Characterization of the microRNA profile in early-stage cervical squamous cell carcinoma by next-generation sequencing

WANSONG LIN, MEI FENG, GUILIN CHEN, ZHIFENG ZHOU, JIEYU LI and YUNBIN YE

1Laboratory of Immuno-Oncology, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital; 2Fujian Key Laboratory of Translational Cancer Medicine; 3Department of Gynecologic Oncology, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital, Fuzhou, Fujian 350014, P.R. China

Received July 6, 2016; Accepted September 1, 2016

DOI: 10.3892/or.2017.5372

Abstract. Squamous cell carcinoma (SCC) is histologically the most prominent type of cervical cancer. There is accumulating evidence suggesting that microRNAs (miRNAs) play important regulatory roles in the biological processes of cervical squamous cell carcinoma (CSCC). Deciphering the miRNA regulatory network in CSCC could deepen our understanding at the molecular level of CSCC initiation and progression. In the present study, we performed next-generation sequencing (NGS) to profile miRNA expression in 3 pairs of early-stage CSCC samples. Quantitative real-time polymerase chain reaction (qRT-PCR) was used to verify primary findings in another 20 pairs of CSCC samples. We identified 37 known miRNAs that exhibited significant alterations in expression (2-fold change or greater), among which 8 miRNAs were upregulated and 29 miRNAs were downregulated. Nine of these miRNAs were selected for further qRT-PCR validation. A novel miRNA candidate was also reported for the first time in the present study to be upregulated. The Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed that its target genes were involved in MAPK, calcium and adherent junction signaling pathways. The present study systematically characterized the miRNA expression variation in early-stage CSCC and provides novel biomarkers for diagnosis and treatment as well as an opportunity for further investigation of the molecular mechanisms underlying the pathogenesis and development of CSCC.

Introduction

In less developed countries, cervical cancer is the second most common cancer and the third leading cause of cancer-related deaths in females. There were an estimated 444,500 new cervical cancer cases and 230,200 deaths in these countries in 2012 (1). Almost 90% of the deaths occurred in developing countries and for these countries it is a huge economic health care burden. The highest cervical cancer death rates are in Asia, with an estimated 144,400 deaths occurring in 2012 (1). Squamous cell carcinoma (SCC) is the most common histological type of cervical cancer (~80% of cases) and is closely correlated with human papillomavirus (HPV) infection (2). Cervical screening can prevent SCC, as it allows for detection and treatment of precancerous lesions, and application of the HPV vaccine also reduces the risk of SCC (3). However, the high cost is a major barrier to making screening and vaccination widely available to all women in developing countries, including China.

Cervical squamous cell carcinoma (CSCC) cases that are classified as International Federation of Gynecology and Obstetrics (FIGO) stages IB and IIA have a greater possibility of being cured and patients have better prognosis (4). Since early detection results in a better prognosis and cervical screening and vaccines are extremely expensive, there is an urgent need to identify more efficient early diagnostic biomarkers and therapeutic targets for CSCC. Therefore, it is necessary to augment our understanding of the molecular mechanisms underlying the pathogenesis and development of CSCC.

Although previous studies have identified chromosomal variations, abnormal expression of oncogenes or tumour-suppressor genes and aberrant promoter methylation in CSCC, few are relevant to microRNAs (miRNAs). miRNAs are a class of small non-coding RNAs (ncRNAs) ~19-25 nucleotides (nts) in length that can regulate gene expression by known classic mechanisms such as binding to messenger RNA (mRNA) at the 3′-untranslated region (3′-UTR) to downregulate protein-coding genes by increasing mRNA degradation (5). They also act through other newly found mechanisms such as binding to promoters, proteins, and by directly interacting with other ncRNAs (5,6). miRNAs play important regulatory roles in various biological processes e.g. differentiation, development, proliferation, apoptosis, cell cycle control and metabolism (7,8). Changes in miRNA expression have been identified in many diseases, such as cardiac and autoimmune disorders, schizophrenia and cancer (9). miRNAs have been found to be differentially expressed between almost all types of
analyzed benign and malignant tumors and non-tumor tissues, including CSCC (10). In addition, miRNAs participate in regulating a variety of neoplastic biological capabilities acquired during the multistep process of tumor development. Using miRNA microarray, Wilting et al. observed that 46 miRNAs showed significantly differential expression between normal cervical squamous epithelium and CSCC (11). Both miR-19a and miR-19b were reported to promote cervical cancer cell proliferation (12), whereas miR-125b inhibited cervical cancer cell apoptosis (13); both processes are involved in cervical carcinogenesis. Other studies reported that miR-9 (14) and miR-135a (15) contributed to HPV-induced cervical cancer formation, and that miR-133b (16) and miR-10a (17) promoted the progression and development of cervical cancer by regulating tumor cell migration, invasion and metastasis. However, miR-424 (18), miR-124 (19) and miR-218 (20) have opposite roles. All of these cited studies analyzed the differences in expression of cervical cancer-associated miRNAs by reverse transcription-polymerase chain reaction (RT-PCR) and hybridization-based microarray that only measure known and relatively abundant miRNAs, but are unable to identify novel miRNAs. To date, no studies have focused on the miRNA profile in early-stage CSCC. Thus, investigating the miRNA profile in early-stage CSCC by high-throughput sequencing may facilitate the analysis of expression of all annotated miRNAs, detection of novel miRNAs and identification of candidate markers for early diagnosis, and treatment of CSCC.

In the present study, we applied next-generation sequencing (NGS) to globally and systematically characterize miRNA alterations in early-stage CSCC by a comparison with paired normal samples. Furthermore, we validated these miRNAs by qRT-PCR and analyzed miRNA target genes and performed function annotation and pathway analysis.

Materials and methods

Clinical sample collection. For sequencing, 3 pairs of human CSCC tumor (T) and paired normal tissue (N) samples in FIGO stages I-II were obtained from patients undergoing surgery of the cervix. Additional, 20 pairs of CSCC and normal tissues were also collected for quantitative real-time polymerase chain reaction (qRT-PCR) validation. All tumor and paired normal tissues were examined by hematoxylin and eosin (H&E) staining, and all pathologic diagnosis were confirmed by 2 independent pathologists; tumor tissues containing >90% tumor cells were selected for further study (Fig. 1). The study protocol was approved by the Ethics Committees on Human Research of the Fujian Cancer Hospital. All patients agreed to join the study and signed a written informed consent. None of the patients had received preoperative adjuvant radiotherapy or chemotherapy. All samples were immediately snap-frozen in liquid nitrogen and then stored at -80°C until RNA isolation.

RNA isolation. Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's recommendations, and quantified using a ND-1000 spectrophotometer (NanoDrop, Rockland, DE, USA). The RNA quality was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and the 28S/18S ratio was determined on 1% agarose gel. Only RNA extracts with an RNA integrity number (RIN) value ≥7 and 28S/18S ratio >1.8 were used for further analysis.

Deep sequencing and bioinformatic analysis. RNA sequencing and bioinformatic analysis were performed at the Beijing Genomics Institute (BGI; Shenzhen, China). In brief, total RNA was ligated to the 5'- and 3'-end adaptors, and sequentially amplified by an RT-PCR reaction. The fragments of ~62-75 bp (small RNA plus adaptors) were isolated from the agarose gel and then purified. These products were directly sequenced with Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. After trimming adaptor sequences, removing contaminated reads and filtering low quality reads, the remaining clean reads were mapped to the human genome using short oligonucleotide alignment program (SOAP). The other small RNAs (rRNA, tRNA, snRNA, snoRNA and piRNA) were detected by screening against Rfam 10.1 and the GenBank database. The known miRNAs were identified by aligning them to the miRBase. The expression differences of the known miRNAs between the CSCCs and paired normal samples were evaluated by comparing the log2 ratio of the 2 groups. The miRNAs with at least a 2-fold-change of expression and a P-value <0.05 between the 2 groups in all three patients were selected for further study. The novel miRNAs were predicted using MIREAP software.

Validation of mature miRNA expression by qRT-PCR. For validation of the selected miRNAs that showed differential expression and predicted novel miRNAs, qRT-PCR was applied using the Applied Biosystems 7500 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. In brief, the miRNAs were reverse transcribed into cDNA using the miScript II RT kit (Qiagen, Valencia, CA, USA) with 5X miScript HiSpec buffer. The mixture was incubated at 37°C for 60 min, and then at 95°C for 5 min and then placed on ice. Quantitative PCR was performed using the miScript SYBR-Br Green PCR kit and 10X miScript Primer Assay (both from Qiagen). The reaction consisted of 1 cycle at 95°C for 15 min, followed by 40 cycles at 94°C for 15 sec and 55°C for 30 sec and 70°C for 30 sec. All reactions were repeated 3 times. The relative expression levels of miRNAs were calculated using the 2^-ΔΔCt method with snRNA U6 as the internal reference. Quantitative PCR data were analyzed by SPSS statistics software (version 19.0; IBM, Armonk, NY, USA). A Student's t-test was used to compare the difference in expression between CSCC and normal tissues. P<0.05 was considered to indicate a statistically significant result.

miRNA target prediction, functional annotation and pathway analysis. The target genes of the selected miRNAs with differential expression and novel miRNAs were predicted using RNAhybrid. For determination of the main biological functions of the candidate target genes, Gene Ontology (GO) enrichment analysis was applied. We mapped all candidates to GO terms in the database (http://www.geneontology.org/), using a hypergeometric test to find significantly enriched GO terms. The GO terms with corrected P-values with an...
enrichment ratio ≤0.05 were defined as significantly enriched target gene candidates. For identification of significantly enriched metabolic pathways or signal transduction pathways in target gene candidates, we mapped these candidates to the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (21). The calculation method is the same as that in the GO analysis. Pathways with corrected P-values with an enrichment ratio ≤0.05 were considered as significantly enriched target gene candidates.

Statistical analysis. All quantitative data are expressed as the mean ± standard deviation (mean ± SD) from at least 3 samples or experiments/data point. Significant differences were analyzed by the Student’s t-test to compare 2 groups of independent samples.

Results

Overview of small RNA sequencing results. To determine the miRNA profile of early-stage CSCC, we performed high-throughput NGS in 3 CSCC and their paired normal samples using Illumina HiSeq 2500. After removing adapters, low quality tags and contaminants, an average of 11,478,411 clean reads (ranging from 10,313,021 to 11,820,021) remained (Table I). The majority of these clean reads were 18-28 nt in length, with 22 nt RNAs the most abundant. These clean reads were mapped to the Human Genome and rRNA, tRNA and snRNA were identified by the alignment to Rfam 10.1 and GenBank databases. Subsequently, an average of 1,726 mature miRNAs (ranging from 1,571 to 1,881) were successfully annotated in the miRBase, and thus identified

Table I. Summary of the small RNA sequencing.

<table>
<thead>
<tr>
<th>Category</th>
<th>T-1</th>
<th>T-2</th>
<th>T-3</th>
<th>N-1</th>
<th>N-2</th>
<th>N-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total reads</td>
<td>12,000,000</td>
<td>10,523,729</td>
<td>12,000,000</td>
<td>12,000,000</td>
<td>12,000,000</td>
<td>12,000,000</td>
</tr>
<tr>
<td>Clean reads</td>
<td>11,713,704</td>
<td>10,313,021</td>
<td>11,519,252</td>
<td>11,799,688</td>
<td>11,820,021</td>
<td>11,704,777</td>
</tr>
<tr>
<td>%</td>
<td>98.40</td>
<td>98.73</td>
<td>96.70</td>
<td>99.02</td>
<td>99.18</td>
<td>98.25</td>
</tr>
<tr>
<td>Unique sRNAs</td>
<td>884,315</td>
<td>673,581</td>
<td>867,499</td>
<td>615,904</td>
<td>466,839</td>
<td>658,536</td>
</tr>
<tr>
<td>miRNAs</td>
<td>4,196</td>
<td>4,545</td>
<td>4,346</td>
<td>3,634</td>
<td>4,035</td>
<td>4,173</td>
</tr>
<tr>
<td>Known miRNAs in miRBase</td>
<td>1,708</td>
<td>1,881</td>
<td>1,810</td>
<td>1,571</td>
<td>1,659</td>
<td>1,724</td>
</tr>
<tr>
<td>miRNAs</td>
<td>200</td>
<td>225</td>
<td>229</td>
<td>171</td>
<td>182</td>
<td>200</td>
</tr>
<tr>
<td>miRNA-5p</td>
<td>414</td>
<td>431</td>
<td>418</td>
<td>385</td>
<td>394</td>
<td>400</td>
</tr>
<tr>
<td>miRNA-3p</td>
<td>306</td>
<td>351</td>
<td>325</td>
<td>298</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>miRNA precursors</td>
<td>788</td>
<td>874</td>
<td>838</td>
<td>717</td>
<td>753</td>
<td>789</td>
</tr>
<tr>
<td>Unannotated sRNAs</td>
<td>232,043</td>
<td>205,460</td>
<td>255,885</td>
<td>291,651</td>
<td>156,254</td>
<td>240,371</td>
</tr>
<tr>
<td>Novel miRNA candidates</td>
<td>28</td>
<td>24</td>
<td>31</td>
<td>20</td>
<td>22</td>
<td>26</td>
</tr>
</tbody>
</table>

T, CSCC tumor tissue; N, paired normal tissue.
as known miRNAs (Table I). The sequences which did not match with any database were considered as unannotated sRNAs (Table I).

**Identification and validation of differentially expressed known miRNAs between CSCC and normal tissues.** To detect the known miRNAs which were differentially expressed between the CSCC and normal tissues, the expression level of each miRNA was first normalized to provide the expression of transcripts per million (TPM), and then the fold-change and P-values were calculated from the normalized expression. There were 424 differentially expressed miRNAs between T1 and N1, 480 between T2 and N2 and 440 between T3 and N3 (Fig. 2). Subsequently, on the basis of the cut-off criteria (fold-change >2 and p-value <0.05), 186, 149 and 126 miRNAs were selected from the data from each of the 3 pairs of patient samples. Finally, only 37 miRNAs exhibited consistent changes in all 3 pairs of samples; among these 37 differentially expressed miRNAs, 8 were significantly upregulated and 29 were significantly downregulated (Fig. 3; Table II).

To further validate the deep sequencing results, qRT-PCR was applied to assess the significantly differentially expressed miRNAs in 3 pairs of samples used for sequencing and the additional 20 pairs of samples. Nine miRNAs were selected from the miRNAs with significantly modified expression and measured by qRT-PCR. All of the 9 miRNAs displayed the same expression pattern consistent with the sequencing data (Fig. 4). This indicated that our high through-put sequencing results were reliable.
Target genes of the significantly altered miRNAs. To demonstrate the differences in gene expression associated with the miRNA profile in CSCC, the putative target genes of the miRNAs that were expected to be significantly differently expressed between CSCC and normal tissues were predicted using RNAhybrid. GO and KEGG pathway analyses were performed to enrich the main biological processes and pathways in which the target gene candidates were involved. However, the P-value of the enrichment ratio failed to achieve filter conditions and the analysis was not continued.

Prediction of novel miRNAs, their target genes and pathways. To detect novel miRNAs, unannotated sRNAs were further processed using MIREAP. The number of novel miRNA candidates obtained from the 6 small RNA libraries ranged from 20 to 31 (Table I). There were 5 common novel miRNA
candidates in the 3 pairs of CSCC and normal samples. Among the 5 common miRNAs, only 1 miRNA, novel_miR_7, displayed a significant difference in expression (fold-change >1 and P-value <0.05) in all 3 pairs of samples: novel_miR_7 was selected as a putative novel miRNA and verified by qRT-PCR (Fig. 5). Its sequence and stem-loop secondary structure are shown in Fig. 5. To further understand the function of the novel miRNA, its target genes were predicted using RNAhybrid and also assessed by GO and KEGG pathway analyses. The GO analysis revealed the main biological processes in which the target genes are involved and is shown in Fig. 6. The KEGG pathway analysis indicated that 5 pathways were enriched; including the MAPK, calcium and adherent junction signaling pathways (Table III).
Discussion

Although CSCC caused by HPV infection can be prevented, it is still a common cancer in women. There is accumulating evidence to indicate that HPV infection alone is insufficient to cause the disease and miRNAs may play critical roles in carcinogenesis (22-24). Therefore, global analysis of miRNA expression comparing CSCC with normal tissue may aid in the better understanding of the function of miRNAs in the course of the disease as well as finding novel biomarkers for diagnosis, treatment and prognostic evaluation. To the best of our knowledge, this is the first study to profile miRNA expression in early-stage CSCC on the NGS platform. We identified 37 significantly differentially regulated known miRNAs in CSCC compared with normal tissue and confirmed a sample of 9 by qRT-PCR. Furthermore, a novel miRNA candidate, novel_miR_7, was identified for the first time and its target genes are involved in MAPK, calcium and adherent junction signaling pathways.

miRNAs are rather stable, even in body fluids such as serum, plasma, urine and saliva (25). The expression of miRNAs is unique to tissues or organs and they can be easily detected (26). These characteristics make them ideal cancer biomarkers. Juan \textit{et al} (27) performed deep sequencing on the serum pools of cervical cancer patients and healthy controls to identify putative novel miRNAs. They found that only 1 of 17 common novel miRNA candidates may distinguish cervical cancer patients from healthy controls. Wang \textit{et al} (28) also constructed expression profiles of miRNAs in the serum of CSCC patients by microarray. They discovered that the levels of 291 of 338 detectable circulating miRNAs were changed >2-fold in serum samples from CSCC patients vs. controls. However, they did not validate these results through any other methodologies. Although detection of miRNAs in serum has been proposed to be a simpler and less invasive diagnostic method than sampling tissue, the consequential biological and technical variability, which is mainly assessed by a series of different normalization processes based on the expression of reference genes, is difficult to avoid. Currently, there is no known suitable extracellular reference RNA for a proper normalization, particularly in cell-free miRNA analysis (26). Commonly used references, including U6 small nuclear RNA (RNu6B) and 5S ribosomal RNA, may be degraded or expressed less stably than others in serum (29). Therefore, we decided to analyze miRNA expression in CSCC tissues. In previous studies, a microarray technique, which is biased

Table III. KEGG pathway analysis for predicted target genes of the novel miRNA in CSCC.

<table>
<thead>
<tr>
<th>KEGG no.</th>
<th>KEGG pathway</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>KO:04010</td>
<td>MAPK signaling pathway</td>
<td>a</td>
</tr>
<tr>
<td>KO:04020</td>
<td>Calcium signaling pathway</td>
<td>b</td>
</tr>
<tr>
<td>KO:04520</td>
<td>Adherent junction</td>
<td>b</td>
</tr>
<tr>
<td>KO:05217</td>
<td>Basal cell carcinoma</td>
<td>b</td>
</tr>
<tr>
<td>KO:04730</td>
<td>Long-term depression</td>
<td>b</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Q-value <0.05 in all 3 pairs of samples; \textsuperscript{b}Q-value <0.01 in all 3 pairs of samples. Q-value, corrected P-value. KEGG, Kyoto Encyclopedia of Genes and Genomes.
towards abundantly expressed known miRNAs, was the major platform for miRNA expression profiling (11, 28). We chose a more sensitive and specific method, NGS, which enables high-throughput analysis and novel miRNA discovery, to measure miRNA expression.

In the present study, we identified 9 miRNAs that exhibited differential expression in CSCC using deep sequencing and confirmed by qRT-PCR evaluation. Six were downregulated, including miR-211-5p, miR-204-5p/3p, miR-218-1-3p/5p and miR-202-5p, and 3 were upregulated, including miR-34b-5p, miR-34c-5p and miR-21-3p. Among these miRNAs, miR-218, miR-21 and miR-34b-5p have previously been discovered and their functions had been verified to be associated with cervical cancer. In comparison to normal cervical tissues the expression of miR-218 was reported to be lower in cervical cancer. In comparison to normal cervical tissues the miR-21 and miR-34b-5p have previously been discovered.

miR-21 was found to be mainly expressed in the tumor histological diagnosis in cervical cancer (30). Meanwhile, miR-21 expression was found to be correlated with poorer prognosis: It suppresses colorectal cancer cell growth, migration, and invasion as well as promotes tumor sensitivity to chemotherapy by inhibiting RAB22A (41). In endometrial carcinoma, miR-204-5p exhibits a similar function through the TrkB-STAT3-miR-204-5p regulatory pathways (42) and miR-204-3p inhibits the growth of hepatocellular carcinoma tumor endothelial cells by blocking the adhesion function of FN1 (43). miR-202 can suppress osteosarcoma cell proliferation and induce cell apoptosis by downregulating Gli2 expression (44). While miR-34c is considered to be a putative tumor suppressor in a majority of tumors (45-48), we found it to be upregulated in CSCC, which is in agreement with previous results that determined expression by miRNA microarray in cervical cancer (11, 28). Thus, we hypothesized that miR-34c may play different roles in cervical cancer and other types of malignancy.

NGS technology allows the rapid discovery of novel miRNAs, for some of which the biological function is unknown. We identified 5 common novel miRNA candidates with differential expression between CSCC and normal tissues. Our focus was narrowed to miRNAs that exhibited a significant change of expression in all 3 pairs of samples. We identified novel_miR_7 as a candidate novel miRNA and confirmed its presence. Functional annotation revealed that its predicted target genes are in the MAPK signaling pathway, which is closely related to cell proliferation, cell death, and the balance between them. Early in 2006, Engelbrecht et al indicated the involvement of MAPK in cervical cancer and demonstrated the correlation between the activity of MAPK and apoptosis in the disease process (49). The HPV 16 E2 protein may induce apoptosis by inhibiting MAPK signaling in CSCC (50). However, playing a role in cervical cancer carcinogenesis, the expression level of MAPK may serve as a marker for cervical cancer progression (51, 52) and its inhibitors have shown the potential to be useful for cervical cancer therapy (53). Therefore, our results provide a solid basis for the characterization of the role of novel_miR_7 in the pathogenesis and development of CSCC and in the treatment of this disease.

In summary, for the first time we identified known and novel miRNAs in early-stage CSCC by high-throughput NGS. Moreover, we validated these miRNAs in other independent samples by qRT-PCR. Our analysis of differentially expressed miRNAs not only support existing findings but also uncovered various differentially expressed miRNAs which had not previously been reported in CSCC, thus providing new molecular signatures of the disease. The prediction of novel miRNAs found that novel_miR_7 may be a promising biomarker for CSCC. These results have implications for exploring the regulatory network of miRNAs, provide new targets for molecular mechanistic studies of CSCC as well as for early detection, targeted therapy and prognostic evaluation of CSCC patients in the future.

Acknowledgements

The present study was supported by grants from the Fujian Province Science and Technology Plan Key Project (no. 2013Y0030) and the National Clinical Key Specialty Construction Program of China.

References


