ORIGINAL RESEARCH ARTICLE

Small RNA sequencing and differential expression of miRNAs in colorectal cancer

Kwan-Liang Lye¹, Shiran Mohd Sidik², Sabariah Abdul Rahman³, Yoke-Kqueen Cheah¹,4,5*

¹ Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
² Faculty of Medicine, National Defence University of Malaysia, 57000 Selangor, Malaysia
³ Faculty of Medicine, Universiti Teknologi Mara, 47000 Sungai Buloh, Malaysia
⁴ Cancer Research Laboratory, Institute of Bioscience, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
⁵ Centre for Diagnostic Nuclear Imaging, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

ABSTRACT

MicroRNAs (miRNAs) are small, non-coding RNAs that are involved in various cellular processes. Many studies have shown that miRNAs are dysregulated in various types of cancer. In Malaysia, there are limited published studies on miRNA expression profile for colorectal cancer, which is among the most common cancer here and other parts of the world. Next-generation sequencing platform was introduced in recent years and has revolutionized the biomedical research settings. In this study, we performed small RNA sequencing on Illumina HiSeq 2000 platform and found that 22 miRNAs were significantly differential expressed between cancer and normal tissues. Further validation on qRT-PCR on 5 of the miRNAs selected showed 3 of them were up-regulated (hsa-miR-106a, hsa-miR-135b, hsa-miR-21) while 2 were down-regulated (hsa-miR1, hsa-miR-504). These findings may lead to a simple profiling method to distinguish the risk of individuals in developing colorectal cancer.

Keywords: sexing; nested multiplex PCR; sex determination; single spermatozoon; bovine

Introduction

Colorectal cancer is the third most commonly diagnosed cancer in the world and are among the top causes of cancer related death. Colorectal cancer incidences worldwide are projected to reach 1.2 million new cases with over 600 thousand deaths[1]. In the multiracial ethnicity of Malaysia, colorectal cancer incidences were the commonest among the males and third most common cancer in females, with Chinese having a higher incidence of colorectal cancer as compared to other ethnicities[2]. Mortality rate for colorectal cancer is above 60% out of those diagnosed with the disease. Most of us are unaware of this disease until it is too late. This is due to the fact that colorectal cancer has few or no signs at all during the early stages of the disease. Hence, detection of the disease at the early stage poses us with some problems to deal with.

MicroRNAs (miRNAs) are small, non-coding RNAs that are found in genomes of invertebrates, vertebrates and plants[3]. They are found to play a key role in various biological processes such as cell proliferation, differentiation and apoptosis[4]. MiRNAs synthesis begins in the nucleus, whereby primary transcripts (pri-miRNAs) are processed into precursor miRNAs (pre-miRNAs) mediated by Drosha and Dicer enzymes[5]. The pre-miRNAs are then exported into the cytoplasm by Exportin-5 and cleaved by Dicer into a mature double-stranded miRNA. This strand is then incorporated into the RNA-induced silencing complex (RISC) from which the miRNA and its mRNA target interact. The miRNA will only interact with the mRNAs that contain anti-sense sequences. However, the interaction can still happen if they are partially complementary to each other[6].
Before the effect of miRNAs on gene expression can be studied, a robust method for profiling these miRNAs is required. Since 2005, a variety of massively parallel sequencing machines such as Illumina, SOLID and Roche platforms have been introduced in the scientific community. These massively parallel sequencing methods or better known as next generation sequencing (NGS) gives a new breath towards the approach of miRNAs profiling. NGS overcomes the limitations such as low throughput, scalability, speed and resolution of the capillary electrophoresis sequencing method\[^7\]. This method has been widely used for miRNAs expression profiling in various types of cancers\[^8\]. To the best of our knowledge, the study of NGS colorectal cancer profiling is limited in Malaysia. Therefore, this study is to provide the insights of the miRNAs profile between colorectal cancer and adjacent normal tissues in Malaysia.

**Materials and Methods**

**Patient samples and ethics statement**

Twelve colorectal cancer (adenocarcinoma) patients were selected from Hospital Kuala Lumpur, Malaysia. Written informed consent was obtained from all patients for the use of tissue samples in this study. This project was approved by the local Ethics Committees (National Medical Research Register, NMRR and Hospital Kuala Lumpur, HKL). The NMRR reference number is NMRR-09-604-4252 and the HKL reference number is HKL/98/AM.882. Tumor tissues and adjacent normal tissues (~5-8cm apart from tumor tissue) were obtained during surgery at Hospital Kuala Lumpur and was immediately stored in RNAlater (Ambion, USA) for 1-2 hours at room temperature before being kept in -80 °C until further processing.

**Sample processing and RNA extraction**

The miRNA-enriched total RNA was extracted using innuPrep micro RNA kit (Analytik Jena, Germany). Prior to extraction, 20 mg of tissue was placed in a sterile 1.5 ml microcentrifuge tube and was homogenized in 450 µl of lysis solution using a rotor tissue homogenizer. After that, the sample was centrifuged at 14,000 × g for 2 minutes to remove the unlysed materials. The supernatant was then subjected to subsequent extraction steps following the manufacturer’s protocol. The concentration and purity of the extracted RNA were measured using Biophotometer 6131 (Eppendorf, USA). RNA integrity was also confirmed by obtaining the RNA integrity number (RIN) by using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). A total of 4 samples (2 cancer + 2 adjacent normal tissue) were chosen for Small RNA Sequencing based on the standard requirement needed for Next-Generation Sequencing (NGS) samples.

**Small RNA sequencing by Illumina HiSeq 2000**

Small RNA sequencing was performed on Illumina HiSeq 2000 (Illumina, USA) platform according to manufacturer’s protocol. Briefly, small RNA that was extracted will undergo adapter ligation and reverse transcribed into cDNA. The cDNA was subsequently amplified by PCR. After PCR purification, the miRNA libraries were sequenced on Illumina HiSeq 2000. Data analysis process started with the data cleaning of the 50 nucleotide (nt) sequence tags from HiSeq sequencing. Low quality reads and reads which is shorter than 18 nt were trimmed and adapter sequences were removed. The clean reads obtained were then mapped to a human reference genome (hg19) using SOAP v2.0 to analyze the miRNAs expression and distribution on the genome. To separate miRNAs from other ncRNA, we annotated the small RNA tags to rRNA, scRNA, snoRNA, and tRNA from Genbank. Then, we mapped the tags with sequences from Rfam and removed matched tags from unannotated tags.

For the differential expression analysis, known miRNA expression of two samples (tumor and adjacent normal tissue) was compared to obtain the differentially expressed miRNA. First, the miRNA expression in the two samples were normalized to get the expression of transcript per million (TPM) as shown below. The fold-change was then calculated from the normalized expression and the log2 ratio plot and scatter plot was generated.

\[
\text{Normalized expression} = \frac{\text{Actual miRNA count}}{(\text{Total count of clean reads} \times 1000000)}
\]

\[
\text{Fold change} = \log_2 (\text{tumor/normal})
\]

**MiRNAs validation by real-time PCR**

Five miRNAs were chosen for validation (Table 1) by real-time PCR. The miRNA-enriched total RNA extracted as mentioned above was converted to cDNA using the NCode™ miRNA first-strand cDNA synthesis kit (Invitrogen, USA) according to the manufacturer’s protocol. The universal qPCR primer provided in the kit was used as the reverse primer. The forward primer was designed from the NCode™ miRNA database (http://esience.invitrogen.com/ncode). The levels of miRNAs expression were
quantified by real-time PCR using Kapa SYBR® Fast qPCR kit (Kapa Biosystems, USA) according to the manufacturer’s protocol. The real-time PCR was run on Rotorgene 6000 (Qiagen, USA) according to manufacturer’s protocol. Rnu48 was used as endogenous control after its expression was found stable when tested across a set of miRNA samples from colorectal cancer patients. A no template control (NTC) was included to assess the specificity of the qPCR reaction. Fold-change of miRNAs expression was calculated using the 2-ΔΔCt method. Prediction of putative targets via in silico analysis

Targets of the differentially expressed miRNAs were predicted using DIANA microT v3.0 with the Score Threshold set at 8.0 and the KEGG (Kyoto Encyclopedia of Genes and Genomes) filter was set for only cancer pathways. DIANA microT v3.0 is an algorithm for the prediction of miRNA target genes and the search was conducted based on either known miRNAs or novel miRNAs.

Results

From the small RNA sequencing by HiSeq 2000, we managed to obtain clean reads ranging from 31-42 million reads. As shown in Figure 1A, the length distribution of the small RNAs mainly falls in the range of 20-25 nucleotides which conforms with the size of small RNAs. A small portion (~13%) of unique small RNAs is present in both the cancer and normal sample. However, they make up the majority number (~95%) of the total small RNA reads (Figure 1B). The distribution of the small RNAs among the different categories after annotation showed that miRNAs are the most abundant small RNA present with nearly 50% of the total reads (Figure 1C). For sample quality check, the rRNA proportion should be less than 40% in high quality samples as was the case for this sample (<20%).

For the differential expression analysis, known miRNAs expression of cancer samples and normal samples were compared and shown in a scatter plot (Figure 2). Using 2-fold expression as the cutoff value, 22 miRNAs were significantly differential expressed between cancer and normal tissues (Table 2). From the list of differentially expressed miRNAs, 5 of them were chosen and further analyzed by qPCR. The results of the qPCR conform to the results obtained from the small RNA sequencing. The up-regulated miRNAs are hsa-miR-106a, hsa-miR-135b and hsa-miR-21, while hsa-miR-1 and hsa-miR-504 are down-regulated (Table 3).

Using DIANA microT v3.0 with Score Threshold set at 8.0 and KEGG filter only for cancer pathways, target genes for the 5 miRNAs were obtained and summarized in Table 4 (Refer to Supplementary File). Among the 5 miRNAs, hsa-miR-106a has the most number of target genes, while hsa-miR-21 has the least. Among the interesting targets are adenomatous polyposis coli (APC) targeted by hsa-miR-135b, PTEN targeted by hsa-miR-106a, Bcl2 targeted by both hsa-miR-106a and hsa-miR-504, and also Kras gene which is targeted by hsa-miR-1.

Discussion

Differences in the expression profiles of miRNAs have been observed in many types of cancers. Dysregulation of specific miRNAs have been shown to be involved in carcinogenesis of colorectal cancer. However, the expression profiles of
Small RNA sequencing and differential expression of miRNAs in colorectal cancer

Figure 1. Small RNA sequencing of selected colorectal cancer and adjacent normal tissue. Figure 1A showed the length distribution of the small RNA obtained from next generation sequencing. Figure 1B showed the total sRNA and unique sRNAs distribution in both the cancer and normal samples. Figure 1C showed the annotation of total sRNAs with miRNAs being the most abundantly present sRNAs in the study.

Figure 2. Scatter plot of miRNA expression level of C38 against N38. In this study, it was shown that the present of up-regulated miRNAs are more in cancer as compared to the adjacent normal tissues.
Table 2. Summary of normalized reads in cancer and normal tissue and the log2 fold change

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Normalized reads in cancer tissue</th>
<th>Normalized reads in adjacent normal tissue</th>
<th>Log2 Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C34</td>
<td>C38</td>
<td>N34</td>
</tr>
<tr>
<td>hsa-miR-1</td>
<td>383.63</td>
<td>134.59</td>
<td>1044.89</td>
</tr>
<tr>
<td>hsa-miR-106a</td>
<td>2.87</td>
<td>5.07</td>
<td>0.12</td>
</tr>
<tr>
<td>hsa-miR-135b</td>
<td>2.57</td>
<td>9.86</td>
<td>0.01</td>
</tr>
<tr>
<td>hsa-miR-21</td>
<td>21340.77</td>
<td>29230.85</td>
<td>1156.70</td>
</tr>
<tr>
<td>hsa-miR-504</td>
<td>2.33</td>
<td>0.10</td>
<td>11.01</td>
</tr>
<tr>
<td>hsa-miR106b</td>
<td>147.04</td>
<td>202.97</td>
<td>17.81</td>
</tr>
<tr>
<td>hsa-miR-133a</td>
<td>4.28</td>
<td>1.18</td>
<td>41.53</td>
</tr>
<tr>
<td>hsa-miR-145</td>
<td>136.74</td>
<td>98.53</td>
<td>703.39</td>
</tr>
<tr>
<td>hsa-miR-17</td>
<td>281.68</td>
<td>491.26</td>
<td>17.63</td>
</tr>
<tr>
<td>hsa-miR-182</td>
<td>6244.02</td>
<td>14471.26</td>
<td>360.12</td>
</tr>
<tr>
<td>hsa-miR-183</td>
<td>4520.81</td>
<td>9650.95</td>
<td>345.86</td>
</tr>
<tr>
<td>hsa-miR-203</td>
<td>1077.91</td>
<td>4620.58</td>
<td>181.02</td>
</tr>
<tr>
<td>hsa-miR-224</td>
<td>215.81</td>
<td>468.39</td>
<td>15.49</td>
</tr>
<tr>
<td>hsa-miR-31</td>
<td>1.47</td>
<td>39.41</td>
<td>0.25</td>
</tr>
<tr>
<td>hsa-miR-3180</td>
<td>129.15</td>
<td>7.50</td>
<td>1.85</td>
</tr>
<tr>
<td>hsa-miR-424</td>
<td>13.96</td>
<td>25.61</td>
<td>0.43</td>
</tr>
<tr>
<td>hsa-miR-452</td>
<td>218.14</td>
<td>649.05</td>
<td>16.05</td>
</tr>
<tr>
<td>hsa-miR-454</td>
<td>4.38</td>
<td>6.79</td>
<td>0.18</td>
</tr>
<tr>
<td>hsa-miR-552</td>
<td>20.68</td>
<td>27.22</td>
<td>1.21</td>
</tr>
<tr>
<td>hsa-miR-627</td>
<td>2.98</td>
<td>18.34</td>
<td>0.15</td>
</tr>
<tr>
<td>hsa-miR-7</td>
<td>13.53</td>
<td>103.93</td>
<td>1.79</td>
</tr>
<tr>
<td>hsa-miR-96</td>
<td>23.09</td>
<td>35.47</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Table 3. RT-qPCR validation on selected hsa-miRNAs

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Fold change log 2 (cancer/normal)</th>
<th>P-value</th>
<th>Cancer vs. normal</th>
<th>Chromosomal location</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-1</td>
<td>-3.2</td>
<td>0</td>
<td>Down-regulated</td>
<td>20:61151513-61151583 (+)</td>
</tr>
<tr>
<td>hsa-miR-106a</td>
<td>1.5</td>
<td>9.5050303795997e-14</td>
<td>Up-regulated</td>
<td>X:133304228-133304308 (-)</td>
</tr>
<tr>
<td>hsa-miR-135b</td>
<td>4.4</td>
<td>1.88969745195566e-78</td>
<td>Up-regulated</td>
<td>1:205417430-205417526 (-)</td>
</tr>
<tr>
<td>hsa-miR-21</td>
<td>2.8</td>
<td>0</td>
<td>Up-regulated</td>
<td>17:57918627-57918698 (+)</td>
</tr>
<tr>
<td>hsa-miR-504</td>
<td>-5.6</td>
<td>1.90129855717238e-50</td>
<td>Down-regulated</td>
<td>X:137749872-137749954 (-)</td>
</tr>
</tbody>
</table>

miRNAs involved in colorectal cancer amongst Malaysian have not been reported. Therefore, this study is aim to elucidate the expression profiles of miRNAs in Malaysian diagnosed with colorectal cancer.

In this study, we found that miRNAs are differentially expressed between tumors and normal tissues. We found 22 miRNAs that are differentially expressed between cancer and normal sample. Out of these, 18 miRNAs were up-regulated and 4 were down-regulated. From these miRNAs, we picked
5 miRNAs that were significantly differentially expressed and 3 of them were up-regulated; hsa-miR-106a, hsa-miR-135b and hsa-miR-21, while 2 of them were down-regulated; hsa-miR-1 and hsa-miR-504. This result is in agreement with previous studies that showed more miRNAs that were up-regulated than down-regulated in cancers profiling.

The down-regulation of hsa-miR-1 in this study is in agreement with a study on colon cancer samples by Sarver et al. Besides, hsa-miR-1 was also found to be down-regulated in a few other types of cancers such as rhabdomyosarcomas, lung cancer, and also thyroid cancer. This finding was similar with Hamford et al., whereby using high-throughput sequencing and comparing between paired tumor and normal tissue, they managed to identify 37 dysregulated miRNAs with miR-1 among the down-regulated miRNAs. It was also shown that hsa-miR-1 inhibits the tumorigenic property of lung cancer cells by down-regulating oncogenic genes such as MET and FoxP1. These findings support the fact that hsa-miR-1 could play a role as tumor suppressor, whereby down-regulation of this miRNA leads to colorectal cancer and other types of cancer formation.

Meanwhile, Hu et al. showed the negative regulation of p53 tumor suppressor gene by hsa-miR-504. Overexpression of hsa-miR-504 leads to the decrease of p53 protein levels and function in cells, thus increasing the tumorigenicity of cells. Another study also found out that hsa-miR-504 was significantly up-regulated, causing the down-regulation of nuclear respiratory factor 1 leading to radio-resistance in nasopharyngeal carcinoma. Expression of hsa-miR-504 was also significantly higher in pancreatic duct adenocarcinoma as compared to normal pancreatic tissue and was correlated with poor prognosis and survival rate in patients. However, the hsa-miR-504 expression in this study showed a down-regulation as compared to the normal tissue. There are a few studies showing the expression of hsa-miR-504 which was down-regulated in cancers such as in the hypopharyngeal squamous cell carcinoma and glioma. This indicated that hsa-miR-504 could be playing different roles in different cancers at different localities.

Besides this, we have also observed up-regulation of hsa-miR-106a in colorectal cancer samples. This finding is in concordance with a few other studies done before. Schetter et al. showed that hsa-miR-106a is among the miRNAs that are up-regulated in his two study cohorts of colon cancer patients. An in vivo study also showed that hsa-miR-106a was overexpressed in T-leukemia. However, the same miRNA was found to be down-regulated in glioma and its low expression was significantly related with the high levels of E2F1 protein and high grade glioma. This further enhanced the fact that miRNAs can act differently in various types of cancer.

Another miRNA, hsa-miR-135b has also been shown to be up-regulated in different types of cancers including breast cancer, colon cancer, prostate cancer and osteosarcoma. McBee et al. reported that hsa-miR-135b was overexpressed in cervical cancer and even serve as potential marker for progression from dysplasia to invasive cervical cancer. In another study using real time PCR, hsa-miR-135b was among the significantly overexpressed miRNAs in comparison between colorectal cancer and adjacent normal tissue. Regulation of APC gene and Wnt pathway by this miRNA was illustrated by Nagel et al. In his study, hsa-miR-135b was shown to target the 3' UTR of APC, thus suppressing its expression and induce downstream Wnt pathway activity.

Among these miRNAs, hsa-miR-21 has been extensively evaluated in literature and was shown to be related to numerous cancers. It is among the first miRNA to be described as oncomiR and targets a host of tumor suppressor genes such as PTEN, Pdcd4, and Bcl. Qi et al. showed that hsa-miR-21 was up-regulated in flat epithelial atypia (FEA) of breast. However, in that study, immunohistochemical staining showed no clear relation between the staining pattern of hsa-miR-21 and the target genes. In another study, hsa-miR-21 was found to down-regulate tumor suppressor Pdc, protein and triggers invasion and metastasis in colorectal cell line. Hsa-miR-21 was also shown to play a role in cell cycle progression and DNA damage checkpoint activation via Cdc25A. This miRNA down-regulates Cdc25A after being induced by serum starvation and DNA damage. Meanwhile, Seike et al. showed that EGFR mutation leads to higher expression of miR-21 that contributes to lung carcinogenesis.

The introduction of next generation sequencing technology has revolutionized the biomedical research settings. From the year 2005, a variety of next generation sequencing platforms were widely used such as 454 Genome Sequencer FLX from Roche, SOLID Sequencer from Applied Biosystem and Genome Analyzer from Illumina. Next generation sequencing has enabled the sequencing of complete set of miRNAs present in the total
RNA. From here, the data obtained can be used to study miRNAs expression profile, identify miRNAs isoforms and also to predict novel miRNAs and their putative targets\[^{40,41}\]. Various studies have utilized this next generation sequencing technology. Chen et al. showed that miRNAs are present in serum and plasma of human and the levels of the miRNAs in serum are stable, reproducible and consistent among individuals of the same species\[^{42}\]. Another study on Illumina’s Genome Analyzer platform showed that by comparing the matched tumor and normal pairs, mutational landscape between individuals can be assessed and identified\[^{43}\]. Wu et al. also utilized the next-generation sequencing technology to profile miRNAs in breast cancer and found 19 miRNAs were upregulated by at least 2-fold expression difference when comparing between cancer and adjacent normal tissue\[^{44}\]. All these discoveries lead towards better diagnosis and prognosis of the diseases.

**Conclusion**

This is among the first study in Malaysia to compare the miRNAs expression profile from colorectal cancer patients utilizing the next-generation sequencing technology followed by qRT-PCR validation. The miRNAs that were differentially expressed between cancer and normal tissue can be potentially used as biomarkers for colorectal cancer. Nevertheless, further investigation has to be carried out on a larger sample size with more diverse stages of cancers to identify miRNAs that are of better diagnostic and prognostic value.

**Acknowledgement**

The authors gratefully acknowledges financial support from the MAKNA (National Cancer Council Malaysia) Research Award for the financial support, and our research collaborators from the Biomedical Sciences Department, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia.

**References**


