (NCE). The expression level of each gene was normalized to the level of β-actin. (A) CDKN1A; (B) CDKN2A; (C) KRT17; (D) KRT4. (E-G) qRT-PCR were performed to confirm the dysregulated signaling nodes in the pathways where DEGs enriched. The value was the Log$_2$ cervicalcancer/NCE for each unigene. (E) Nodes in PI3K/Akt pathway; (F) Nodes in Ras-Raf-MAPK signaling; (G) Nodes in mTOR pathway. Statistical analysis was conducted by t-test and $p < 0.05$ was considered significant.
many malignancies. In cervical cancer, genetic alterations in the K-Ras gene were associated with radiation resistance [24]. But the mRNA level of K-Ras was similar in normal tissues, CIN and cervical cancer, different from those of H-Ras and N-Ras which were significantly increased in cancerous samples [25]. Our study found upregulated transcripts of H-Ras and N-Ras and downregulated transcript of K-Ras in HR HPV infected cervical cancer, which generally align with the previous reports. After Ras activation, Raf is recruited to the cell membrane and associates with Ras to activate MEK1 and MEK2. MEK1 and MEK2 thereafter phosphorylate threonine and tyrosine residues of the Erk1 and Erk2 proteins associated with cell proliferation, survival, metabolism, migration, and transcription [26, 27]. The present study shows that Raf1, MAPK1, MAPK3, and MAPK20 were significantly upregulated in cancer samples, thereby indicating the Ras-Raf-MAPK pathway may be transcriptionally activated in cervical cancers.

The mammalian target of rapamycin (mTOR) is a critical effector in cell-signaling pathways commonly dysregulated in cancers. The activation of mTOR complexes is closely associated with PI3K/Akt and Ras-Raf-MAPK signaling pathways. For example, activated Akt can promote the activity of mTOR complex 1 (mTORC1) by phosphorylating TSC1/2 factors which are characterized as tumor-suppressor [28]. Positive growth signals from the Ras-Raf-MAPK pathway also inhibit TSC1/2 factors and the mTOR complex during the activation of mTOR [29]. Our study revealed the mRNA levels of many positive factors in the PI3K/Akt and Ras-Raf-MAPK pathways were upregulated. Doing so could facilitate the activation of Akt and mTORC1 and then the mTOR complex. The transcription activity of RPS6KB1 encoding S6K1 protein downstream of mTORC1 was upregulated. While the mRNA levels of ATG13 and ULK1 encoding key factors in the autophagy pathway were also upregulated, though the activity of their protein products is negatively regulated by mTORC1 [29], revealing other factors may exist to regulate autophagy besides mTOR. Recently, a relevant study by Babion et al. indicated that mTOR signaling was one of the three altered pathways in HPV-induced carcinogenesis, in which dysregulation was induced by the copy number changes of respective genes [30]. This provides a novel strategy to study the dysregulated expression of key genes in tumors.

Different HPVs show diverse potentiality to induce carcinogenesis. Though HPV-16 and HPV-18 are the most carcinogenic worldwide, HPV-58 and HPV-52 were very prevalent in Asia, Africa and some regions of the world [8]. We screened and compared the transcriptome characteristics among HPV-16, HPV-58 and the other HPVs infected paraffin-embedded samples. It revealed the numbers of DEGs identified from other-HPVs infected samples were smaller than those from HPV-16 or HPV-58 infected samples. The GO classification, KEGG analysis and DEGs profiles of HPV-58 infected samples aligned closely to those of samples infected with HPV-16 and differed from those infected with the other HPVs. Indeed, HPV-58 was the second most prevalent HPV in southwestern China, which frequency in CIN3+ lesions was much higher than those of HPV-52, 33 and 35 [9, 31]. Considering the prevalence of HPVs significantly varied from region to region, we estimated the women in southwestern China may be more susceptible to HPV-58 than HPV-52, 33 or 35, thus suffered more serious dysregulation of gene expression in the HPV-infected tissues.

In this study, we identified four pathways including MAPK, mTOR, PI3K-Akt, and Ras signaling pathway and many key genes differentially expressed in cervical cancer tissues; notably Akt1, N-Ras, H-Ras, and mTOR. Samples with HPV-16 or 58 infection were induced more dysregulated genes than those with the other HPVs infection. The present study may provide novel mRNA biomarkers for diagnosis of cervical lesions or potential therapeutic targets for cancer treatment. Further studies should validate the biological functions of the identified pathways and the practical uses of these as candidate biomarkers.

Authors’ contributions

Dr. Wenbo Long designed the research, analyzed the data, wrote the manuscript and edited the manuscript. Qianxin Chen performed the experiments and collected the specimens. Jiajia Hu performed the qPCR experiments. Qianxin Chen performed the experiments and collected the data, wrote the manuscript and edited the manuscript.

Ethics approval and consent to participate

The project was examined by the ethics committee of the Affiliated Hospital of Southwest Medical University. The research content and design of the project are in line with the ethical norms. All the authors declare that they have no conflict of interest.

Funding

Funding in this work were received from the following: PhD Research Fund of the Affiliated Hospital of the Southwest Medical University (grant no. 16238); Luzhou city: Southwest Medical University Joint Project (grant no. 2018LZXNYD-ZK22); Scientific Research Fund of Southwest Medical University (grant no. 2017-ZRQN-088)

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Acknowledgments

We would like to express my gratitude to all those who helped us during the writing of this manuscript.
Conflict of interest

The authors declare that they have no competing interests.

Submitted: November 05, 2019
Accepted: May 26, 2020
Published: December 15, 2020

References


Corresponding Author: WENBO LONG, Ph.D.
Taiping Street in Sichuan province, Luzhou City, Jiangyang District No. 25, China
e-mail address: wenbolong@swmu.edu.cn