Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Genetics and Genomics 38 (2011) 505−513

JOURNAL OF GENETICS AND GENOMICS

www.jgenetgenomics.org

Small RNA transcriptome investigation based on next-generation sequencing technology

Linglin Zhou [a], Xueying Li [a], Qi Liu [a], Fangqing Zhao [b,*], Jinyu Wu [a,*]

[a] *Institute of Genomic Medicine, Wenzhou Medical College, Wenzhou 325035, China*
[b] *Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China*

Received 17 June 2011; revised 25 August 2011; accepted 25 August 2011

Abstract

Over the past decade, there has been a growing realization that studying the small RNA transcriptome is essential for understanding the complexity of transcriptional regulation. With an increased throughput and a reduced cost, next-generation sequencing technology has provided an unprecedented opportunity to measure the extent and complexity of small RNA transcriptome. Meanwhile, the large amount of obtained data and varied technology platforms have also posed multiple challenges for effective data analysis and mining. To provide some insight into the small RNA transcriptome investigation, this review describes the major small RNA classes, experimental methods to identify small RNAs, and available bioinformatics tools and databases.

Keywords: Next-generation sequencing; Small RNA transcriptome; Category; Bioinformatics

1. Introduction

In almost all organisms, ranging from bacteria to mammals, our knowledge of small RNAs has grown at a breathtaking pace (Ghildiyal and Zamore, 2009), especially in the last few years since the first small non-coding RNA was identified. With the identification of extensive and diverse populations of small RNAs, increasing efforts have been made to reveal their wide range of biological functions. As potential regulators of gene expression at both transcriptional and post-transcriptional levels, small RNAs are found to be involved in regulation of gene expression through translational repression, mRNA degradation and chromatin modification (Czech and Hannon, 2011). Being the currently best-characterized, most conserved and prominent non-coding small RNAs (Bartel, 2009), microRNAs (miRNAs) are to date over 10,000 in number and found in 115 species across invertebrates, vertebrates, plants, yeasts, and even protists and viruses (Liu et al., 2004a). lin-4 and let-7, originally identified by genetic analysis of *Caenorhabditis elegans* development timing, are the two founding members of miRNAs (Lee et al., 1993; Reinhart et al., 2000).

Due to an increased throughput and a reduced cost, next-generation sequencing (NGS) technology, such as Roche 454, ABI SOLiD and Illumina Solexa, represents an attractive alternative to traditional methods for a wide range of applications (Zhang et al., 2011). At present, NGS technology has been widely used in the fields of *de novo* genome sequencing, genome re-sequencing and transcriptome studies. In particular, deep sequencing has shown significant promise for small RNA discovery and genome-wide transcriptome profiling because of its ability to sequence many different small RNAs at single-base resolution, on a genome-wide scale and in a single instrument run. Therefore, NGS has been widely used to study small RNA transcriptome. For example, using the Solexa platform, the NK cell miRNA transcriptome has been investigated to study miRNA roles in NK cell biology, and 21 novel miRNA genes have been discovered (Kapranov et al., 2007). Moreover, using deep sequencing, Jima et al. (2010) have provided a comprehensive quantitative catalog of miRNA

* Corresponding authors. Tel/fax: +86 10 6488 0586 (F. Zhao); +86 577 8883 1309 (J. Wu).
E-mail addresses: zhfq@mail.biols.ac.cn (F. Zhao), iamwujy@yahoo.com.cn (J. Wu).

doi:10.1016/j.jgg.2011.08.006

expression by normal and malignant B cells, paving the way for characterization of the biogenesis and functions of expressed miRNA genes (Kim, 2005).

However, with the generation of billions of short sequencing reads, the infrastructure of existing bioinformatics technology has been greatly challenged in the field of data transfer, storage, read alignment, assembly and subsequent processing management despite that lots of user-friendly and efficient analysis software have been developed to fill the gaps. In this review, we firstly give an overview of the small RNA categories. Then, we introduce various experimental methods adopted to investigate the small RNAs, with focus on the application of NGS technologies in miRNAs. Finally, we discuss recently available small RNA transcriptome databases and bioinformatics tools for NGS platforms.

2. Classes of small RNAs

Small RNAs, also called nonconventional, functional or regulatory RNAs, are defined as any non-coding RNAs <200 nucleotides (Kapranov et al., 2007), which are extensively involved in regulation. For example, they induce gene silencing through base-pairing with specific targets, and collectively regulate a substantial fraction of the transcriptome (Czech and Hannon, 2011). Their important roles in transcription, RNA stability, and translation are becoming increasingly evident. On the whole, based on their biogenesis, biological functions and associated protein partners, small RNAs can be categorized into a few broad classes (Kim, 2005) (Table 1).

Derived from the stem-loop regions of longer RNA precursors, miRNAs, ~21−24 nt in size, influence mRNA stability and translation by binding to them (Bartel, 2004). miRNA-mediated translational repression represents a distinct mechanism of post-transcriptional regulation, and modulates more than one-third of the human gene transcriptomes according to bioinformatics analysis (Lewis et al., 2005). To date, most of them have been identified to be involved in vital biological processes, such as development, apoptosis, neural patterning, and viral infection. Interestingly, miRNAs found in plants are different from those in animals in terms of gene structures, biogenesis and mechanisms as well as target properties, indicating that miRNAs in animals and plants may have independent origins (Jones-Rhoades et al., 2006; Molnar et al., 2007; Zhao et al., 2007). Deep sequencing has contributed considerably to the discovery of many expressed small RNAs and the identification of a large number of candidate miRNAs. Over the past 5 years, more than 8000 miRNA genes have been identified in animals and plants (miRBase release version 16), and the number is still growing (Witten et al., 2010). An up-to-date repository of miRNAs and miRNA genes from various organisms is publicly available in the miRNA registry (miRBase Sequence Database) (http://www.mirbase.org/).

Short-interfering RNAs (siRNAs), another major class of endogenous non-coding small RNAs, are shown to be important regulators of gene expression in eukaryotes (Vazquez, 2006). siRNAs are usually derived from long double-stranded RNA (dsRNA) processed by RNase III-like enzymes called DICERs and mainly involved in defense against molecular parasites including viruses, transposons, and transgenes through RNAi (Sijen and Plasterk, 2003). In general, siRNAs trigger mRNA degradation through perfect-matching to their targets (Meister and Tuschl, 2004). siRNAs are subdivided into tasi-RNA, ra-siRNA, nat-siRNAs, scnRNA and hc-siRNA (Table 1).

According to the report of Aravin et al. (2001), a distinct class of small RNAs, named PIWI-interacting RNAs (piRNAs), typically 24−31 nt in size, has been proved to be present in *Drosophila melanogaster*. Since then, piRNAs are identified and preliminarily characterized as a novel class of small RNAs. Recent reports indicated that piRNAs specifically bind to the Piwi clade of Argonaute proteins (germline-specific members of the AGO protein family) to form the RNA-induced silencing complex (RISC), which silences transposable elements (TEs), thus protecting genomic integrity in the germ line (Li et al., 2009). Although much progress has been made in the investigation of their distinct biogenesis mechanisms (Girard et al., 2006; Aravin et al., 2007) and molecular functions (Siomi et al., 2011), the piRNA pathway remains poorly understood and needs to be further explored.

Recently, researchers have made a groundbreaking discovery of a new class of bidirectionally transcribed, non-polyadenylated non-coding RNA − enhancer RNA or eRNA (Kim et al., 2010). However, their biological functions remain unclear and need further investigation.

Discoveries will continue to exponentially increase the knowledge of new small RNA classes and new examples of existing classes, and thus our understanding of gene regulation will be expanded significantly as well.

3. Experimental methods to identify small RNAs

3.1. Conventional experimental approaches

Various experimental strategies including cloning, Northern blotting, RNase protection assay and primer extension methodologies have been employed for quantification and identification of novel small RNAs. A new miRNA gene prediction method combined with small-RNA cloning was used to study a subset of pathogenic viruses, and the family of herpes viruses was demonstrated most likely to encode miRNAs (Pfeffer et al., 2005). Many small RNAs with the potential to arise from fold-back structures characteristic of the lin-4 and let-7 hairpins were identified by cloning and sequencing (Lagos-Quintana et al., 2001). Furthermore, by using pyrosequencing technology, Ruby et al. (2006) identified thousands of endogenous small RNAs in *C. elegans*, including miRNAs, siRNAs, and 21U-RNAs. Although cloning and sequencing is a useful approach for the identification of individual novel small RNAs, there are several shortcomings in its application. For example, it requires a large amount of total RNA, but it is not practical in many cases. Due to its low coverage, small RNAs of low abundance may very often be missed. Additionally, it is difficult to use this method to distinguish miRNAs from degradation products of other ncRNAs, such as

Table 1
The classification of small RNAs.

Class	Sub-class	Length (nt)	Biogenesis	Protein partner (associated Argonaute superfamily proteins)	Action mechanism	Biological function	Reference
MicroRNA (miRNA)		~22 (19–25)	Two-step cleavage of hairpin precursors by Drosha and Dicer	Argonaute subfamily	Post-transcriptional regulation of transcripts	Diverse functions such as development and cell differentiation	Bartel, 2004; Meister and Tuschl, 2004
Short-interfering RNA (siRNA)	Endogenous trans-acting siRNA (tasi-RNA)	21–22 (nematode) and 21 (plants)	Cleavage of long endogenous dsRNAs precursors by Dicer	Ago1*	Post-transcriptional regulation of transcripts (mRNA cleavage)	Related to plant development	Peragine et al., 2004; Vazquez et al., 2004; Baumberger and Baulcombe, 2005; Zhao et al., 2007
	Repeat-associated siRNA (ra-siRNA)	24–26 (plants) and 24–27 (fruit flies)	Cleavage of long dsRNAs derived from genomic repeats and retrotransposons by Dicer	Ziwi	Modification of histone and/or DNA	Silence of transposable element and viruses	Ketting et al., 1999; Tabara et al., 1999; Djikeng et al., 2001; Mette et al., 2002; Schramke and Allshire, 2003; Xie et al., 2004;
	Natural-antisense transcript derived siRNAs (nat-siRNAs)		Dicer processing of dsRNA arising from sense- and antisense-transcript pairs		Post-transcriptional regulation	Pathogen defense and regulation of stress responses in plants	Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Vazquez, 2006
	Small scan RNA (scnRNA)	27–30	Cleavage of long dsRNAs by Dicer	Twi1	Guide heterochromatin formation and DNA elimination	Genome rearrangement during conjugation	Mochizuki et al., 2002; Liu et al., 2004b; Mochizuki and Gorovsky, 2004
	Heterochromatic siRNA (hc-siRNA)	~24	Derived from long dsRNA precursors	Ago1*	DNA and histone methylation	Transposon control and regulation of heterochromatin structures	Djikeng et al., 2001; Shi et al., 2004a, 2004b
Tiny non-coding RNA (tncRNA)		~18	Produced by Dicer from specific position near the 5′ end of transcribed genes			Unknown	Ambros et al., 2003; Taft et al., 2009
Small modulatory RNA (smRNA)		~20			Transcriptional transactivation	Regulate neuron specific gene expression and neuronal differentiation	Kuwabara et al., 2004
Endogenous siRNA (endo-siRNA)		~21	Produced from endogenous dsRNA precursors by Dicer-2		Post-transcriptional regulation	Silence of transposable element and possibly endogenous mRNA	Tabara et al., 1999; Okamura and Lai, 2008; Watanabe et al., 2008; Fagegaltier et al., 2009
Piwi-associated RNA (piRNA)		24–31	Produced from long ssRNA precursors in gonads, which is Argonaute-dependent but Dicer-independent	Piwi, Aubergine (Aub) and Ago3* HILI, HIWI1, HIWI2 and HIWI3	Post-transcriptional and chromatin regulation	Transposons and retroelements silence in the germline	Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Grimson et al., 2008; Klattenhoff and Theurkauf, 2008

(*continued on next page*)

Table 1 (*continued*)

Class	Sub-class	Length (nt)	Biogenesis	Protein partner (associated Argonaute superfamily proteins)	Action mechanism	Biological function	Reference
Small nuclear RNA (snRNA)					Post-transcriptional modification	Involved in mRNA splicing	Kiss, 2001
Small nucleolar RNA (snoRNA)	C/D box snoRNAs	~70−120				Directs the enzymatic modification of ribosomal RNAs	Taft et al., 2009
	H/ACA box snoRNAs	~100−200				Guide pseudouridulation	Taft et al., 2010
qiRNAs		20−21	Originate mostly from the ribosomal DNA locus	Argonaute protein QDE-2	Post-transcriptional regulation	Play a role in the DNA damage response by inhibiting protein translation	Lee et al., 2009
Enhancer RNAs (eRNAs)						Unknown	Kim et al., 2010

rRNA or tRNA. To circumvent these limitations, many other methods for small RNA transcriptome analysis have been developed. Various studies have showed that miRNAs are efficiently detected by Northern blotting analysis (Valoczi et al., 2004). Using Northern blotting analysis, a subset of brain-expressed miRNAs were characterized, indicating their potential roles in murine and human neuronal differentiation (Pfeffer et al., 2005). RNase protection assay was used to detect mature miRNA (Lee et al., 2003). Previous research proposed a new real-time RT-PCR scheme (Stem-loop RT-PCR) for miRNA quantification (Kramer, 2011), which enables fast, accurate and sensitive detection and expression profiling of miRNA and other small RNAs.

3.2. Microarray

Microarray technology is a further step toward high throughput quantification of miRNA expression, and has been successfully developed to explore various miRNA expression patterns during development (Rosa and Brivanlou, 2009), differentiation (Asli et al., 2008), oncogenesis (Croce, 2009), and disease progression (Latronico and Condorelli, 2009). Moreover, this technology has also been developed for the identification of novel miRNAs predicted by bioinformatics approaches (Bentwich et al., 2005; Li et al., 2006). In a recent study, a combined approach of computational prediction and microarray analysis was used to identify and quantify a baboon's miRNAs (Karere et al., 2010). Using custom-designed dual-channel microarrays, researchers also conducted miRNA profiling studies of staged embryos to analyze microRNA expression levels (Thomson et al., 2004). In addition, microarray analysis was used to determine the expression patterns of a subset of miRNAs in the brain (Krichevsky et al., 2003) and many other human tissues (Barad et al., 2004). Therefore, microarrays can be used to study small RNA expression in addition to mRNA expression. A good case in point is miRNA microchip, which is specific for miRNA profiling on a global scale and serves as a robust tool not only for detection of miRNA expression patterns but also for expression quantification (Liu et al., 2008). In general, miRNA-specific oligonucleotide microarray system has several advantages in comparison with the currently available methods for detection and characterization of miRNAs. Firstly, the global expression of multiple genes can be identified in the same sample at one time point. Secondly, the expression of both mature and precursor molecules can be identified simultaneously using carefully synthesized oligonucleotide probes. Thirdly, in comparison to Northern blotting analysis, less amount of RNA is needed and more reproducible results can be obtained. Finally, lower false-negative and false-positive rates can be achieved.

3.3. Next-generation sequencing

Although cloning and sequencing of small RNAs enables the discovery of new miRNAs, it was time-consuming and limited to the most abundant sRNAs. Real-time PCR enables

a rapid detection of miRNAs and precursors but has limitations on novel miRNA detection. RNA hybridization has low throughput and limited sensitivity. miRNA arrays are not suitable for detecting novel miRNAs and have radioactive isotopes in probe design. Moreover, all these methods have major limitations on comparability and accuracy of the derived data (Cummins and Velculescu, 2006).

Traditional methods mentioned above are not generally useful for discovering and profiling low-abundance or novel small RNAs. In contrast, deep sequencing approaches, such as Illumina Solexa, ABI SOLiD and Roche 454, have several advantages, including high resolution, dramatically increased throughput and sequencing depth, and reduced complexity of experimental procedures. With the availability of high-throughput sequencing technology, new discoveries of species- or tissue-specific or low-abundant small RNAs have become possible.

Since there are some differences among these novel sequencing technologies, proper sequencing method should be selected based on different research goals (Metzker, 2010). For example, 454 technology offers the longest, high quality reads (with read lengths of 400 bp, 500 Mb/run), whereas the Solexa platform yields a much larger amount of data (with read lengths of 35−75 bp, 12 Gb/run) (McCormick et al., 2011). Therefore, for sequencing of shorter reads (up to 35 bp), the Solexa platform will be a better choice (Szittya et al., 2008). Unlike platforms that produce longer read lengths, SOLiD can currently give an enormous volume of reads with sequencing depth of 300 Gb/run and produce maximum read lengths of 75 bp (McCormick et al., 2011). Given that commercially available NGS platforms coexist in the marketplace with various sequencing capabilities in terms of throughput, read length, error rate, etc, a preferable platform is often selected according to several factors, e.g., cost, sequencing depth and read length. For example, considering that sRNA is in the range of approximately 18−30 nt and that low sequencing depth may fail to detect rare species, Illumina or SOLiD is currently the ideal choice for small RNA transcriptome investigations.

These revolutionized technologies have resulted in a massive expansion of transcriptomics in all fields. Particularly, for species whose whole genomic information is unavailable, deep sequencing method shows remarkable superiority over traditional approaches. We envision that such approaches will gain even more popularity in the near future. However, for the lack of efficient and flexible tools, how to handle and analyze the data yielded by deep sequencing is the computational and analytical challenge now facing us.

4. Currently available bioinformatics databases and tools

4.1. Databases

deepBase (Yang et al., 2010), a comprehensive database developed for annotation and discovery of small RNAs from transcriptomic data, for one thing, provides a comprehensive annotation and mining of deep sequencing data from 185 small RNA libraries covering diverse tissues and cell lines of seven organisms. Another two databases related to deep sequencing small RNA datasets are Gene Expression Omnibus (GEO), which is a web-based analysis tool established for the efficient large-scale functional genomic data retrieval and mining (Barrett et al., 2007), and FANTOM4 EdgeExpressDB, which is a novel database and set of interfaces for interpreting biological networks and comparing large high-throughput expression datasets that requires minimal development for new data types and search patterns (Severin et al., 2009). Using intuitive gene-centric and sub-network views, the FANTOM4 EdgeExpressDB summarizes gene expression patterns in the contexts of alternative promoter structures and regulatory transcription factors and miRNAs, and is a crucial resource for elucidation of the gene regulation mechanisms in acute myeloid leukemia, monocyte/macrophage differentiation as well as human transcriptional networks.

With a large number of miRNAs data being generated and miRNA target genes being discovered, many public resources have been constructed. miRBase (Griffiths-Jones et al., 2008), a database for miRNA research, is a central online repository focusing on miRNA nomenclature, annotation and target prediction. TarBase (Sethupathy et al., 2006) is developed for a comprehensive collection and description of experiment-supported miRNA targets. miR2Disease (Jiang et al., 2009) is designed for the study of miRNA-disease relationships in various human diseases. PMRD (Zhang et al., 2010), which integrates the latest plant miRNA data with a user-friendly web interface, is a useful tool for studying miRNAs and their target genes of model plants and major crops.

4.2. Bioinformatics analysis pipeline

Next-generation sequencing platforms pose increasing challenges for the efficiency, accuracy and usability of data analysis software. Now more than ever, intuitive and efficient data exploration and bioinformatics analysis tools are urgently required. In an effort to relieve growing pressure on handling massive quantities of short-read sequences, several NGS-based small RNA transcriptome bioinformatics analysis tools, together with useful databases of small RNAs, have been developed (Table 2).

MicroRazerS (Emde et al., 2010) is a stand-alone software optimized for rapid alignment of small RNA reads onto a reference genome, which is a basic step for further identification and quantification of miRNAs obtained from deep sequencing. CASHX pipeline (Fahlgren et al., 2009) is used to process read data generated from Illumina and other high-throughput sequencing technologies. The pipeline can be used to parse, map, quantify and manage large quantities of sequence data. Several computational methods were developed to detect miRNAs. Publicly available tools for miRNA discovery include miRDeep, miRTRAP, MIREAP, mirTools, miRanalyzer, MIReNA, miRExpress, UEA sRNA toolkit, SeqBuster and miRNAkey. miRDeep (Friedlander et al., 2008), to our knowledge the first stand-alone tool used for deep sequencing miRNA data, was developed to analyze

Table 2
NGS-based small RNA transcriptome bioinformatics databases and analysis tools.

Name	Link	Function	Service type	Reference
MicroRazerS	http://www.seqan.de/projects/MicroRazerS.html	Small RNA reads alignment	Stand-alone	Emde et al., 2010
CASHX	http://jcclab.science.oregonstate.edu/?q=node/view/54596	Small RNA reads alignment Quantify and manage sequence data	Stand-alone	Fahlgren et al., 2009
mirDeep	http://www.mdc-berlin.de/rajewsky/miRDeep.html	Detect known and novel miRNAs	Stand-alone	Friedlander et al., 2008
miRTRAP	http://flybuzz.berkeley.edu/miRTRAP.html	Detect novel miRNAs	Stand-alone	Hendrix et al., 2010
MIREAP	https://sourceforge.net/projects/mireap/	Detect known and novel miRNAs	Stand-alone	
mirTools	http://centre.bioinformatics.zj.cn/mirtools/	Small RNA read alignment Comparative analysis of two or more miRNA expression data Classification and annotation of known miRNAs Detect novel miRNAs	Web server	Zhu et al., 2010
miRanalyzer	http://bioinfo2.ugr.es/miRanalyzer/miRanalyzer.php	Small RNA read alignment Detect known and unknown miRNAs Detect undetected mature-star miRNAs	Web server Stand-alone	Hackenberg et al., 2011
UEA sRNA toolkit	http://srna-tools.cmp.uea.ac.uk/	Predict miRNAs and their targets Compare expression levels in sRNA loci Find transactin siRNA loci	Web server	Moxon et al., 2008
SeqBuster	http://estivill_lab.crg.es/seqbuster	miRNA isoforms analysis	Web server	Pantano et al., 2010
miRNAkey	http://ibis.tau.ac.il/miRNAkey	Small RNA read alignment Comparative analysis of miRNA expression data	Stand-alone	Ronen et al., 2010
MIReNA	http://www.ihes.fr/∼carbone/data8/	Detect novel miRNAs pre-miRNAs validation	Stand-alone	
miRExpress	http://mirexpress.mbc.nctu.edu.tw/index.php	miRNA expression profiling	Stand-alone	Wang et al., 2009
deepBase	http://deepbase.sysu.edu.cn/	Small RNA reads alignment Small RNAs transcriptome annotation Data storage, retrieval, mining and visualization	Database	Yang et al., 2010
GEO short-read archive	http://www.ncbi.nlm.nih.gov/geo/	Capture, storage and retrieval of large-scale genomic data	Database	Barrett et al., 2007
FANTOM4 EdgeExpressDB	http://fantom.gsc.riken.jp/4/edgeexpress	Gene express patterns of promoters, genes and microRNAs	Database	Severin et al., 2009
miRBase	http://www.mirbase.org	Search for miRNA Analyze genomic coordinates and context Mine relationships between miRNAs	Database	Griffiths-Jones et al., 2008
TarBase	http://diana.cslab.ece.ntua.gr/tarbase/	Collection of experimentally tested miRNA targets information	Database	Sethupathy et al., 2006
miR2Disease	http://www.mir2disease.org/	Collection of microRNA-disease relationship information Data submission or retrieval	Database	Jiang et al., 2009
PMRD	http://bioinformatics.cau.edu.cn/PMRD/	Integrations of plant miRNA data	Database	Zhang et al., 2010

large-scale small RNA sequencing data and detect both known and novel miRNAs with high confidence, and employs stringent statistical controls to estimate the false-positive rate and the sensitivity of predictions to ensure its reliability. MiRTRAP (Hendrix et al., 2010) is a novel computational strategy developed for comprehensive identification of the full repertoire of miRNAs (known and novel miRNAs) from deep sequencing. In brief, MiRTRAP has powerful strength to systematically identify all possible miRNAs from Illumina Solexa sequencing data. MIREAP (https://sourceforge.net/projects/mireap/) is a tool for identifying both known and novel miRNAs from small RNA libraries generated by Illumina/454/SOLiD technology. mirTools (Zhu et al., 2010), a comprehensive web server developed to characterize multiple small RNA transcriptomes in 15 most commonly studied model organisms, has several advantages such as high resolution, high yield and reduced complexity of experimental procedures. One feature of mirTools is that it runs in two-mode architecture on a centralized platform supporting miRNA identification not only in single but between paired samples. miRanalyzer (Hackenberg et al., 2011), a web server and stand-alone tool based on a highly accurate machine-learning approach, enables detection of all known miRNA sequences annotated in miRBase and prediction of novel miRNAs. Additionally, the new version of miRanalyzer has been developed with differential expression module, allowing comparative analysis of differential expression of known and predicted microRNAs. MIReNA (http://www.ihes.fr/∼carbone/data8/) is a genome-wide search algorithm for the detection of new miRNAs starting from known ones or from deep sequencing data, and for pre-miRNAs validation, with high sensitivity, specificity and accuracy. Compared with miRDeep, MIReNA shows high specific predictive power for miRNAs. miRExpress (Wang et al., 2009) is a stand-alone software

package initially designed for detecting miRNA expression profile from HTS reads of small-RNAs without the need for sequenced genomes. UEA sRNA toolkits (Moxon et al., 2008) were designed to analyze a large number of short sequences, such as the identification of miRNAs and their targets and the detection of differential expression levels of specific small RNA loci. SeqBuster (Pantano et al., 2010) is a web-based tool for handling and analyzing the miRNA variants or isomers hidden in large-scale small RNA datasets. miRNAkey (Ronen et al., 2010) is a software package with user-friendly graphical interface that was developed to analyze miRNA deep sequencing data, including comparative analysis of differentially expressed miRNAs in paired samples.

4.3. Limitations of current approaches

The generation of expression profiles and identification of novel small RNAs from deep sequencing are dependent on tools. The bioinformatics tools introduced above can be used for i) data mapping, annotation and visualization; ii) data storage and retrieval; iii) integration and interpretation of data from multiple technological platforms, tissues and cell lines; and iv) custom analyses that help address a variety of biological questions. However, these tools still have some limitations on, for example, the identification of novel miRNAs, piRNAs, etc. Due to a large number of variable sequences and the nuclear localization of mammalian piRNAs, currently available tools are not effective in identifying novel piRNAs from tremendously abundant RNA species. Future studies should place emphasis upon identification of small RNAs which were previously thought to be mere degradation products. There are great challenges in the development of both more efficient automated pipelines for deciphering deep sequencing data and validation pipelines for checking the results. Additionally, there's an impending need for the analysis of large-scale data generated under a user-friendly environment in which large-scale short reads can be displayed from next-generation sequencing. Several bioinformatics tools have been specifically designed for visualizing next-generation sequencing data. For example, MapView has been designed for general short reads alignment (Bao et al., 2009), and MagicViewer for SNP annotation and visualization (Hou et al., 2010). However, no ideal next-generation sequencing data visualization tools have been presently developed for the small RNA transcriptome. Most notably, because bioinformatics tools, both web-based and stand-alone, differ somewhat in mapping and counting strategies, expression levels of miRNA estimated by different tools may vary to some degree. For example, using the same dataset (http://ibis.tau.ac.il/miRNAkey/downloads/sample_data_small.tar.gz), a pair-wise comparison analysis between miRNAkey, UEA sRNA toolkits and miRExpress showed that there is some discrepancy between these tools in the estimated expression levels of miRNA (Supplementary Fig. 1). Since each tool bears inherent biases, to parse and compare sequencing data across platforms, further estimation and correction of systematic errors should be made.

5. Conclusions and perspectives

In summary, high-throughput sequencing, on the one hand, provides unprecedented opportunities to generate a large amount of sequencing data and boosts studies on the discovery and quantification of small RNAs; on the other hand, it poses challenges for subsequent data manipulation and interpretation. The development of tools for computational analysis of large amount of data generated from NGS is still in its infancy. Full-scale efforts are being made to improve our abilities to handle and analyze sequencing data. As miRNA research is an exciting and growing field, many open questions are yet to be answered and future research directions yet to be discussed. To solve these problems, the integration is required of genomics, genetics, molecular and structural biology, biochemistry and bioinformatics. With all the achievements described above, we have made a step further toward a comprehensive knowledge of the small RNAs.

Acknowledgements

This work was supported by the National High Technology Research and Development Program of China (No. 2006AA02A304) and Major State Basic Research Development Program of China (No. 2007CB512302).

Supplementary data

Supplementary data associated with this article can be found online at doi:10.1016/j.jgg.2011.08.006.

References

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., Jewell, D., 2003. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr. Biol. 13, 807−818.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., Gvozdev, V.A., 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. Curr. Biol. 11, 1017−1027.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., Hannon, G.J., 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316, 744−747.

Asli, N.S., Pitulescu, M.E., Kessel, M., 2008. MicroRNAs in organogenesis and disease. Curr. Mol. Med. 8, 698−710.

Bao, H., Guo, H., Wang, J., Zhou, R., Lu, X., Shi, S., 2009. MapView: visualization of short reads alignment on a desktop computer. Bioinformatics 25, 1554−1555.

Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y., Lobenhofer, E.K., Sharon, E., Shiboleth, Y.M., Shtutman, M., Bentwich, Z., Einat, P., 2004. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res. 14, 2486−2494.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., Edgar, R., 2007. NCBI GEO: mining tens of millions of expression profiles − database and tools update. Nucleic Acids Res. 35, D760−D765.

Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281−297.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215−233.

Baumberger, N., Baulcombe, D.C., 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA 102, 11928−11933.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y., Bentwich, Z., 2005. Identification of hundreds of conserved and non-conserved human microRNAs. Nat. Genet. 37, 766−770.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., Zhu, J.K., 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell 123, 1279−1291.

Croce, C.M., 2009. Causes and consequences of microRNA dysregulation in cancer. Nat. Rev. Genet. 10, 704−714.

Cummins, J.M., Velculescu, V.E., 2006. Implications of micro-RNA profiling for cancer diagnosis. Oncogene 25, 6220−6227.

Czech, B., Hannon, G.J., 2011. Small RNA sorting: matchmaking for Argonautes. Nat. Rev. Genet. 12, 19−31.

Djikeng, A., Shi, H., Tschudi, C., Ullu, E., 2001. RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24−26-nucleotide RNAs. RNA 7, 1522−1530.

Emde, A.K., Grunert, M., Weese, D., Reinert, K., Sperling, S.R., 2010. MicroRazerS: rapid alignment of small RNA reads. Bioinformatics 26, 123−124.

Fagegaltier, D., Bouge, A.L., Berry, B., Poisot, E., Sismeiro, O., Coppee, J.Y., Theodore, L., Voinnet, O., Antoniewski, C., 2009. The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*. Proc. Natl. Acad. Sci. USA 106, 21258−21263.

Fahlgren, N., Sullivan, C.M., Kasschau, K.D., Chapman, E.J., Cumbie, J.S., Montgomery, T.A., Gilbert, S.D., Dasenko, M., Backman, T.W., Givan, S.A., Carrington, J.C., 2009. Computational and analytical framework for small RNA profiling by high-throughput sequencing. RNA 15, 992−1002.

Friedlander, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., Rajewsky, N., 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat. Biotechnol. 26, 407−415.

Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. Nat. Rev. Genet. 10, 94−108.

Girard, A., Sachidanandam, R., Hannon, G.J., Carmell, M.A., 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199−202.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154−D158.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., Bartel, D.P., 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature 455, 1193−1197.

Grivna, S.T., Pyhtila, B., Lin, H., 2006. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc. Natl. Acad. Sci. USA 103, 13415−13420.

Hackenberg, M., Rodriguez-Ezpeleta, N., Aransay, A.M., 2011. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res. 39, W132−W138.

Hendrix, D., Levine, M., Shi, W., 2010. miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol. 11, R39.

Hou, H., Zhao, F., Zhou, L., Zhu, E., Teng, H., Li, X., Bao, Q., Wu, J., Sun, Z., 2010. MagicViewer: integrated solution for next-generation sequencing data visualization and genetic variation detection and annotation. Nucleic Acids Res. 38, W732−W736.

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., Liu, Y., 2009. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 37, D98−D104.

Jima, D.D., Zhang, J., Jacobs, C., Richards, K.L., Dunphy, C.H., Choi, W.W., Yan Au, W., Srivastava, G., Czader, M.B., Rizzieri, D.A., Lagoo, A.S., Lugar, P.L., Mann, K.P., Flowers, C.R., Bernal-Mizrachi, L., Naresh, K.N., Evens, A.M., Gordon, L.I., Luftig, M., Friedman, D.R., Weinberg, J.B., Thompson, M.A., Gill, J.I., Liu, Q., How, T., Grubor, V., Gao, Y., Patel, A., Wu, H., Zhu, J., Blobe, G.C., Lipsky, P.E., Chadburn, A., Dave, S.S., 2010. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood 116, e118−e127.

Jones-Rhoades, M.W., Bartel, D.P., Bartel, B., 2006. MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. 57, 19−53.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R., 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484−1488.

Karere, G.M., Glenn, J.P., VandeBerg, J.L., Cox, L.A., 2010. Identification of baboon microRNAs expressed in liver and lymphocytes. J. Biomed. Sci. 17, 54.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas Jr., A., Zhu, J.K., Staskawicz, B.J., Jin, H., 2006. A pathogen-inducible endogenous siRNA in plant immunity. Proc. Natl. Acad. Sci. USA 103, 18002−18007.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., Plasterk, R.H., 1999. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell 99, 133−141.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G., Greenberg, M.E., 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182−187.

Kim, V.N., 2005. Small RNAs: classification, biogenesis, and function. Mol. Cells 19, 1−15.

Kiss, T., 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J. 20, 3617−3622.

Klattenhoff, C., Theurkauf, W., 2008. Biogenesis and germline functions of piRNAs. Development 135, 3−9.

Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K., Kosik, K.S., 2003. A microRNA array reveals extensive regulation of microRNAs during brain development. RNA 9, 1274−1281.

Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., Gage, F.H., 2004. A small modulatory dsRNA specifies the fate of adult neural stem cells. Cell 116, 779−793.

Kramer, M.F., Jul 2011. Stem-Loop RT-qPCR for miRNAs. Curr. Protoc. Mol. Biol. Chapter 15: Unit15.10.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294, 853−858.

Latronico, M.V., Condorelli, G., 2009. MicroRNAs and cardiac pathology. Nat. Rev. Cardiol. 6, 419−429.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., Kingston, R.E., 2006. Characterization of the piRNA complex from rat testes. Science 313, 363−367.

Lee, H.C., Chang, S.S., Choudhary, S., Aalto, A.P., Maiti, M., Bamford, D.H., Liu, Y., 2009. qiRNA is a new type of small interfering RNA induced by DNA damage. Nature 459, 274−277.

Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843−854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., Kim, V.N., 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415−419.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15−20.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., Kittler, E.L., Zapp, M.L., Klattenhoff, C., Schulz, N., Theurkauf, W.E., Weng, Z., Zamore, P.D., 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 137, 509−521.

Li, S.C., Pan, C.Y., Lin, W.C., 2006. Bioinformatic discovery of microRNA precursors from human ESTs and introns. BMC Genomics 7, 164.

Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M., Alder, H., Bullrich, F.,

Negrini, M., Croce, C.M., 2004a. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc. Natl. Acad. Sci. USA 101, 9740−9744.

Liu, C.G., Calin, G.A., Volinia, S., Croce, C.M., 2008. MicroRNA expression profiling using microarrays. Nat. Protoc. 3, 563−578.

Liu, Y., Mochizuki, K., Gorovsky, M.A., 2004b. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. Proc. Natl. Acad. Sci. USA 101, 1679−1684.

McCormick, K.P., Willmann, M.R., Meyers, B.C., 2011. Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. Silence 2, 2.

Meister, G., Tuschl, T., 2004. Mechanisms of gene silencing by double-stranded RNA. Nature 431, 343−349.

Mette, M.F., van der Winden, J., Matzke, M., Matzke, A.J., 2002. Short RNAs can identify new candidate transposable element families in *Arabidopsis*. Plant Physiol. 130, 6−9.

Metzker, M.L., 2010. Sequencing technologies − the next generation. Nat. Rev. Genet. 11, 31−46.

Mochizuki, K., Fine, N.A., Fujisawa, T., Gorovsky, M.A., 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell 110, 689−699.

Mochizuki, K., Gorovsky, M.A., 2004. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. Genes Dev. 18, 2068−2073.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., Baulcombe, D.C., 2007. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. Nature 447, 1126−1129.

Moxon, S., Schwach, F., Dalmay, T., Maclean, D., Studholme, D.J., Moulton, V., 2008. A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics 24, 2252−2253.

Okamura, K., Lai, E.C., 2008. Endogenous small interfering RNAs in animals. Nat. Rev. Mol. Cell Biol. 9, 673−678.

Pantano, L., Estivill, X., Marti, E., 2010. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res. 38, e34.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., Poethig, R.S., 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. Genes Dev. 18, 2368−2379.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., Russo, J.J., Ju, J., Randall, G., Lindenbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M., Tuschl, T., 2005. Identification of microRNAs of the herpesvirus family. Nat. Methods 2, 269−276.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403, 901−906.

Ronen, R., Gan, I., Modai, S., Sukacheov, A., Dror, G., Halperin, E., Shomron, N., 2010. miRNAkey: a software for microRNA deep sequencing analysis. Bioinformatics 26, 2615−2616.

Rosa, A., Brivanlou, A.H., 2009. microRNAs in early vertebrate development. Cell Cycle 8, 3513−3520.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., Bartel, D.P., 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell 127, 1193−1207.

Schramke, V., Allshire, R., 2003. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science 301, 1069−1074.

Sethupathy, P., Corda, B., Hatzigeorgiou, A.G., 2006. TarBase: a comprehensive database of experimentally supported animal microRNA targets. RNA 12, 192−197.

Severin, J., Waterhouse, A.M., Kawaji, H., Lassmann, T., van Nimwegen, E., Balwierz, P.J., de Hoon, M.J., Hume, D.A., Carninci, P., Hayashizaki, Y., Suzuki, H., Daub, C.O., Forrest, A.R., 2009. FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions. Genome Biol. 10, R39.

Shi, H., Djikeng, A., Tschudi, C., Ullu, E., 2004a. Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. Mol. Cell. Biol. 24, 420−427.

Shi, H., Ullu, E., Tschudi, C., 2004b. Function of the Trypanosome Argonaute 1 protein in RNA interference requires the N-terminal RGG domain and arginine 735 in the Piwi domain. J. Biol. Chem. 279, 49889−49893.

Sijen, T., Plasterk, R.H., 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature 426, 310−314.

Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A., 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. 12, 246−258.

Szittya, G., Moxon, S., Santos, D.M., Jing, R., Fevereiro, M.P., Moulton, V., Dalmay, T., 2008. High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. BMC Genomics 9, 593.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., Mello, C.C., 1999. The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. Cell 99, 123−132.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., Irvine, K., Arakawa, T., Nakamura, M., Kubosaki, A., Hayashida, K., Kawazu, C., Murata, M., Nishiyori, H., Fukuda, S., Kawai, J., Daub, C.O., Hume, D.A., Suzuki, H., Orlando, V., Carninci, P., Hayashizaki, Y., Mattick, J.S., 2009. Tiny RNAs associated with transcription start sites in animals. Nat. Genet. 41, 572−578.

Taft, R.J., Simons, C., Nahkuri, S., Oey, H., Korbie, D.J., Mercer, T.R., Holst, J., Ritchie, W., Wong, J.J., Rasko, J.E., Rokhsar, D.S., Degnan, B.M., Mattick, J.S., 2010. Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. Nat. Struct. Mol. Biol. 17, 1030−1034.

Thomson, J.M., Parker, J., Perou, C.M., Hammond, S.M., 2004. A custom microarray platform for analysis of microRNA gene expression. Nat. Methods 1, 47−53.

Valoczi, A., Hornyik, C., Varga, N., Burgyan, J., Kauppinen, S., Havelda, Z., 2004. Sensitive and specific detection of microRNAs by Northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res. 32, e175.

Vazquez, F., 2006. *Arabidopsis* endogenous small RNAs: highways and byways. Trends Plant Sci. 11, 460−468.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., Crete, P., 2004. Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. Mol. Cell 16, 69−79.

Wang, W.C., Lin, F.M., Chang, W.C., Lin, K.Y., Huang, H.D., Lin, N.S., 2009. miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. BMC Bioinformatics 10, 328.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y., Sasaki, H., 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453, 539−543.

Witten, D., Tibshirani, R., Gu, S.G., Fire, A., Lui, W.O., 2010. Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. BMC Biol. 8, 58.

Xie, Z., Romano, D.M., Kovacs, D.M., Tanzi, R.E., 2004. Effects of RNA interference-mediated silencing of gamma-secretase complex components on cell sensitivity to caspase-3 activation. J. Biol. Chem. 279, 34130−34137.

Yang, J.H., Shao, P., Zhou, H., Chen, Y.Q., Qu, L.H., 2010. deepBase: a database for deeply annotating and mining deep sequencing data. Nucleic Acids Res. 38, D123−D130.

Zhang, J., Chiodini, R., Badr, A., Zhang, G., 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics 38, 95−109.

Zhang, Z., Yu, J., Li, D., Liu, F., Zhou, X., Wang, T., Ling, Y., Su, Z., 2010. PMRD: plant microRNA database. Nucleic Acids Res. 38, D806−D813.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.J., Qi, Y., 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Genes Dev. 21, 1190−1203.

Zhu, E., Zhao, F., Xu, G., Hou, H., Zhou, L., Li, X., Sun, Z., Wu, J., 2010. mirTools: microRNA profiling and discovery based on high-throughput sequencing. Nucleic Acids Res. 38, W392−W397.