Computational and Experimental Tools of miRNAs in Cancer

Esen Çakmak

Kirşehir Ahi Evran University, Mucur Health Services Vocational School, Department of Medical Services and Techniques, Medical Laboratory Techniques Program, Mucur Campus, Kirşehir, Turkey

Abstract

MicroRNAs (miRNAs) are short non-protein coding and single-stranded small RNA molecules with a critical role in the regulation of gene expression. These molecules are crucial regulatory elements in diverse biological processes such as apoptosis, development, and progression. miRNA genes have been associated with various human diseases, particularly cancer, and considered as a new biomarker. After the discovery of miRNAs, many researches have focused on identifying and characterizing miRNA genes in cancer. The various expression levels of miRNAs between cancer cells and normal cells are very crucial to diagnosis, prognosis, and treatment of many cancers. Many computational and experimental tools have been employed to characterize miRNAs. However, there exist some challenges in identifying miRNA using both computational and experimental tools due to miRNA features. The present review briefly introduced miRNA biology and certain computational and experimental tools for identifying and profiling miRNAs in cancer. Furthermore, we presented the advantages and challenges of these tools.

Keywords: miRNAs, Cancer, Computational tools, Experimental tools

Introduction

MicroRNAs (miRNAs) are non-protein coding, 18-to-25-nucleotide long and single-stranded small RNA molecules. Ambros and colleagues detected the first miRNA (lin-4) in a nematode, *Caenorhabditis elegans.*

miRNAs negatively regulate gene expression in transcriptional or post-transcriptional levels and cause the degradation of target mRNA. These molecules are crucial regulatory elements of biological processes such as apoptosis, development, disease generation, and progression. It is believed that miRNAs regulate at least one third of all human genes.

Biogenesis of miRNAs commences in nucleus with the synthesis of primary miRNA (pri-miRNA), a long transcript of miRNA genes. RNase III enzyme Drosophila processes pri-miRNAs in order to generate 60-70 nucleotide hairpin structures known as precursor miRNA (pre-miRNA). After pre-
miRNAs are exported into the cytoplasm, Dicer enzyme cleaves them, forming mature miRNAs. One strand of the double-strand RNAs is miRNA; other strands are complementary miRNAs with inactive forms, called miRNA*. miRNAs are functionally incorporated into the RNA-induced silencing complex (RISC). These small RNAs are mostly bound partially (first 2-7 nucleotides, called seed site) to specific sequence in the 3’ untranslated region of target mRNA while they are rarely observed in the 5’ untranslated region or the coding region. Although one miRNA regulates the expression of many genes, several different miRNAs are bound to a single mRNA transcript. Therefore, gene expression is negatively regulated to degrade, destabilize, or translationally inhibit mRNAs.

In the genome, miRNA genes are mostly located in the intronic parts of protein-coding genes; however, they are rarely seen in exonic parts and among genes. More than 50% of miRNA genes have been identified in cancer-related genomic regions or fragile sites. miRNA genes can be considered as a new biomarker for various human diseases. The research associated with miRNAs is based on different expression levels between cancer cells and normal cells. Knockdown or overexpression of miRNAs is an important approach to cancer pathogenesis. The expression levels of miRNAs indicate whether they are functionally oncogenes or tumor suppressor genes. If the expression levels of miRNAs decrease, they are tumor suppressor genes preventing cancer and controlling oncogenes, cell differentiation, or apoptosis. Overexpression of miRNAs supports tumor development through inhibiting tumor suppressor genes. These miRNAs are known as oncogenes which lead to cancer. Role of miRNAs in cancer was primarily reported in B cell chronic lymphocytic leukemia (CLL). Many miRNAs have been identified as oncogenes or tumor suppressor genes in various cancer cells. Two miRNAs, namely miR-15a and miR-16-1, are located at chromosome 13q14. These miRNAs negatively regulate B cell lymphoma 2 (Bcl2) overexpression in CLL patients at post-transcriptional levels. It is estimated that these miRNAs are tumor suppressor genes in CLL. However, miR-155 overexpression in human B cell lymphomas, lung cancer, and breast cancer has features of an oncogene.

Detection, profiling, and quantification of miRNA expression in cancer

Since miRNAs have been reported to regulate gene expression, many studies have been conducted on their functions. Specifically, the expression level of miRNAs is a crucial factor for the diagnosis and prognosis of many cancers, including chronic lymphocytic leukemia, lung, breast, ovarian, prostate, bladder, kidney, colorectal, pancreatic, and gastric cancers. Comparison of miRNA expression levels between cancer and normal tissues is a very effective approach for an early detection, diagnosis, prognostic, and therapy. Therefore, miRNAs are supposed to be a new biomarker in cancer.

To date, many approaches have been performed to detect and quantify miRNAs. These approaches can be classified into three categories, namely experimental, computational, and forward genetics. Forward genetics approach is rarely used in miRNA identification. In this classical approach, the phenotype is known, while the genotype associated with the phenotype is not. Experimental approaches include main tools such as northern blotting, direct cloning, and sequencing, in-situ hybridization, next generation sequencing, real-time PCR, and microarray techniques. Each miRNA approach has its own advantages and disadvantages. The ideal method for miRNA detection should fulfill several requirements such as being sensitive, specific, reproducible, quantitative, easy to perform, and low cost. The experimental and computational approaches are described in the following.

Computational approaches

miRNAs are expressed at low levels and their expression levels show important variability in different tissues and under environmental conditions. They are identified using experimental approaches in various organisms. It is difficult
miRNAs in Cancer

to determine miRNA features because they have small sizes, and their experiments are time-consuming, expensive, and labor-intensive. Therefore, computational approaches have been developed as alternative methods for the identification of miRNAs.25, 26 However, there are some computational challenges in identifying miRNAs. Millions of putative miRNAs with appropriate sequence lengths can be obtained in a genome by computational approaches. Folding these sequences may increase the complexity. Many putative miRNAs may represent false positive results. Therefore, it is recommended that putative miRNAs be validated with experimental approaches.26

Computational approaches use some properties of miRNAs to predict putative miRNAs such as conservation of miRNAs among species, minimal folding free energy, and formation of stable stem-loop structure by pre-miRNAs.27, 28 Initial computational approaches are based on sequence or structure conservation of miRNAs. In earlier studies, miRNAs were cloned in many organisms, resulting in the identification of new miRNAs based on homology searches with known miRNAs.29-31 The main principle in these approaches is to compare genomes and filter out hairpins that are phylogenetically non-conserved in the related species. However, finding miRNAs that are species specific and irrelevant to previously known organisms is a major issue in these approaches.24, 28 To surmount this problem, computational approaches such as MFOLD32 and RNAfold33, use the features of RNA secondary structure such as hairpin-like secondary structures and free-energy minimization.34

MiRscan35 and miRseeker36 computational approaches have been improved for the prediction of miRNA genes based on sequence, structure, and closely related species conservation. MiRscan uses seven miRNA features associated with weights and has a specificity of 70% and a sensitivity of 50%. A 21-nucleotide window is slid along putative stem-loops, and putative hairpins are given scores in this approach.34,37 MiRseeker is another approach which predicts miRNA genes in Drosophila genome and has concepts similar to MiRscan.38 Researchers that developed miRseeker and MiRscan have verified new miRNAs using northern blotting techniques. Also, miRseeker relies on thermodynamics location, while MiRscan is based on base pairing, hairpins, bulges/loops, and the mature location of secondary structure.34

Phylogenetic shadowing is a computational approach for identification of novel human miRNAs. This approach is based on multiple sequence comparisons of closely-related species in both conserved and non-conserved regions.39-41

Machine learning approaches have also been successfully developed for miRNA identification. These approaches employ a specific computer program together with training data based on the algorithms and methods. The training data are the sequence and structure knowledge of miRNAs obtained from different species. These data are classified into positive examples (miRNAs downloaded from the database miRBase)42 and negative examples (the random control set of non-miRNA sequences). These approaches include several algorithms such as naive Bayes, support vector machines (SVM), and hidden Markov models (HMM).37,41 We summarize some of these algorithms below.

Nam et al. (2005) introduced a probabilistic colearning model, known as ProIMR, for miRNA discovery. ProIMR simultaneously considers both the structure and sequence of miRNA precursors and employs a HMM, which is a statistical model. HMM topology and states are dependent on prior information and assumptions. Moreover, the exact probabilities of HMM are provided from the accumulated data.38, 41, 43

Yousef et al. described a new machine learning algorithm that employs a simple Naive Bayes classifier to predict miRNA genes.44 This approach automatically generates a model with training data collected from verified miRNAs. Therefore, it reduces the false positive rate with comparative analysis in order to predict miRNA genes. In addition, it allows the prediction of non-conserved miRNAs.44, 45

RNAmicro45 is an SVM based on multiple sequence alignments. RNAmicro separates pre-
miRNAs from other types of non-coding RNAs (ncRNAs).\textsuperscript{46} RNAmicro was applied to genome wide RNAz surveys in mammals, urochordates, and nematodes.\textsuperscript{45,47} MiRFinder is a computational learning method SVM, built for pre-miRNA prediction. This tool is able to determine the differences between miRNA and non-miRNA hairpins. This computational approach comprises of two main stage as search for hairpin candidates and exclusion of non-robust structures in genome. MiRFinder has a good performance compared with other tools such as RNAmicro and miRscan.\textsuperscript{48}

Experimental approaches

Computational approaches can only identify conserved miRNAs in various organisms. To predict miRNAs, these approaches require nucleotide sequence for algorithms.\textsuperscript{25,49} The computational methods predict millions of putative miRNAs in the genome. However, these putative miRNAs can cause false positive results.\textsuperscript{26} Therefore, experimental approaches are required to both validate the computational results and identify novel miRNAs.

A wide range of experimental approaches have been performed for detection, profiling, and quantification of miRNA in cancer. This section highlights certain important techniques for miRNA identification such as direct cloning and sequences, northern blotting, microarrays and quantitative RT-PCR.

Direct cloning and sequencing

Direct cloning and sequencing is a widely used molecular technique for detecting RNAs and studying gene expressions. This technique is modified to identify miRNAs as miRNA serial analysis of gene expression (miRAGE). miRAGE consists of direct miRNA cloning and serial analysis of gene expression (SAGE).\textsuperscript{50} miRAGE method includes isolation of total RNA, adaptor ligation, reverse transcription, tag purification, concatenation, cloning and sequencing steps.\textsuperscript{31,51} For instance, samples obtained from colorectal cancer tissues and normal tissues were analyzed using miRAGE and 200 known mature miRNAs, 133 putative miRNA and 112 uncharacterized miRNA* forms were identified.\textsuperscript{52} Many of the initial miRNAs identified and determined the expression levels with miRAGE in various species. However, miRAGE requires a large amount (1 mg) total RNA, has a low throughput and is very labor intensive.\textsuperscript{51}

Northern blotting

miRNAs were primarily determined by use of northern blotting technique.\textsuperscript{53} In this technique, the expression properties and sizes of miRNAs are specified and putative miRNAs are verified. The technique, developed in 1977,\textsuperscript{54} is similar to the general blotting technique. Northern blotting usually includes the following steps: the extraction of the total RNA from a homogenized tissue sample, separation of RNA on agarose gel, transfer of separated RNA onto the membrane, synthesis of labeled cDNA probe, hybridization of denatured probe with the membrane, washing of the membrane, and exposure with the film.\textsuperscript{51} Although this technique has been widely used for miRNA expressions, it has several challenges such as low throughput, low sensitivity, and time consumption. However, improved northern blotting protocols are more sensitive and safe in detecting miRNAs. These protocols are generally different in terms of labeling and designing the probes. Also, a new northern blotting method has been developed by use of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) to cross-link RNAs into membranes. The new method is more specific compared with the traditional cross-linking methods.\textsuperscript{55-57}

In situ hybridization

In situ hybridization (ISH) is an important tool for determining miRNA expressions. ISH utilizes a labeled complementary DNA sequence or complementary RNA sequence for the localization of a particular DNA or RNA sequence in tissues or cells. There are several different applications of ISH, including DNA ISH, RNA ISH, and fluorescent DNA ISH (FISH); these applications are basically the same except for a few features such as the probes. The probe is labeled with either radio-labeled, fluorescent-labeled, or
antigen-labeled bases. The labeled probe is hybridized to the particular DNA or RNA sequence at elevated temperatures and the excess probe is removed with washing procedure. Afterwards, localization and quantification steps are performed via autoradiography, fluorescence microscopy, or immunohistochemistry, respectively. The major advantage of ISH over other methods is its ability to determine miRNA expression profiling at the cellular level. Furthermore, ISH enables a semi-quantitative analysis of miRNA expressions. Expression profiling approaches such as northern blots or PCR are limited to determining cellular sources or localization of miRNA expressions due to different cell types. Thus, ISH is an ideal method for determining the cellular localization of different RNAs. The major challenge of ISH is its low efficiency because miRNAs are very small in length. The challenge can be eliminated using LNA technique in ISH. LNA modified oligonucleotide probe increases hybridization affinities and significantly improves specific miRNA detection. However, ISH performs some techniques for the detection of miRNAs (e.g. in formalin-fixed and paraffin embedded tissue specimens (FFPE) using LNA probes).

Quantitative RT-PCR

Real-time reverse transcriptase PCR (qRT-PCR) is a powerful technology with high sensitivity, accuracy and practical ease for quantification of gene expression. Detection and quantification of miRNAs are technically difficult due to their small size and variable expression levels. Northern hybridization, cloning, and microarray technology have been commonly performed in the detection and quantification of miRNAs; however, these applications are more limited in terms of sensitivity and specificity compared with real-time reverse PCR.

The first miRNA discovery using real-time PCR technology occurred with the characterization of miRNA precursors. However, miRNA precursors are unconformable with mature miRNAs in terms of cellular concentrations. To overcome this challenge, a novel miRNA quantification method has been developed consisting of stem-loop RT and TaqMan PCR approaches for the analysis of mature miRNAs. Stem-loop RT primers are more efficient and more specific compared with conventional primers. TaqMan miRNA assays can even determine one nucleotide difference among the related miRNAs. Furthermore, genomic DNA contamination does not affect TaqMan miRNA assays. This method is comprised of reverse transcriptase (RT) and real time PCR steps. In one step, the stem-loop RT primers are hybridized to a particular miRNA and reverse transcribed with a MultiScribe reverse transcriptase. Next, the RT products are quantified by TaqMan PCR. The stem-loop RT primer is an important invention for the characterization of mature miRNAs. This primer is also able to provide some advantages. Stem-loop RT primer contains a stem-loop and a 3’ overhang complementary to the 3’ miRNA sequence within 5-8 nucleotides. This method presents a fast, accurate, highly specific, sensitive, and wide dynamic range for the detection and quantification of miRNAs. It is also a unique tool for the early diagnosis of diseases such as cancer. Different techniques such as SYBR green and miR-Q can replace the TaqMan technique to analyze miRNAs in the RT-PCR.

Microarray analysis

Microarray technology, firstly developed in 1995, is a powerful high-throughput approach to identify genes in gene expression. Microarray approach has been adapted for the identification of miRNA genes and performed to screen a genome-wide miRNA profiling in normal and disease tissues. Moreover, it can reveal miRNA expression signatures associated with diagnosis, prognosis, and therapeutic applications.

In microarray techniques, nucleic acid hybridization is basically performed between the target molecule and its complementary probe. Antisense probes using microarray applications are spotted and immobilized onto a solid surface such as a glass or a quartz support platform. Microarray technology simultaneously performs multiple nucleic acid hybridizations such as
genotyping and measures the expression levels of many genes. In the microarray procedures, miRNAs are primarily isolated from a given sample and labeled with fluorescent dyes. The labeled miRNAs are hybridized with the miRNA microarray. Next, miRNAs emit the fluorescence signals binding to the corresponding probes. The emitted signals are then analyzed for detection and relative quantification of miRNAs. In microarray analysis, sample labeling and probe design are the most important points in miRNA profiling. The difference between mature miRNAs and their precursors, the short length of miRNAs, the sequence similarity among miRNAs, low abundance of certain miRNAs, and melting temperatures of mature miRNAs pose some challenges in miRNAs microarrays.

The major factor is probe design for miRNA microarrays. The oligonucleotide probes generally include a linker sequence and capture sequence regions. The linker sequences contain an amine-modified terminus together with poly (dT) or poly (dA); the capture sequences are usually complementary sequences in immobilized probes. Hundreds to thousands of oligonucleotide probes are employed in the miRNA microarray. Furthermore, nucleotide hybridization is performed at a single temperature. miRNA probes have a wide range of melting temperatures (Tm), hence the challenges occurring in hybridization. Furthermore, due to the short length of miRNAs, it is impossible to regulate the probe sequences for the normalization of melting temperature. In the miRNA microarray, several methods have been developed to eliminate these issues. LNA is an improved method in miRNA microarrays performed for the normalization of melting temperature in the capture probes.

The other major factor is the labeling of miRNA microarrays. A number of strategies have been employed in miRNA labeling, which can be divided into two groups, namely direct labeling and indirect labeling. Direct labeling is when miRNA molecules directly interact with fluorescent dye. This method is easy to perform and able to eliminate the problems caused by reverse transcription and PCR amplification. Indirect labeling is also labeled miRNA reverse transcript, the RT-PCR product of miRNA, or the in vitro transcript of miRNA instead of miRNA molecules. Microarray expression profiling of miRNAs has become a widely used technique, especially LNA approach, and is a high throughput method for the detection of many miRNAs in a single run. However, this technique has some limitations regarding the usage of miRNA profiling: it is insufficient to provide quantitative data and requires a large quantity of RNA for analysis. Moreover, due to the short length of miRNAs, this technique reduces the sensitivity and specificity of miRNA microarrays.

Next-generation sequencing (NGS)

In the sequencing technologies, Sanger and colleagues made the initial breakthrough in 1977, called Sanger sequencing. This technology, which is based on chain termination methods, is considered as a ‘first-generation’ technology used to complete human genome sequencing. After the human genome project, novel strategies were improved for field sequencing. NGS is one of these strategies which has significantly impacted molecular research. NGS technology is carried out by repeated cycles of polymerase-mediated nucleotide extensions or iterative cycles of oligonucleotide ligation. The major upsides of NGS are more nucleotide sequences (up to 3 GB per run), short read lengths (about 30-300 bp), and lower cost and manpower. NGS uses various platforms depending upon different founder companies such as Roche/454 System, Illumina/Solexa, and ABI/Solid. Data analysis of NGS platforms have similar strategies. However, there are different details in the initial data processing of NGS platforms. NGS is a high throughput tool within technologies of miRNA profiling. Despite the low abundance transcripts, this tool has high sensitivity and reproducibility. Also, it is independent on species.

NGS has been widely used to determine miRNA, sequence isoforms, novel miRNAs, and potential mRNA targets. In miRNA sequencing procedure, total RNAs are firstly isolated from the given sample and then fractionated by gel electrophoresis. Small
miRNAs in Cancer

RNAs are purified from gels and ligated by an adapter. cDNA libraries are created with reverse transcription and run once again in the gel electrophoresis for size selection. Finally, cDNA fragments are sequenced and analyzed.77

Conclusions
Since miRNA discovery, researchers have focused on the characterization of miRNA to determine their roles in various human diseases. Particularly, different expression levels of miRNAs between cancer cells and normal cells are important factors for cancer diagnosis, prognosis, and treatment. miRNAs detection approaches include computational and experimental approaches with subcategories. Due to the features of miRNAs, there are advantages and disadvantages in the usage of each approach. Conducting new studies on the identification of new miRNAs, miRNA action mechanism and miRNA-target interaction, mRNA may present a significant contribution to the treatments of cancer and other diseases.

Conflict of Interest
None declared.

References


