Screening of biomarkers in cervical squamous cell carcinomas via gene expression profiling

BING CHEN, CHUNDONG LI, LEI ZHANG, JIAHUI LV and YING TONG

Department of Gynaecology and Obstetrics, General Hospital of The Air Force, Beijing 100142, P.R. China

Received December 9, 2014; Accepted August 27, 2015

DOI: 10.3892/mmr.2015.4322

Abstract. In the present study, gene expression profiles of high-grade squamous intraepithelial lesions (HSIL) and invasive cervical squamous cell carcinomas (CSCC) were analyzed using bioinformatic tools to identify key genes and potential biomarkers. Analyses of differentially expressed genes (DEGs) were performed for HSIL vs. normal control and invasive CSCC vs. normal control tissues using the Limma package in R. Pathway enrichment analysis was performed using KOBAS. A protein-protein interaction (PPI) network for the DEGs in invasive CSCC was constructed using String. Functional enrichment analysis was performed for the DEGs in the PPI network using DAVID. Relevant small molecules were predicted using Cmap. A total of 633 and 881 DEGs were identified in HSIL and invasive CSCC, respectively, and the two groups had 305 DEGs in common. Genes associated with the mitogen-activated protein kinase signaling pathway were enriched in the HSIL, while cell cycle-associated genes were over-represented in invasive CSCC. The PPI network, containing 72 upregulated genes and 434 edges, was illustrated. Functional enrichment analysis showed that the cell cycle was the most significant gene ontology term. A total of six small molecules associated with the pathology of CSCC were identified, including the anti-cancer drug piperlongumine, which showed a negative correlation. The findings of the present study not only enhanced the current understanding of the pathogenesis of CSCC, but may also be a basis for the development of novel therapies.

Introduction

Cervical cancer is the second most common cancer in women worldwide with the fifth highest mortality rate (1,2). Squamous cell carcinomas are the most common type, accounting for 80-85% of all cervical cancers (3). Infection with human papilloma virus is the greatest risk factor for cervical cancer (4), followed by smoking (5). The five-year relative survival rate for the earliest stage of invasive cervical cancer is 92%; however, the prognosis is significantly lower when metastasis is present, suggesting the importance of early diagnosis.

Studies have identified pathways associated with the pathogenesis of cervical cancer, particularly the molecular mechanisms underlying its invasiveness. Wnt signaling was reported to be involved in the pathogenesis of cervical cancer (6). Tumor necrosis factor-α (TNF-α) is a pro-inflammatory cytokine, which has been implicated in several cancers. Duarte et al (7) reported that G-308A TNF-α polymorphism is associated with an increased risk of invasive cervical cancer. Chan et al (8) indicated that overexpression of forkhead box M1 transcription factor is associated with cervical cancer progression and pathogenesis. Murphy et al (9) performed an immunocytochemical analysis to reveal that p16INK4A, CDC6 and MCM5 are predictive biomarkers in cervical pre-invasive neoplasia and cervical cancer. Microarray technology is also widely adopted in the discovery of crucial genes. Song et al (10) identified several candidate genes associated with invasion of cervical cancer via microarray analysis of normal cervix, in situ carcinoma and invasive cervical cancer tissues. Zhai et al (11) identified genes contributing to the invasive properties of cervical carcinoma cells. However, as these findings have not resulted in an improved outcome for patients with cervical carcinoma, additional study is required.

The present study, analyzed gene expression profiles of high-grade squamous intraepithelial lesions (HSIL) and invasive cervical squamous cell carcinomas (CSCC) with currently available bioinformatic tools, attempting to identify crucial genes in the pathogenesis of CSCC as well as potential biomarkers for diagnosis or prognosis.

Materials and methods

Microarray data. A gene expression data set [accession no. GSE7803 (11)], including 10 normal squamous cervical epithelial, 7 HSIL and 21 invasive CSCC samples, was downloaded from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). Gene expression levels were determined using the Affymetrix Human Genome U133A Array (no. HG-U133A; Affymetrix Inc., Santa Clara, CA, USA), and probe annotations were acquired.
Screening of DEGs. The 22,283 probes were mapped into 20,967 genes. A log² transformation was applied on the gene expression levels (12). Analysis of differentially expressed genes (DEGs) was performed for pre-invasive cervical squamous cell carcinomas vs. normal control and invasive cervical squamous cell carcinomas vs. normal control groups using the Limma package (13) in R. Multiple testing correction according to the Benjamini-Hochberg (BH) method (14) was applied to the P-values and the false discovery rate (FDR) was calculated. FDR<0.05 was set as the cut-off value to screen out significant DEGs.

To identify genes associated with the invasiveness of CSCC, DEGs in HSIL were compared with those in invasive CSCC.

Cluster analysis. Two-way cluster analysis was performed using the expression levels of the DEGs with package pheatmap in R (15). An Euclidean distance was adopted in the analysis.

Pathway enrichment analysis. Pathway enrichment analysis was performed for the DEGs using KOBAS (16). The statistical method is based on cumulative hypergeometric distribution and P<0.05 was set as the threshold to filter out significantly over-represented biological pathways.

Construction of a protein-protein interaction (PPI) network. Proteins are involved in complex interaction networks to fulfill certain biological functions. Therefore, revealing the PPI is a useful method to identify molecular mechanisms. A PPI network was constructed for the DEGs of invasive CSCC using String (17), which was then visualized by Cytoscape (18).

Functional enrichment analysis. Functional enrichment analysis was performed for the DEGs in the PPI network using the Database for Annotation, Visualization and Integration Discovery (DAVID; http://david.abcc.ncifcrf.gov/) (19) online tool. The statistical method is based on hypergeometric distribution. FDR<0.05 was set as the cut-off value.

Prediction of relevant small molecules. Connectivity map (Cmap) was designed to link gene patterns associated with disease to corresponding patterns produced by drug candidates (20,21). Relevant small molecules were predicted using the DEGs and those with |score| >0.9 were retained.

Results

Differentially expressed genes. Compared with normal controls, 633 and 881 DEGs were identified in HSIL and invasive CSCC, respectively. The two groups of DEGs were compared and 305 genes were found to be common between HSIL and invasive CSCC (Fig. 1).

Cluster analysis. To verify the reliability of the DEG results, two-way cluster analysis was performed with unique DEGs of HSIL and invasive CSCC (Fig. 2). HSIL as well as invasive CSCC were clearly separated from normal controls, confirming the reliability of the DEG analysis.

Pathway enrichment analysis. Pathway enrichment analysis was performed for the unique DEGs and common DEGs of HSIL and invasive CSCC using KOBAS (Table I). the mitogen-activated protein kinase (MAPK) signaling pathway was significantly enriched in the unique DEGs of HSIL, while the cell cycle was overrepresented in the unique DEGs of invasive CSCC.

PPI network of DEGs in invasive CSCC. A PPI network was constructed for the DEGs in invasive CSCC (Fig. 3). The network consisted of 72 upregulated genes and 434 edges.
Functional enrichment analysis. Functional enrichment analysis was performed for the genes in the PPI network using the DAVID online tool. The top 10 gene ontology (GO) terms are listed in Table II. All of these terms were associated with the cell cycle, which was in accordance with the results of the pathway enrichment analysis. A total of 41 DEGs were involved in the cell cycle, including DBF4, TTK, PTTG1, CDC45, CDK1, CDC20, MCM2, MCM6, CCNB1 and MAD2L1.

Relevant small molecules. A total of six relevant small molecules were predicted by Cmap with |score| >0.9 (Table III). Piperlongumine was the most negatively correlated molecule. Previous studies have indicated that piperlongumine has anti-tumor activity (22,23).

Discussion

In the present study, a comparative analysis of gene expression profiles was performed between HSIL, invasive CSCC and normal controls. A total of 633 and 881 DEGs were identified in HSIL and invasive CSCC, respectively. Comparison of the two groups of DEGs showed that the HSIL and CSCC groups had 305 DEGs in common. Cluster analysis results verified the confidence of the DEGs. Pathway enrichment analysis revealed that the MAPK signaling pathway was significantly enriched in the unique DEGs of HSIL, while the cell cycle was overrepresented among the unique DEGs of invasive CSCC. The MAPK pathway can be activated by diverse extracellular and intracellular stimuli and regulates a variety of cellular activities, including proliferation, differentiation, survival and death. Deregulation of MAPK pathways has been implicated in numerous human diseases, including cancer (24,25). Dysregulation of the cell cycle is the most common feature of cancer (20,26) and the analysis of the present study revealed that it was most significantly enriched in invasive CSCC.

To further investigate the molecular mechanisms underlying invasive CSCC, the PPI network was constructed, which included 72 upregulated genes and 434 edges. Functional enrichment analysis revealed that the cell cycle and GO terms
associated with the cell cycle were enriched in the genes from
the network. This finding was consistent with the results of
the pathway enrichment analysis. Several of the genes identi-
fied were key genes or potential targets in invasive CSCC.
NEK2 is a serine/threonine-protein kinase that is involved in
mitotic regulation. Upregulation of NEK2 is observed in cell
lines derived from breast cancer (27) and cervical cancer (28).
Hayward and Fry (28) suggested that NEK2 contributes
to chromosome instability and may be a target for chemo-
therapeutic intervention. DBF4 is involved in cell adhesion and
migration, possibly through its regulation of the arrangement of
the actin cytoskeleton (29). The overexpression of DBF4 has
been reported in numerous cancer types (30), and the present
study revealed that it was upregulated in invasive CSCC and
may contribute to the metastasis of CSCC. PTTG1 has trans-
foming activity in vitro and tumorigenic activity in vivo, and is
highly expressed in various tumor types. Depletion of PTTG1
has anti-proliferative effects in multiple tumor types (31). It also
increases cell motility and promotes lymph node metastasis in
esophageal squamous cell carcinoma (32). Hence, the present
study speculated that PTTG1 may have a crucial role in the
proliferation and mobility of cervical cancer cells. Elevated
expression levels of MCM2 and MCM6 has been reported in
cervical neoplasia (33), suggesting that these genes may be
implicated in the development of CSCC.

Furthermore, the present study predicted associated small
molecules using the expression levels of DEGs in invasive CSCC
by using Cmap. Piperlongumine was the most negatively corre-
lated molecule, which is a bioactive compound isolated from
long peppers that shows selective toxicity towards a variety of
cancer cell types (34). The cytotoxicity of piperlongumine has
been attributed to increases in reactive oxygen species in cancer
cells. Jarvius et al (35) reported that it induces inhibition of the
ubiquitin-proteasome system in cancer cells. Ginzburg
et al (36) further reported that piperlongumine inhibits nuclear factor-κB
activity and attenuates aggressive growth characteristics of
prostate cancer cells. Piperlongumine may therefore be suitable

Table II. Significantly enriched GO terms in the genes from the network.

<table>
<thead>
<tr>
<th>GO term</th>
<th>Function</th>
<th>Count</th>
<th>P-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0007049</td>
<td>Cell cycle</td>
<td>41</td>
<td>1.35x10^{-33}</td>
<td>1.96x10^{-30}</td>
</tr>
<tr>
<td>0000279</td>
<td>M phase</td>
<td>31</td>
<td>1.37x10^{-33}</td>
<td>1.99x10^{-28}</td>
</tr>
<tr>
<td>0000278</td>
<td>Mitotic cell cycle</td>
<td>32</td>
<td>1.43x10^{-33}</td>
<td>2.07x10^{-28}</td>
</tr>
<tr>
<td>0022403</td>
<td>Cell cycle phase</td>
<td>33</td>
<td>1.49x10^{-33}</td>
<td>2.16x10^{-28}</td>
</tr>
<tr>
<td>0007067</td>
<td>Mitosis</td>
<td>27</td>
<td>3.88x10^{-30}</td>
<td>5.63x10^{-27}</td>
</tr>
<tr>
<td>0000280</td>
<td>Nuclear division</td>
<td>27</td>
<td>3.88x10^{-30}</td>
<td>5.63x10^{-27}</td>
</tr>
<tr>
<td>0022402</td>
<td>Cell cycle process</td>
<td>35</td>
<td>4.74x10^{-30}</td>
<td>6.88x10^{-27}</td>
</tr>
<tr>
<td>0000087</td>
<td>M phase of mitotic cell cycle</td>
<td>27</td>
<td>6.31x10^{-30}</td>
<td>9.15x10^{-27}</td>
</tr>
<tr>
<td>0048285</td>
<td>Organelle fission</td>
<td>27</td>
<td>1.14x10^{-29}</td>
<td>1.66x10^{-26}</td>
</tr>
<tr>
<td>0051301</td>
<td>Cell division</td>
<td>27</td>
<td>9.56x10^{-27}</td>
<td>1.39x10^{-23}</td>
</tr>
</tbody>
</table>

FDR, false discovery rate; GO, gene ontology.

Table III. Small molecules associated with the pathology of cervical squamous cell carcinomas.

<table>
<thead>
<tr>
<th>Cmap name</th>
<th>Enrichment</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piperlongumine</td>
<td>-0.927</td>
<td>0.01099</td>
</tr>
<tr>
<td>GW-8510</td>
<td>-0.915</td>
<td>0.00010</td>
</tr>
<tr>
<td>Alsterpaullone</td>
<td>-0.911</td>
<td>0.00128</td>
</tr>
<tr>
<td>Quinostatin</td>
<td>-0.901</td>
<td>0.01976</td>
</tr>
<tr>
<td>Prestwick-692</td>
<td>0.943</td>
<td>&lt;0.00010</td>
</tr>
<tr>
<td>Isoflupredone</td>
<td>0.959</td>
<td>0.00010</td>
</tr>
</tbody>
</table>
for controlling invasive CSCC. This result may be useful for the development of drugs for invasive CSCC.

In conclusion, the present study identified a number of DEGs in HSIL and invasive CSCC, which may provide direction for future studies. Potential biomarkers and associated small molecules for CSCC were revealed, which may contribute to the development of novel diagnostic markers and therapeutics for CSCC.

References


