The dysregulation of microarray gene expression in cervical cancer is associated with overexpression of a unique messenger RNA signature

Seyedeh Zahra Mousavi¹, Vahdat Poortahmasebi¹, Talat Mokhtari-Azad¹, Shohreh Shahmahoodi¹, Mohammad Farahmand¹, Mahdieh Farzanehpour¹, Somayeh Jalilvand¹∗

¹Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
2Infectious and Tropical Diseases Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
3Department of Bacteriology and Virology, School of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

ABSTRACT

Background and Objectives: Human papillomavirus (HPV) is the fourth most common cause of cervical cancer (CC). The aim of the present study was to investigate gene expression levels of previously identified transcriptional signatures for malignant and non-malignant CC.

Materials and Methods: To validate of previously analyzed microarray gene expression data, we selected two hub genes (CDK1 and PLK1) and four differentially expressed mRNAs that were common in pre-malignant-normal and malignant-pre-malignant networks (SMS, NNT, UHMK1 and DEPDC1). To this purpose, the study included women diagnosed histologically with malignant CC (n=15), pre-malignant (n=15), and normal subjects (n=15). The expression of six host genes and viral E6/E7 genes were measured by quantitative Real-Time PCR.

Results: The results showed higher expression of CDK1/PLK1 hub genes and SMS, NNT and UHMK1 genes in malignant CC group than non-malignant CC group and normal group. A positive correlation was observed between gene expression of viral E6/E7 oncogenes and UHMK1 gene.

Conclusion: Dysregulation of several mRNA signatures are a common feature of CC and can be potentially used as diagnostic and prognostic biomarkers as well as can be applied to therapeutic targets for CC treatment.

Keywords: Human papillomavirus; Cervical cancer; Microarray; Gene expression

INTRODUCTION

Cervical cancer (CC) is the fourth most common cancer among women worldwide, as 570000 new cases were reported in 2018 that accounting for 6.6% of all female cancers (1). Human papillomavirus (HPV) is the key player for the development of CC (2). Although 40 HPV types can infect the genital tract, only 14 types are frequently found in CC, in-
process. Serine/threonine kinases Polo-like kinase 1 (PLK1) and cyclin-dependent kinase 1 (CDK1) were the most important hub genes in the protein-protein interaction (PPI) network (7). These proteins play critical roles in dysregulation of the cell cycle and consequently could lead to cervical cancer. These hub genes may be used as biomarkers for the diagnosis and prognosis as well as treatment of this cancer.

In our previous study, microarray gene expression data were obtained from the Gene Expression Omnibus (GEO) database with the accession number GSE67522. This dataset consist of 40 cervical tissue samples evaluating gene expression in three individual arrangements including histologically normal (HPV negative, n=12), HPV 16 positive non-malignant lesions (n=10), and HPV 16 positive invasive cervical cancer (n=18). By comparing gene expression profile of normal subjects with non-malignant patients, 208 DEGs were uncovered. Moreover, 221 genes were differentially expressed between non-malignant and malignant patients. For evaluation of common genes in two aforementioned networks, Venn diagram analysis revealed that four common genes including spermine synthase (SMS), nicotinamide nucleotide transhydrogenase (NNT), U2AF homology motif (UHM) kinase 1 (UHMKI), and DEP domain-containing protein 1A (DEPDC1) found in two groups. SMS, NNT and UHMKI genes have been upregulated in malignant-non-malignant network in comparison with non-malignant HPV-normal network. However, these results were obtained from in silico analysis and validation of these results are needed. To support this notion, in the present study, the expression levels of PLK1/CDK1 as hub genes and four above-mentioned gene signatures have investigated among Iranian patients with cervical cancer.

MATERIALS AND METHODS

Study population. Forty-five fresh uterine cervix biopsies, including 15 normal samples, 15 precancerous specimens, and 15 cervical cancer samples, were collected after hysterectomy at Imam Khomeini Complex Hospital (Tehran, Iran) and kept in RNALater (Qiagen) at -70°C to stabilize RNA. None of patients with cervical cancer received any chemotherapy or radiotherapy. All samples were collected following ethical committee approval of Tehran University of Medical Sciences (Approval Code: IR.TUMS.REC.1396.293). The informed consent was obtained for all participants before surgery. All samples were stained with hematoxylin–eosin on 5μm paraffin sections and evaluated by the pathologist.

Detection of HPV genome and genotyping. DNA was extracted with the High Pure Viral Nucleic Acid Kit according to manufacturer’s instruction (Roche, Germany). The integrity of the extracted DNA was evaluated using PCR to amplify a 110 bp segment of human β-globin gene by PC03/PC04 primers. The HPV genome was detected using a nested-PCR by two sets of primers, including MY09/MY11 and GP5+/GP6+ to target a 450 bp and 150 bp region of L1 gene, respectively (Table 1). The PCR products were run on a 1.5% agarose gel. All HPV positive samples were sequenced using the 3130 Genetic Analyzer (Genetic Analyzer ABI- 3130 DNA Sequencer, Foster city, CA, USA). The sequences were edited by the BioEdit software Package (https://bioedit.software.informer.com/7.2/). GHVP genotypes were characterized by phylogenetic analysis with reference sequences, which were obtained from the GenBank (https://www.ncbi.nlm.nih.gov/genbank/).

RNA extraction from cervical tissue samples. Total cellular RNA was extracted using the TRIZOL reagent (Ambion, Life Technologies, Carlsbad, California, USA) according to the protocol provided by the manufacturer. RNA concentrations were quantified spectrophotometrically at 260/280 nm using a NanoDrop spectrophotometer (Agilent Technologies, Santa Clara, CA, USA). Quality of the extracted RNA was assessed by the electrophoresis on a 1% agarose gel containing ethidium bromide. Total RNA was reverse transcribed into cDNA using SuperScript III First-strand Synthesis System (Invitrogen) using ran-
dom hexamer primers according to the manufacturer’s instructions.

Validation test for two hub genes and four common genes using quantitative Real-Time PCR. The expression of six host genes, including CDK1, PLK1, SMS, NNT, DEPDC1 and UHMK1 were measured by quantitative Real-Time PCR using the SYBR Green kit (TAKARA Bio INC., Otsu, Japan) on an Applied Biosystems®StepOnePlus™ Real-Time PCR system (Applied Biosystems, Foster city, CA, USA). The GAPDH gene was applied to normalize the relative quantity. The sequences of primers that used in this study are shown in Table 2. The Real-Time PCR data were calculated using the 2-ΔΔCt method (8). The relative gene expression of each marker were expressed as fold change to compare the mRNA levels between three different groups.

Expression measurement of viral E6 and E7 genes using quantitative Real-Time PCR. The expression levels of two viral genes, including E6 and E7, were measured by quantitative Real-Time PCR using the SYBR Green kit on a Real-Time PCR system. GAPDH was selected as the endogenous standard to normalize the relative quantity. The relative fold change was analysed using the 2-ΔΔCt method.

Statistical analysis. The Mann-Witney non-parametric test and t-test were applied to analyze the statistical discrepancy between studied groups using GraphPad Prism software (7.0.1). Statistical analysis was carried out with ANOVA for multiple comparisons. A P-value less than 0.05 was considered significant.

RESULTS

Patient characteristics and HPV genotyping. A total of 45 fresh uterine cervix biopsies, including 15 normal, 15 precancerous, and 15 cervical cancer specimens from Imam Khomeini Hospital of Tehran, Iran were included in this study (Table 3). The mean age of pre-malignant and malignant CC was 48.41 ± 7.33 and 54.44 ± 9.32 years, respectively. The mean age was 38.88 ± 7.11 years for control group. The difference of mean age between groups was statistically significant (P-value < 0.05).

Table 3. The prevalence of HPV and HPV types in studied groups

<table>
<thead>
<tr>
<th>Studied groups (N)</th>
<th>HPV positive N (%)</th>
<th>HPV genotypes N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal subject (n= 15)</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Pre-malignant patients (n= 15)</td>
<td>10 (67)</td>
<td>HPV 16 = 9 (90)</td>
</tr>
<tr>
<td>Malignant patients (n= 15)</td>
<td>15 (100)</td>
<td>HPV 16 = 15 (100)</td>
</tr>
</tbody>
</table>

As shown in Table 3, HPV DNA was detected in 67% and 100% of pre-malignant CC and CC cases, respectively. None of normal samples were infected with HPV. Direct sequencing data from 25 L1 sequences revealed 24 samples as being genotype 16. HPV-53 was observed in pre-malignant patients only.

Validation of hub genes (CDK1 and PLK1) and common genes of two networks (SMS, NNT, UHMK1, and DEPDC1) by quantitative RT-PCR. To confirm the reliability and validity of the microarray gene expression data, we selected two differentially expressed hub genes (CDK1 and PLK1) and four differentially expressed mRNAs that were common
in pre-malignant-normal and malignant-pre-malignant networks (SMS, NNT, UHMK1, and DEPDC1: Fig. 1) (7). Our previous analysis was indicated that the mRNA level of SMS, NNT and UHMK1 were significantly downregulated in pre-malignant patients compared to histologically normal subjects. However, gene expression analysis showed that the SMS, NNT and UHMK1 were significantly upregulated in malignant patients compared to pre-malignant patients. The mRNA level of DEPDC1 was upregulated in both networks. In the current study, 45 samples (15 per group) were subjected to qRT-PCR. We next investigated these transcriptional signatures in malignant cervical cancer when compared to healthy subjects and pre-malignant patients.

The qRT-PCR data revealed that mRNA levels of SMS, NNT and UHMK1 genes were significantly upregulated in patients with malignant patients compared to pre-malignant patients and normal subjects, which was consistent with the results of the initial microarray analysis (Fig. 2). However, no significant differences were found between pre-malignant patients and normal subjects, except for UHMK1 that the mRNA level of UHMK1 was significantly elevated in both networks. Nevertheless, our initial screening had shown that expression level of UHMK1 was downregulated in pre-malignant patients in comparison to histologically normal subjects. Opposite of our previous microarray result, no significant mRNA level of DEPDC1 was found among three different groups.

As a mentioned earlier, our previous investigation indicated the CDK1 and PLK1 were determined as the most important upregulated key genes of both malignant-pre-malignant and pre-malignant-normal networks. Analysis of mRNA levels of our study groups revealed statistically significant differences in transcriptional levels of CDK1 and PLK1 between malignant cervical cancer with pre-cancer patients and normal subjects (Fig. 3).

**E6 and E7 expression profiles in the tissue samples and correlation analysis between validated genes.** E6 and E7 viral gene expression profiles were analyzed in all HPV positive samples by qReal-Time PCR. The HPV 16 E6 mRNA was detected in all patients with malignant cervical cancer and in 88.9% (8 out of 9) of pre-malignant cancer, respectively. The E7 oncogene was highly expressed in all malignant cancer (100%) and in 77.8% (7 out of 9) of pre-malignant cases (Fig. 4). The relative expression values, adjusted to GAPDH expression, indicated the E6 and E7 mRNA levels were significantly higher in malignant cervical cancer than pre-malignant patients (P-value< 0.001).

To comprehend the associations among the E6 and E7 gene expressions, the correlation values were analysed. No remarkable associations were observed between viral oncogenes and each CDK1, PLK1, SMS, NNT and DEPDC1 genes in the tissue samples.
of the studied groups (Table 3). However, a positive correlation was observed in malignant CC between E6 and UHMK1 expression level \( R = 0.6, P\text{-value} = 0.0017 \).

**DISCUSSION**

Transcriptome analysis of the tissue samples from patients with cervical cancer is important to determine the molecular mechanisms underlying pathogenesis of cervical cancer and to identify the cancer specific hub genes. Microarray technique is a powerful tool for gene expression analysis and for identifying several clinically relevant molecular gene signatures for disease diagnosis and therapeutic approaches \( (9) \). Gene Expression Omnibus is a major source of microarray data and huge public repositories of transcriptome profiles \( (10) \). In our previous investigation, we had provided a network-based analysis of HPV 16 infection in histologically normal women and malignant CC/pre-malignant CC subjects \( (7) \). The CDK1 and PLK1 were determined as the most important hub genes with the highest centrality score of pre-malignant CC vs normal and malignant CC vs pre-malignant CC networks. Both hub genes of topological networks were involved in the progression of the cell cycle process.

Consistent with the microarray data analysis, the results of the present study indicated higher significant mRNA expressions of CDK1 and PLK1 in the tissue samples of 15 patients with malignant CC than those in the pre-malignant CC \( (n=15) \) and normal samples \( (n=15) \). Cellular cyclin dependent kinases include protein kinases are the positive key players that regulate activities throughout the cell cycle process \( (11) \). The CDK1 is the critical protein for G1 progression and G1 to S transition via association with several interphase cyclins \( (12-14) \). CDK1 protein is highly expressed in E7-expressing cells upon DNA damage \( (15) \). Systematic gene expression analysis by Luo et al. showed that CDK1 plays a comprehensive role in regulating genetic networks involved in the progression of CC \( (16) \). Previously, another independent microarray gene expression dataset \( (GSE9750) \) was used to further verification of our results obtained from the original dataset. Interestingly, CDK1 identified as a main hub gene in both original and validation networks \( (7, 17) \). In parallel, PLK1 is evidently a critical hub gene that play an important role in the CC progression. PLK1 is a key initiator for G2 to M transition \( (18) \) and is a candidate molecule as a biomarker \( (19) \). Regarding this fact, the inhibition of CDK1 and PLK1 genes also has been suggested as a highly specific potential therapeutic targets for cancer treatment \( (18) \).

In the present study, we find four DEGs including SMS, NNT, UHMK1 and DEPDC1 that were common in both pre-malignant CC vs normal and malignant CC vs pre-malignant CC patients and most of the DEGs of both networks were mainly specific for each group. These genes were either associated with important biological activities or oncogenic functions. Analysis of microarray data from the GSE67522 dataset showed higher expression of SMS and NNT in the malignant CC group than in pre-CC group and lower expression in pre-CC group than in normal group. Our quantitative Real-Time PCR results indicated that mRNA levels of SMS and NNT were more abundant in malignant CC compared to pre-malignant CC and normal groups. However, the mRNA levels were not shown significantly differences in malignant CC and pre-malignant CC. SMS have been shown to play an essential role in the cell growth and survival of several cancers, such as cervical cancer \( (20) \). Regard to this point, the increased mRNA level of SMS could be applied as a useful screening marker for the progression of human cancers \( (21) \). NNT located in the mitochondrial inner membrane and is critical to homeostasis of NADH and NADPH \( (22) \). It is suggested that dysregulation of NNT may
play a cell proliferation activity in cancers. Duan et al. revealed that NNT acts as a novel prognostic biomarker in hepatocellular carcinoma (HCC) (23). We found no statistically significant associations of DEPDC1 expression in patients groups. DEPDC1 is considered as a new cancer-associated gene which is aberrantly expressed in different cancers (24). Our microarray data obtained from US patients (17), whereas we evaluated the candidate genes in Iranian patients with cervical cancer. Ethnicity stems not just from differences in genetic sequence, but also from differences in the expression of genes shared by ethnic group (25). However, using qPCR analysis, we revealed the expression of UHMK1 was elevated in malignant CC compared to pre-malignant CC and histologically normal subjects. The mRNA level of UHMK1 was also upregulated in pre-malignant CC patients in comparison to normal subject. Moreover, we found statistically significant correlation between expression of E6 viral oncogene and UHMK1 gene. UHMK1 is important for phosphorylation of splicing factor 1 (SF1) and may be involved in the RNA splicing (26). Another critical function defined for UHMK1 is its ability to positively modulate the cell cycle progression in G1 phase (27).

Patients with transcriptionally active HPV are considered at higher risk for cancer progression. Several studies revealed that viral gene expression pattern could be useful as HPV molecular biomarker of cancer development (28-30). Analysis of E6 and E7 expression levels was considered useful to determine the risk of cervical cancer development (30, 31). In the present study, the E6 and E7 mRNA levels in malignant CC was much higher than in pre-malignant CC. These results were consistent with previous studies reporting that HPV is always biologically active in cervical cancer patients, suggesting that the E6 and E7 genes are the main tumourogenic factors (30). However, except of UHMK1, we observed that there were no significant correlation between expression of CDK1, PLK1, SMS, NNT, UHMK1, and DEPDC1 genes with E6/E7 viral mRNA levels in both malignant CC and pre-malignant CC.

CONCLUSION

Using quantitative Real-Time PCR analysis, we suggested that the measurement expressions of CDK1, PLK1, SMS, NNT and UHMK1 genes in CC patients are a simple and useful tool for diagnosis of disease. CDK1 and PLK1 also can be potentially used as diagnostic and prognostic biomarkers as well as can be applied to therapeutic targets for CC treatment.

ACKNOWLEDGEMENTS

This study has been funded and supported by Tehran University of Medical Sciences (TUMS), Grant no. 39293. It has also been part of a Ph.D. thesis supported by TUMS.

REFERENCES