microRNA: emerging biomarkers in human disease and profiling challenges

J. Eunice Zhang
(University of Liverpool, UK)

Over the last decade, a rapid growth in the field of micro(mi)RNA research has been evident. Profiles of miRNAs have been shown to reflect pathological conditions and have been implicated as potential biomarkers in the diagnosis and evaluation of human diseases. In this article, an overview of miRNA discovery and biogenesis is provided. Recent advances in using circulating miRNA as biomarkers in human diseases are discussed, along with the challenges researchers are facing in miRNA profiling and future diagnostic prospects.

Discovery of miRNA

miRNAs encompass an extensive class of short, single-stranded, non-coding RNA molecules which are approximately 22 nucleotides in length. The first miRNA, *lin-4*, was discovered over 20 years ago in the roundworm *Caenorhabditis elegans*[1,2]. It was found that the gene *lin-4*, which controls developmental timing in *Caenorhabditis elegans*, did not encode a protein but was instead a small RNA molecule that down-regulates the gene expression of *lin-14* by partial complementary sequence pairing. However, the importance of miRNA was not fully appreciated at the time. It was not until years later, after the second characterized miRNA, *let-7*, was discovered to be highly conserved across species[3,4], that concerted efforts commenced among researchers to identify other miRNAs and elucidate their function. To date, more than 28,000 miRNAs have been identified in a wide variety of organisms. Approximately 400 miRNAs have been defined in plants, roughly 2000 in mouse and over 2500 miRNAs in humans[5].

miRNA biogenesis and gene regulation

The vast majority of functional miRNAs are generated via a classical route termed the canonical pathway (Figure 1). Starting in the nucleus, miRNA genes are transcribed by RNA polymerase II (Pol II) as primary miRNAs (pri-miRNAs), which range between hundreds and thousands of nucleotides in length, containing imperfectly base-paired hairpin structures with a 5′ cap and a 3′ poly(A) tail. These long pri-miRNAs are subsequently excised in the nucleus to produce shorter 60–70-nucleotide hairpin structures termed precursor miRNAs (pre-miRNAs) by a protein complex called Microprocessor, which includes Drosha, a ribonuclease III enzyme, and a double-stranded RNA-specific ribonuclease named DiGeorge syndrome critical region 8 protein (DGCR8). The pre-miRNAs are then transported out of the nucleus by exportin 5 (XPO5) into the cytoplasm in a Ras-related nuclear protein (Ran) GTP-dependent manner, where Dicer1 (another ribonuclease III enzyme) in complex with the transactivation-responsive RNA-binding protein (TRBP), a double-stranded RNA-binding protein, cleaves the pre-miRNA hairpins into mature duplex miRNAs (~19–24 nucleotides). The mature miRNA consists of the guide (antisense) strand and the passage (sense) strand. The guide strand is incorporated into the miRNA-induced silencing complex (miRISC), which contains Dicer1 and Argonaute (Ago) proteins, whereas the passage strand is unwound and is degraded.

miRNA biogenesis also occurs via alternative non-canonical pathways in the absence of ribonuclease III enzymes such as Dicer or Drosha. These non-canonical miRNAs, however, are poorly conserved and are of low abundance and are not discussed in this article.

It is now well established that miRISCs regulate genes at the post-transcriptional level in a sequence-specific manner. The miRISC interacts with the gene target through partial sequence complementation, usually within the 3′ untranslated region (UTR) of the target gene. This target recognition occurs predominantly via a crucial domain located in the 5′ end of the miRNA consisting of nucleotides 2–7, which is termed the 'miRNA seed'. The downstream nucleotides of miRNA (particularly nucleotide 8 and, less importantly, nucleotides 13–16) also contribute to

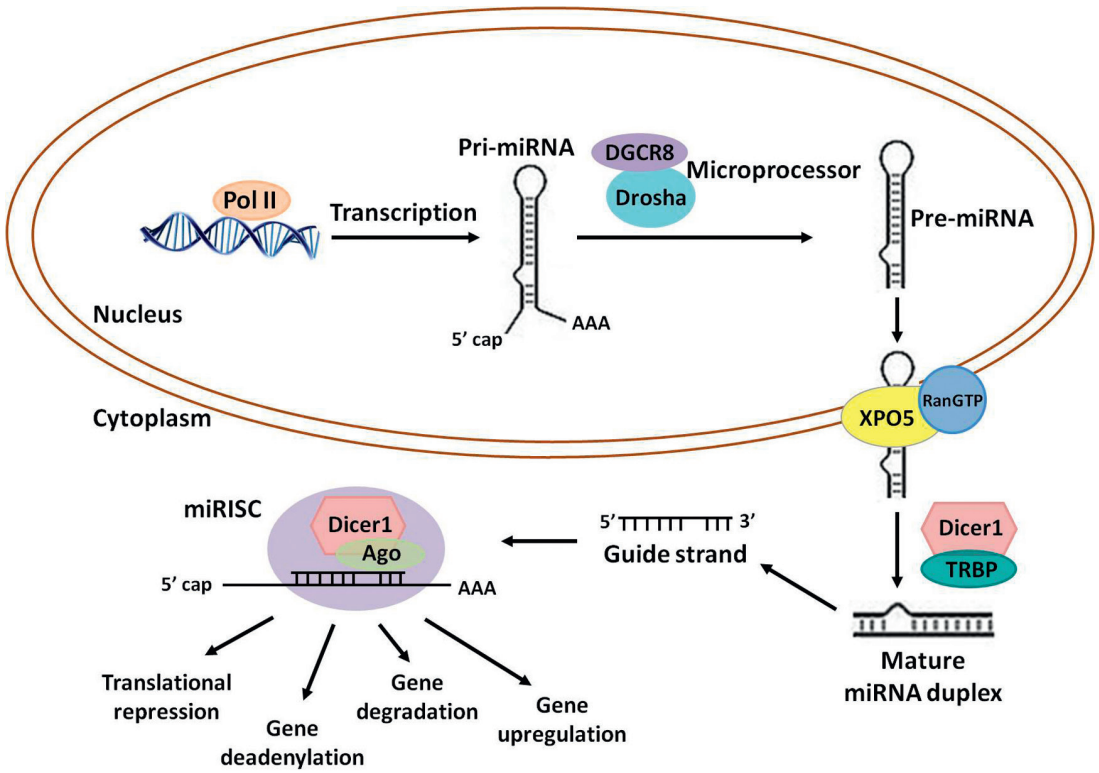

Figure 1. The biogenesis of miRNA via the canonical pathway

base-pairing with the gene. Nevertheless, a small subset of miRNAs also targets the 5′ UTR and/or the coding region of genes.

The extent of base-pairing between the miRNA and its gene target determines whether the gene is silenced through cleavage, translational repression or deadenylation. More recently, increasing experimental evidence indicates that miRNAs do not solely modulate down-regulation of genes, miRNAs can also stimulate gene expression by direct and indirect mechanisms in response to specific gene sequences, cofactors and cellular conditions[6].

Given that most target sites on the gene have partial base-pairing complementarity with their corresponding miRNA, individual miRNAs may target hundreds or thousands of different genes. Computational prediction has estimated that over 60% of human protein-coding genes contain at least one conserved miRNA-binding site[7], and a multitude of non-conserved target sites also exist on genes for different miRNAs, resulting in a complex regulatory network.

Studies have shown that the expression patterns of miRNAs are highly tissue-specific and miRNAs regulate diverse aspects of biological processes, including developmental and physiological processes. Aberration of individual or subset of miRNAs has been associated with many human diseases, such as cancer, heart diseases, neurological disorders, viral infections and metabolic diseases.

Circulating miRNA as biomarkers in human diseases

In addition to being primarily expressed intracellularly, numerous studies have recently shown that miRNAs are detectable outside cells in body fluids such as blood serum, plasma, urine and saliva. In contrast with the general assumption that RNA are unstable molecules, biochemical analyses indicated that circulating miRNAs are, in fact, highly stable, demonstrating resistance to ribonuclease digestion, extreme pH and temperature. Although the origin and biological mechanisms driving the secretion of extracellular miRNAs are still under debate, studies have shown that circulating miRNAs are shielded from degradation via three potential approaches: (i) by being packaged in microparticles such as exosomes, microvesicles and apoptotic bodies[8]; (ii) by being present in lipid vesicles such as high-density lipoprotein[9]; or (iii) by association with RNA-binding proteins (e.g. Ago2). More intriguingly, miRNAs have been reported to mediate cell–cell communication[10].

Interestingly, the expression patterns of circulating miRNAs have shown tight correlation with diseases such as heart disorders, cancer and drug-induced

Table 1. Examples of circulating miRNAs as biomarkers in human diseases

Disease	miRNA biomarker	Source
Acute myocardial Infarction	*miR-30a*	Plasma
Coronary arterial disease	*miR-126*	Plasma
Prostate cancer	*miR-141*	Serum
Bladder cancer	*miR-126, miR-182*	Urine
Acetaminophen-induced acute liver injury	*miR-122, miR-192*	Serum

liver injury (Table 1), heralding circulating miRNAs as promising non-invasive candidates for diagnostic and prognostic biomarkers, as well as predictors of drug response.

Challenges of miRNA profiling

The challenges of studying miRNAs are three-fold. First, because of the short length of miRNAs, traditional DNA-based methods are not sensitive enough to detect these sequences with any reliability. Secondly, highly homologous sequences across miRNA family members may differ by as little as one nucleotide, emphasizing the need for high specificity and the ability to discriminate between single nucleotide mismatches. Thirdly, compounded by the relatively low concentration of miRNA present in the circulation (typical levels of circulating miRNAs in serum were estimated to range between 200 aM and 20 pM[11]), accurate and consistent comprehensive measurement is difficult. To date, qPCR (quantitative PCR), microarray and sequencing-based approaches are the three major platforms that have been used for miRNA profiling.

Currently, the most commonly used technique for quantifying miRNAs is qPCR which is sensitive and quantitative. Amplification is initiated with a miRNA-specific primer and a stem–loop/poly(A) primer. Detection of amplified products uses either SYBR® Green or a TaqMan® probe. Significant drawbacks of this method lie with the specificity and differential amplification efficiency of the designed primers. Unlike conventional qPCR, only one flanking primer can be specific to the miRNA, so extreme care must therefore be taken to ensure that only one product is being amplified, especially when using SYBR® Green.

Microarray generally requires more starting material than qPCR and is the least quantitative of the three miRNA profiling platforms. Thousands of probes on the array enable parallel analysis of all known miRNAs, using fluorescence detection. Similarly to qPCR, distinguishing between miRNAs may be problematic owing to the sequence similarity between miRNAs.

Finally, next-generation sequencing (NGS) resolves some of the issues related to short and highly similar sequences of miRNAs as it does not rely on the design of primers or probes specific to each miRNA, allowing the measurements of a broad spectrum of miRNAs and the identification of novel miRNA sequences. However, of the three methods discussed, NGS requires the most input material, and there is a potential sizeable source of bias in the construction of sequencing libraries. Data analysis can also be relatively challenging, requiring a trained bioinformatician.

Future perspectives

More recently, electrochemical biosensors have been developed to allow for rapid and ultrasensitive detection of miRNAs using low input samples. Furthermore, measurements can be undertaken directly from crude samples, with very low detection limits. This approach holds a great promise to serve as point-of-care diagnostics.

As the functional roles of circulating miRNAs in human diseases are uncovered further, and coupled with improved methods of circulating miRNA detection, the clinical management of human diseases including therapeutic efficacy prediction could be revolutionized. ■

Eunice Zhang received her PhD in Pharmacology at the University of Liverpool working on the pharmacogenetics of warfarin. She is currently working as a postdoctoral research associate in Professor Sir Munir Pirmohamed's laboratory at the Wolfson Centre for Personalised Medicine at the University of Liverpool. Her main research areas are in the Pharmacogenomics of Oral Anticoagulants and Adverse Drug Reactions, with particular interests in the fields of genomics and transcriptomics and their roles in improving prescription choices and the development of diagnostic tests for personalized medicine. Email: eunzhang@liverpool.ac.uk

Glossary

Term	Definition
Agov	Argonaute, a family of proteins which binds to miRNA and is essential catalytic components of the miRNA-induced silencing complex
Canonical pathway	A well-established classical pathway
DGCR8	DiGeorge syndrome critical region 8 protein, a double-stranded RNA-specific ribonuclease
Dicer	A ribonuclease III enzyme which cleaves pre-miRNA into miRNA
Drosha	A ribonuclease III enzyme which initiates miRNA processing in the nuclease
GTP	Guanidine triphosphate
Microprocessor	A protein complex consisting of Drosha and a DGCR8
miRISC	miRNA-induced silencing complex which contains Dicer1 and Ago proteins
Non-canonical pathway	An alternative less generalized pathway
Pol II	RNA polymerase II
Poly(A)	A stretch of RNA that has only adenine bases
Pre-miRNA	Precursor miRNA
Pri-miRNA	Primary miRNA
Ran	Ras-related nuclear protein, a small GTP-binding nuclear protein involved in the translocation of RNA from the nuclease into the cytoplasm
Ribonuclease	A class of enzymes that catalyses the degradation of RNA into smaller components, commonly abbreviated RNase
RIII	Ribonuclease III
TRBP	Transactivation-responsive RNA-binding protein
XPO5	Exportin 5, a protein involved in transporting molecules between the cytoplasm and the nucleus of a eukaryotic cell
5′ cap	Five-prime cap is a specially altered nucleotide (a methylated guanosine) on the 5′ end of the transcript

References

1. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell **75**, 843–854
2. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. Cell **75**, 855–862
3. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. Science **294**, 853–858
4. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science **294**, 858–862
5. http://www.mirbase.org/
6. Vasudevan, S. (2012) Posttranscriptional upregulation by microRNAs. Wiley Interdiscip. Rev. RNA **3**, 311–330
7. Friedman, R.C., Farh, K.K., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. **19**, 92–105
8. Arroyo, J.D., Chevillet, J.R., Kroh, E.M. et al. (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc. Natl. Acad. Sci. U.S.A. **108**, 5003–5008
9. Vickers, K.C., Palmisano, B.T., Shoucri, B.M., Shamburek, R.D. and Remaley, A.T. (2011) MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat. Cell Biol. **13**, 423–433
10. Muralidharan-Chari, V., Clancy, J.W., Sedgwick, A. and D'Souza-Schorey C. (2010) Microvesicles: mediators of extracellular communication during cancer progression. J. Cell Sci. **123**, 1603–1611
11. Tsujiura, M., Ichikawa, D., Komatsu, S. et al. (2010) Circulating microRNAs in plasma of patients with gastric cancers. Br. J. Cancer **102**, 1174–1179