Max Delbrück Center for Molecular Medicine (MDC)

Global profiling of miRNA and the hairpin precursor: insights into miRNA processing and novel miRNA discovery

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)

eingereicht am Fachbereich Biologie-Chemie-Pharmazie
der Freien Universität Berlin

vorgelegt von

Na Li 李娜

aus Shanxi, V. R. China

Berlin
2011

1. Gutachter: Prof. Dr. Constance Scharff

2. Gutachter: Prof. Dr. Dr. Ralf Einspanier

Disputation am: 21.11.2011

ACKNOWLEDGEMENT

I wish to thank....

...My supervisor Dr. Wei Chen for inviting me to join his lab, for being an excellent mentor and teacher, for all his supervision, discussion and support on my PhD program during the past four years, as well as for his correction of this dissertation

...Professor Constance Scharff for giving me the opportunity to join the Free University, Berlin, for her generous support during my stay, and for critical review of my work

...Professor Ralf Einspanier for being my second supervisor and for reviewing my thesis

...Xintian You (Arthur) for his irreplaceable contribution on this project, especially for setting up the pipline for novel miRNA prediction, the figures and tables…

...Dr. Hao Hu, Dr. Marc Friedlaender and Dr. Zhuo Fang for their stimulating discussions and valuable suggestions, Sebastian Mackowiak from Prof. Nikolaus Rajewsky's group for mirDeep analysis supporting and many helpful discussions

...Sebastian Mackowiak and Marvin Jens from Rajewsky's group for the translation of the Zusammenfassung

...Dr. Yu Shi and Dr. Matthew Poy for providing me the mouse tissues and mouse N2a cells, and Nadine Thierfelder from Rajewsky's group for providing me the C.elegans sample

...All the members from Chen's lab for scientific and non scientific discussion, support and for making everyday work more enjoyable, particularly Ana Babic, Sun Wei, Yongbo Wang, Sun Wei(Sunny), Gangcai Xie, Martina Weight and Anna-Maria Stroehl. Dr. Yuhui Hu, Dr. Andreas Gogol-döring for their inspiring discussions and suggestions. Mirjam Feldkamp and Claudia Langnick for their technical support on sequencing

...My colleagues from BIMSB, especially those from Prof. Nikolaus Rajewsky's group, for the friendly atmosphere and for many scientific and non-scientific help and discussions

…The China Scholarship Council (CSC) for providing me with three-years-PhD scholarship

...Last, but above all, I owe my deepest gratitude to my parents and my family in China for their everlasting love and care. Over these years I could not visit them as often as I wished and should. But no matter where I am, I always know that they are right there supporting me, proud of me and believing in me.

TABLE OF CONTENTS

SUMMARY

MicroRNAs (miRNAs) constitute an important class of small regulatory RNAs that are derived from distinct hairpin precursors (pre-miRNAs). In contrast to mature miRNAs, which have been well characterized in numerous genome-wide studies of different organisms, research on global profiling of pre-miRNAs is limited.

Here, using massive parallel sequencing, we have for the first time performed global characterization of both mouse miRNAs and pre-miRNAs. In total, 87,369,704 and 252,003 sequencing reads derived from 887 miRNAs and 281 pre-miRNAs were obtained, respectively. With the sequence information of both, especially the pre-miRNAs which to our knowledge have not before been sequenced in genome-wide manner, several new aspects of processing and modification of known mouse miRNAs, including Ago2-cleaved pre-miRNAs, new instances of miRNA editing events, untemplated nucleotide additions at the 3' end of both miRNAs and the hairpin precursors, as well as exclusively 5' tailed mirtrons, were revealed.

Furthermore, based on the sequences of both mature and precursor miRNAs, we developed a novel miRNA discovery strategy that did not rely on the availability of genome reference sequences. With this strategy 238 known mouse pre-miRNAs could be recovered and 69 novel ones were predicted with high confidence. Similar to the known ones, the mature miRNAs derived from most of these novel loci showed reduced abundance following Dicer knock down. Evaluation on another dataset from *C. elegans* demonstrated that our pipeline could be applied for miRNA discovery in different organisms, especially in the absence of a reference genome.

We believe our method could be widely used in the study of miRNAs not only in the organisms whose genome has not yet been sequenced, but also in samples where the genome differs significantly from the reference sequences, such as cancer.

ZUSAMMENFASSUNG

MicroRNAs (miRNAs) stellen eine Klasse kleiner, regulatorischer RNA Moleküle dar, die aus längeren Vorläufer Molekülen hergestellt werden. Diese sogenannten 'precursor miRNAs' (pre-miRNAs) haben eine charakteristische Haarnadel Sekundärstruktur. Obwohl diese pre-miRNAs eine substanzielle Rolle bei der Entstehung der miRNAs spielen, sind diese im Vergleich zu miRNAs nur wenig untersucht worden. Dies läßt sich auch anhand der zahlreichen Publikationen über genomweite Untersuchungen von miRNAs in verschiedenen Organismen nachvollziehen.

In dieser Arbeit wird zum ersten Mal mit Hilfe von so genanntem 'massive parallel sequencing' eine genomweite Analyse beschrieben, in der miRNAs und pre-miRNAs simultan in dem selben Organismus untersucht werden. Insgesammt wurden 87.369.704 miRNA Moleküle und 252.003 Vorlauefer Moleküle sequenziert. Diese Moleküle konnten 887 miRNAs bzw. 281 verschiedenen pre-miRNAs zugeordnet werden. Mit dem Wissen über die Menge der Moleküle von miRNAs und pre-miRNAs konnten neue Einsichten bezüglich der Prozessierung und Modifikation von annotierten Maus miRNAs gewonnen werden. Unter anderem erhielten wir neue Informationen über durch Ago2 geschnittene pre-miRNAs, neue miRNA Modifikationen, zusätzliche Nukleotidvorkommmen am 3'-Ende von miRNAs und pre-miRNAs sowie mirtrons, deren 3'-Ende das Resultat von pre-mRNA Splicing ist.

Desweiteren haben wir eine computergestütze Methode entwickelt, die die miRNA und pre-miRNA Sequenzierdaten benutzt um neue miRNAs zu identifizieren. Im Vergleich zu anderen Methoden benötigt unser Ansatz kein Referenzgenom. Insgesamt haben wir 238 bekannte Maus pre-miRNAs identifiziert und 69 neue vorhergesagt. Durch einen sogenannten 'Knockdown' des Dicer Gens konnten wir eine ähnliche Verminderung der vorhergesagten miRNAs feststellen, wie dies auch bei den bekannten miRNAs der Fall war. Eine Evaluierung unserer Methode auf *C. elegans* Daten hat deutlich gezeigt, dass unser Ansatz auch in anderen Organismen gut funktioniert. Die Tatsache, dass kein Referenzgenom benötigt wird, macht unsere Methode auch nützlich für Organismen ohne sequenziertes Genom.

Wir sind der Überzeugung, dass unsere Methode sehr gut für die Identifikation von miRNAs in Organismus sowohl mit bereits sequenziertem Genom als auch nicht sequenziertem Genom geeignet ist. Darüber hinaus ist dieser Ansatz auch auf stark veränderte Genome, wie dies z.B. bei den meisten Krebszellen der Fall ist, anwendbar.

1 Introduction

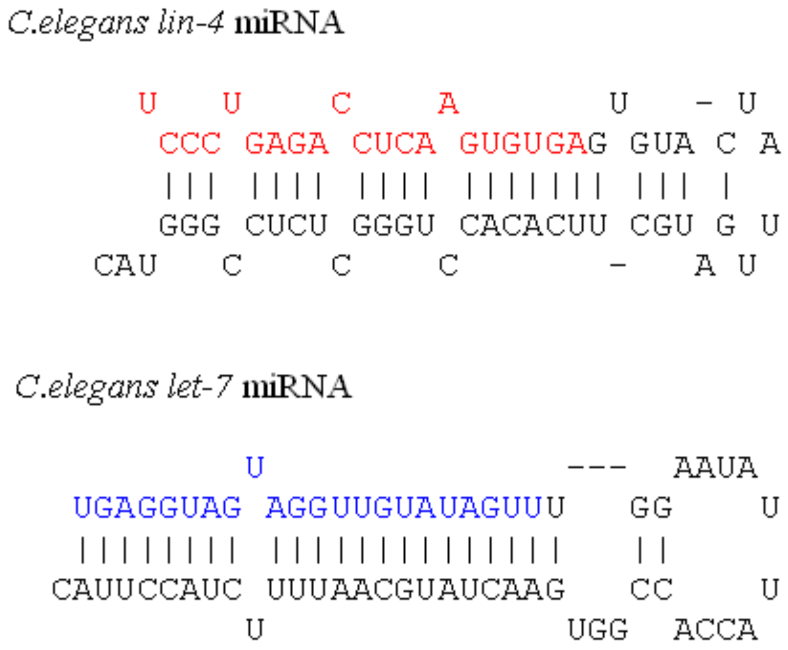
MicroRNAs (miRNAs) constitute an important class of small non-coding RNAs that regulate gene expression at the post-transcriptional level through sequence-specific base pairing. Most miRNAs are transcribed by the RNA polymerase II to generate primary miRNA (pri-miRNA) transcripts. For canonical miRNAs, the pri-miRNAs bearing one or more imperfect inverted repeats are cleaved by the RNase III enzyme Drosha to yield hairpin-shaped precursor miRNAs (pre-miRNAs). Alternatively, pre-miRNAs can be generated from debranched short introns with hairpin potential (mirtron) by the spliceosome complex, or can be derived from other small non-coding RNAs such as small nucleolar RNAs (snoRNAs). After being transported into the cytoplasm by the exportin-5 complex, pre-miRNAs are further processed by another RNase III enzyme Dicer into double stranded miRNA:miRNA* duplexes, of which one strand is incorporated into the miRNA-induced silencing complex (miRISC) and guides the effector complex to silence target mRNAs through mRNA cleavage, translational repression or deadenylation.

1.1 A short history of miRNA

The founding members of the miRNA family, *lin-4* and *let-7*, were both identified by genetic screens for the temporal control of post-embryonic developmental timing in *C.elegans*. Loss-of-function mutation of *lin-4* or *let-7* gene activity causes reiteration of early fate during late developmental stages (Chalfie, Horvitz et al. 1981; Reinhart, Slack et al. 2000). In 1993, *lin-4* was first reported as a 22-nucleotide (nt) non-coding RNA partially complementary to 7 repeated sequences in the 3' untranslated region (UTR) of the *lin-14* mRNA (Lee, Feinbaum et al. 1993). The studies suggested that small non-coding RNA *lin-4* acts as a translational repressor of *lin-14* mRNA via RNA-RNA interaction in its 3' UTR.

The discovery of small RNA *lin-4* and its target-specific translational regulation suggested a new mechanism of post-transcriptional gene regulation during development. In 2000, the second miRNA, *let-7*, was identified in *C.elegans* by Ruvkun lab (Reinhart, Slack et al. 2000). Like *lin-4*, the 21-nt *let-7* is also generated from a double-stranded hairpin precursor (Figure 1.1), and it controls late temporal transitions during development through the functional binding to the 3' UTR of *lin-41* and *hbl-1* (*lin-57*), thereby inhibiting their translation (Abrahante, Daul et al. 2003; Vella, Choi et al. 2004). Unlike *lin-4*, *let-7* is evolutionarily conserved among a wide range of animal species, including vertebrate, ascidian, hemichordate, mollusc, annelid and arthropod (Pasquinelli, Reinhart et al. 2000), which indicates its general role in gene regulation.

In 2001, intrigued by the idea that *lin-4* and *let-7* might represent the first memebers of an abundant class of yet unknown small regulatory RNAs, three groups started to clone small RNAs from different species and found hundreds of so-called miRNAs in worm, fly and mammal (Lagos-Quintana, Rauhut et al. 2001; Lau, Lim et al. 2001; Lee and Ambros 2001). Since then, even more miRNAs have been discovered in almost all metazoan genomes and their important role as post-transcriptional regulators of gene expression has been revealed.


C.elegans lin-4 miRNA

C.elegans let-7 miRNA

U U C A U - U
CCC GAGA CUCA GUGUGAG GUA C A
||| |||| |||| ||||||| ||| |
GGG CUCU GGGU CACACUU CGU G U
CAU C C C - A U
U --- AAUA
UGAGGUAG AGGUUGUAUAGUUU GG U
|||||||| |||||||||||||| ||
CAUUCCAUC UUUAACGUAUCAAG CC U
U UGG ACCA

Figure 1.1 *C.elegans lin-4* and *let-7* miRNAs. *lin-4* precursor hairpin (top), with sequence of the 22 nt mature miRNA in red and *let-7* precursor hairpin (bottom), with sequence of the 22 nt mature miRNA in blue.

1.2 miRNA genes and their genomic distribution

Deep-sequencing technologies provide a great opportunity in accelerating the rate of novel miRNA discovery. Nowadays, the miRNA database (http://www.mirbase.org/index.shtml) (Kozomara and Griffiths-Jones 2011) contains more than 15,000 miRNA gene loci in over 140 species. In mouse genome, miRNA genes are scattered on all chromosomes except for the Y chromosome. Some miRNA genes have multiple paralogues that might arise from gene duplication, for example, the mouse let-7 family contains 13 members. In many cases, miRNA genes are found in clusters and transcribed from a single polycistronic transcription unit (TU) (Lagos-Quintana, Rauhut et al. 2001; Lau, Lim et al. 2001; Lee, Jeon et al. 2002; Mourelatos, Dostie et al. 2002). There are 46 miRNA clusters in mouse genome, 35 of which are generated from non-coding TUs, whereas others are overlapped with protein coding genes (http://www.diana.pcbi.upenn.edu/miRGen/v3/miRGen.html) (Megraw, Sethupathy et al. 2007).

Mammalian miRNAs can be categorized into four groups according to their genomic locations (Kim, Han et al. 2009).

- ***miRNAs in intron of non-coding transcripts (~40%).*** E.g. the mir-15a and mir-16-1 are clustered and located within the intron of a non-coding RNA referred to as deleted in lymphocytic leukemia 2 (DLEU2).
- ***miRNAs in exon of non-coding transcripts (~10%).*** E.g. mir-155 is found in the BIC non-coding mRNA.
- ***miRNAs in intron of protein-coding transcripts (~40%).*** E.g. mir-186 is located in intron 8 of the zinc finger protein 265 (Znf265). The mir-25~93~106b cluster is embedded in the intron 13 of the DNA replication licensing factor minichromosome maintenance protein 7 (MCM7) transcript.
- ***miRNAs in exon of protein-coding transcripts (~10%).*** E.g. mir-671 is located in the last exon of chondroitin polymerizing factor 2 (Chpf2), mir-686 is found in the second exon of Proteasome subunit beta type-5 (Psmb5).

Some ‘mixed’ miRNA genes can be assigned to either exonic or intronic regions depending on the alternative splicing pattern of the host transcripts.

1.3 miRNA biogenesis

miRNA biogenesis consists of multiple processing steps involving several protein complexes in nuclear and cytoplasm (Figure 1.2).

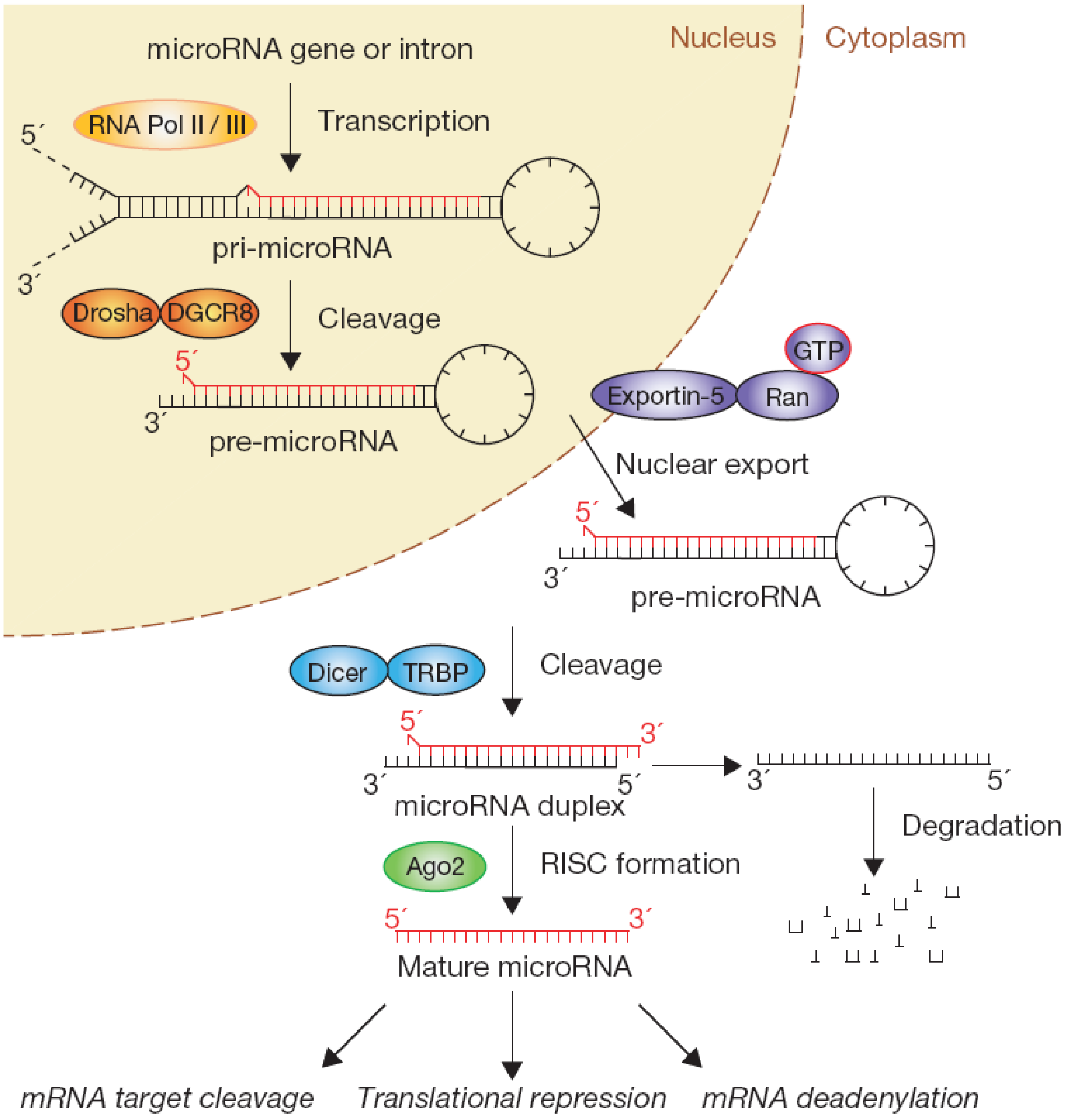


Figure 1.2 The canonical pathway of microRNA biogenesis. The canonical miRNA gene is transcribed by RNA polymerase II or III to generate the pri-miRNA. In the nucleus, the initiation step is mediated by the microprocessor complex Drosha–DGCR8 (Pasha) to generate the hairpin shaped pre-miRNA. The pre-miRNA is exported to the cytoplasm by Exportin-5–Ran-GTP, where the RNase III Dicer in complex with the double-stranded RNA-binding protein TRBP cleaves the pre-miRNA hairpin to miRNA duplex. One strand is loaded into the RNA-induced silencing complex (RISC), where it guides RISC to silence target mRNAs through mRNA cleavage, translational repression or deadenylation, whereas the other strand is degraded. Adapted from (Winter, Jung et al. 2009).

1.3.1 The Microprocessor complex

Most miRNA genes are transcribed by RNA polymerase II to generate the pri-miRNA transcripts which contain 5'-cap structure, poly(A) tail and even are spliced (Bracht, Hunter et al. 2004; Cai, Hagedorn et al. 2004; Lee, Kim et al. 2004). However, a minor group of miRNAs that are associated with Alu repeats can be transcribed by RNA polymerase III (Borchert, Lanier et al. 2006). The pri-miRNA transcripts range in size from hundreds of nucleotides to several hundred kilobases and contain one or more stem-loop structures corresponding to individual miRNAs.

The first step of miRNA maturation takes place in the nucleus by a multiprotein complex called the Microprocessor complex, which is composed of an RNase III enzyme Drosha and its cofactor, the DiGeorge syndrome critical region gene 8 (DGCR8) in mammalians, Pasha in *D.melanogaster* and *C.elegans* (Lee, Ahn et al. 2003; Denli, Tops et al. 2004; Gregory, Yan et al. 2004; Han, Lee et al. 2004; Landthaler, Yalcin et al. 2004). This complex recognizes the stem of hairpin structure in pri-miRNA and cleaves at a site ~11 bp away from the single-stranded/double-stranded RNA junction (SD junction) (Han, Lee et al. 2006), producing a ~70 nt hairpin-shaped pre-miRNA with ~2 nt 3' overhang (Figure 1.3).

Drosha is the key nuclease that initiates the miRNA processing, in which two tandem RNase III domains (RIIIDs) form an intramolecular structure with each domain cleaving one strand of pri-miRNA to generate a pre-miRNA with 2 nt 3' overhang. However, neither Drosha nor DGCR8 alone can cleave the pri-miRNA precisely, which indicates that the Microprocessor complex containing both Drosha and DGCR8 is necessary and sufficient for the processing of pri-miRNA to pre-miRNA (Gregory, Yan et al. 2004).

DGCR8 is an evolutionarily conserved protein that contains two double-stranded RNA-binding domains (dsRBDs) and a WW domain (Figure 1.3B). DGCR8 is located in the chromosomal region 22q11.2 whose heterozygous deletion results in the most common human genetic deletion syndrome, known as DiGeorge syndrome (Shiohama, Sasaki et al. 2003; Yamagishi and Srivastava 2003). DGCR8 acts as the molecular ruler and measures the distance (~11 nt) from the dsRNA-ssRNA junction of pri-miRNA, whereas Drosha cleaves the 5' and 3' arms of the pri-miRNA hairpin.

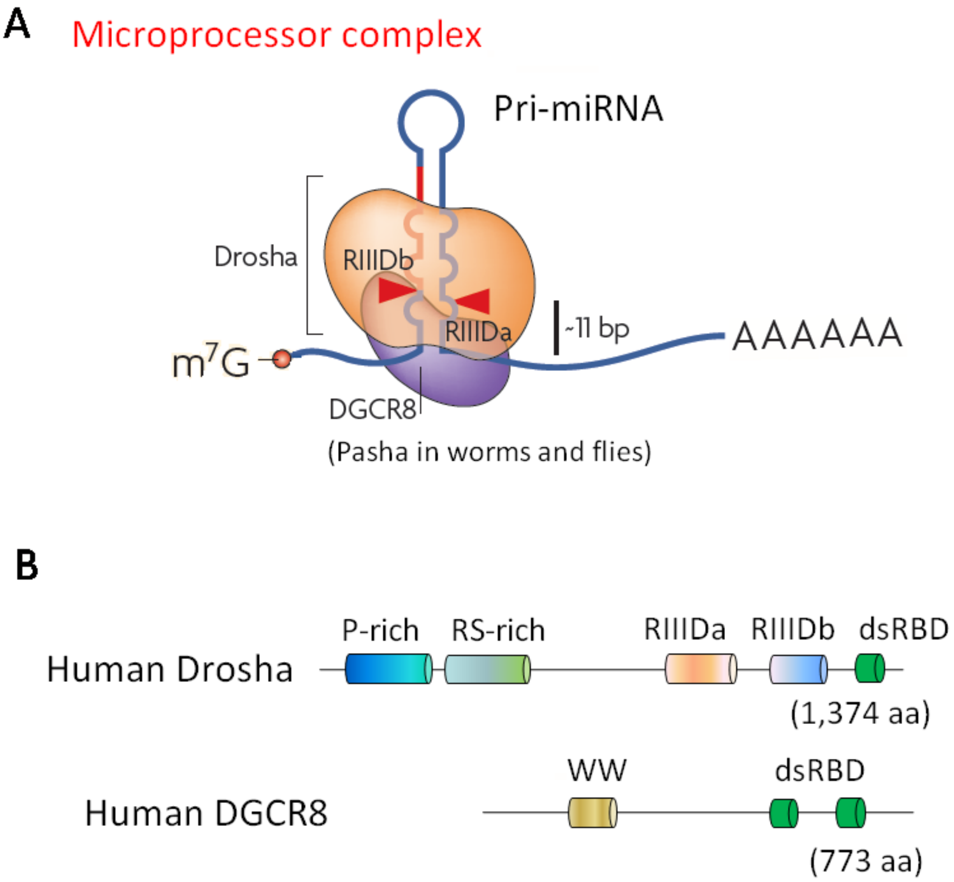


Figure1.3 Pri-miRNA is processed by Drosha and its partner DGCR8. The pri-miRNA is recognized by the RNase III enzyme Drosha and its cofactor DGCR8 protein in mammalians (Pasha in *D.melanogaster* and *C.elegans*). Drosha interacts with a dsRNA-binding protein, DGCR8 in mammalians (Pasha in *D.melanogaster* and *C.elegans*), through its middle region. This Drosha-DGCR8 complex is known as the Microprocessor complex (A). Domain structure of Drosha and DGCR8 are shown in B. Adapted from (Kim, Han et al. 2009).

1.3.2 Alternative routes into the miRNA hairpin precursors

mirtron

'mirtron' pathway, the first alternative miRNA biogenesis pathway, has been discovered in flies, worms and mammals (Berezikov, Chung et al. 2007; Kim and Kim 2007; Okamura, Hagen et al. 2007). Mirtrons are derived from short intron with hairpin potential. Following splicing, the branch point of the lariat-shaped intron is resolved and the debranched intron forms a pre-miRNA hairpin, which can be exported into cytoplasm via Exportin-5 and cleaved by Dicer, thus bypassing the processing by Microprocessor. Beside the canonical mitrons, some mirtron-like loci have been identified, in which the small RNA hairpin resides at one end of the large intron (Ruby, Jan et al. 2007; Babiarz, Ruby et al. 2008), and therefore requires exonucleolytic trimming to form pre-miRNA hairpin (Figure 1.4).

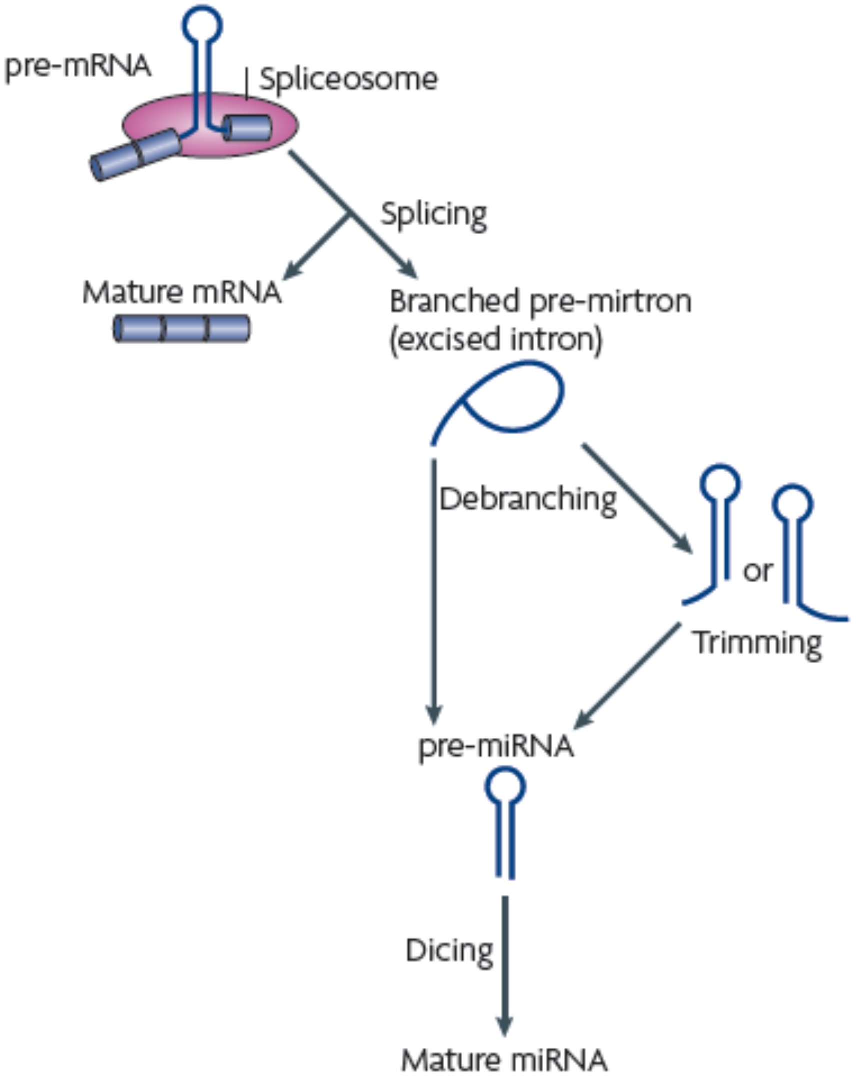


Figure 1.4 Non-canonical intronic small RNAs are produced from spliced introns and debranching, such small RNAs (called mirtrons) can derive from small introns that resemble pre-miRNAs that bypass the Drosha-processing step. Some introns have tails at either the 5' or 3' end, which need to be trimmed before pre-miRNA export. Adapted from (Kim, Han et al. 2009).

miRNAs derived from snoRNAs

Some snoRNAs can generate pre-miRNAs independent of the Microprocessor-mediate processing (Ender, Krek et al. 2008; Saraiya and Wang 2008). Deep-sequencing of small RNAs associated with immunopurified human Ago1 and Ago2 revealed that some miRNA-like molecules originate from particular snoRNAs. Processing of these snoRNAs requires

Dicer activity but is independent of Drosha/DGCR8. Recently more studies supported that snoRNAs could be involved in miRNA biogenesis (Taft, Glazov et al. 2009; Brameier, Herwig et al. 2011; Ono, Scott et al. 2011).

1.3.3 The Exportin-5 complex

Following the processing in nucleus, pre-miRNAs are exported to the cytoplasm, by the transporter Exportin 5 (Exp-5) (Figure 1.5). Exp-5 is a member of the nuclear karyopherin β transporter receptor family. Like other nuclear transport receptors, Exp-5 cooperates with the small GTPase Ran to mediate directional export. In the nucleus, pre-miRNA binds to Exp-5 and the GTP-bound form of the cofactor Ran. After exported to the cytoplasm, GTP is hydrolyzed and the pre-miRNA is released from Exp-5. Exp-5 recognizes the >14 bp dsRNA stem and 2 nt 3' overhang (Basyuk, Suavet et al. 2003; Gwizdek, Ossareh-Nazari et al. 2003; Zeng and Cullen 2004; Okada, Yamashita et al. 2009).

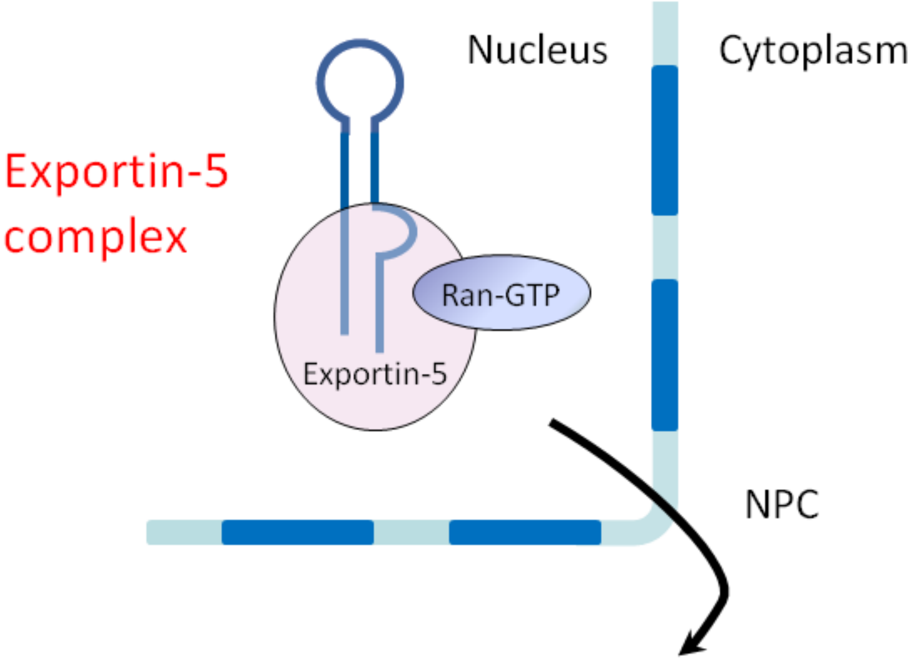


Figure 1.5 The exportin-5 complex. Pre-miRNA exported from nucleus into cytoplasm through nuclear pore complex (NPC) by the Exportin-5 complex involving RanGTP hydrolysis.

1.3.4 The pre-miRNA processing complex

Following export from the nucleus, pre-miRNAs are recognized by the RNA-induced silencing complex (RISC) containing Dicer-TRBP-Ago2, and are cleaved near the terminal loop into ~22 nt miRNA:miRNA* duplexes by the RNase III protein Dicer in the cytoplasm (Bernstein, Caudy et al. 2001; Grishok, Pasquinelli et al. 2001; Hutvagner, McLachlan et al. 2001; Ketting, Fischer et al. 2001; Knight and Bass 2001).

Dicer contains two copies of the universally conserved catalytic RNase III domain (RIIID), an N-terminal ATPase/DExDhelicase domain, a small domain of unknown function (DUF283), a C-terminal dsRNA-binding domain (dsRBD) and a central PIWI/Ago/Zwille (PAZ) domain upstream of the two RIIIDs (Figure1.6 A). Dicer is a highly conserved protein among almost all eukaryotic organisms, including *Schizosaccharomyces pombe*, plants and animals (Figure1.6 B).

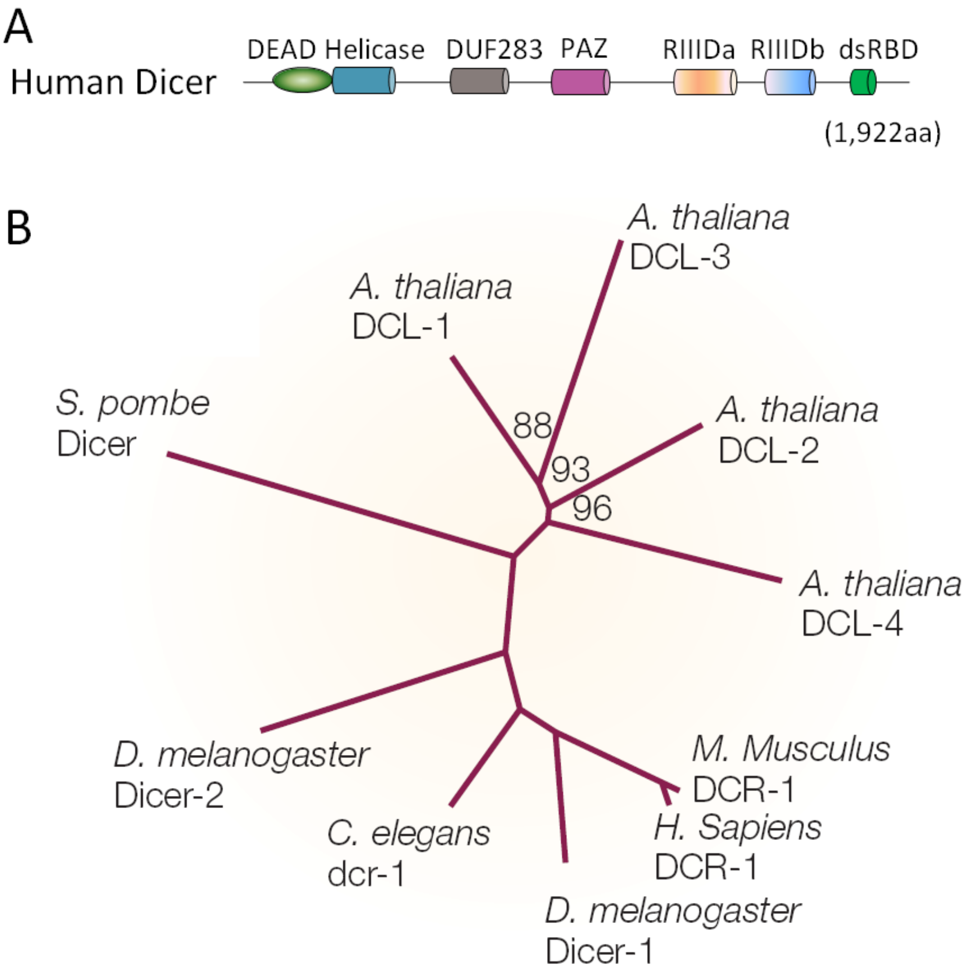


Figure 1.6 The structure and function of the Dicer family. A. The domain structure of Human Dicer. B. The phylogenetic tree of the Dicer protein family. *A. thaliana, Arabidopsis thaliana; C. elegans, Caenorhabditis elegans; D. melanogaster, Drosophila melanogaster; H. sapiens, Homo sapiens; M. musculus, Mus musculus;* RBD, RNA binding domain; *S. pombe, Schizosaccharomyces pombe.* Adapted from (He and Hannon 2004).

Several models have been proposed for how Dicer generates RNA fragments with the specific size of ~22 nt, based on either an X-ray structure of the *Aquifex aeolicus* RNase III and biochemical data obtained from recombinant human Dicer or by studying the crystal structure of an intact and fully active Dicer from *Giardia intestinalis* combined with biochemical analyses (Zhang, Kolb et al. 2004; Macrae, Zhou et al. 2006; MacRae, Zhou et al. 2007). In the model, the PAZ domain of Dicer specifically recognizes the 3' overhang present in the pre-miRNA terminus, and the connector helix is responsible for the measurement of 21-23 bp length and determines the positioning at the cleavage site of the two RNase III domains. Similar to Drosha, each RNase III domain of Dicer cleaves one strand of the miRNA duplex at the opposite end of the extremity produced by Drosha (Figure 1.7).

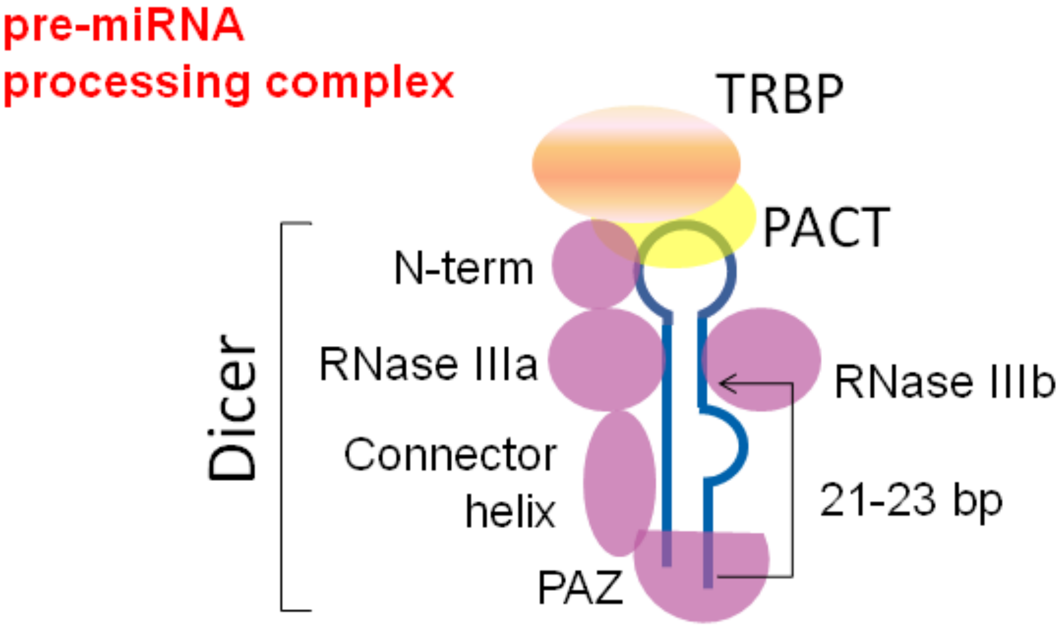


Figure 1.7 The model for how Dicer works in pre-miRNA processing. The PAZ domain of Dicer specifically recognizes the 3' overhang present in the pre-miRNA terminus, and that the connector helix is responsible for the measurement of the 21-23 bp length and determines the positioning at the cleavage site of the two RNase III domains. Each RNase III domain of Dicer cleaves one strand of the miRNA duplex, at the opposite end of the extremity produced by Drosha.

TRBP (the human immunodeficiency virus transactivating response RNA-binding protein) was reported to interact with Dicer and Argonaute 2 (Ago2). The depletion of TRBP by RNAi caused defects of siRNA- or miRNA-mediated RNA silencing processes in human cell lines (Chendrimada, Gregory et al. 2005; Haase, Jaskiewicz et al. 2005).

PACT (the PKR-activating protein) has been reported to be associated with a complex containing Dicer, Ago2, and TRBP (Lee, Hur et al. 2006). Although, neither TRBP nor PACT is required for Dicer processing itself, they seem to directly interact with each other and further associate with Dicer to facilitate siRNAs production (Kok, Ng et al. 2007).

Ago2-mediated pre-miRNA cleavage: the ac-pre-miRNA

For some miRNAs with perfect complementarity in the hairpin stem, an additional processing intermediate product, named as Ago2-cleaved pre-miRNA (ac-pre-miRNA), can be generated by Ago2 before Dicer-mediated cleavage. Ago2, which has robust RNaseH-like endonuclease activity, cleaves the 3' arm of the hairpin in the middle to generate a nicked hairpin. Dicer can process this nicked precursor as efficient as the full-length pre-miRNA. The Ago2-mediated step may facilitate the following strand dissociation and RISC activation (Diederichs and Haber 2007).

Processing of pre-mir-451 also requires cleavage by Ago2, which is independent of Dicer (Cheloufi, Dos Santos et al. 2010; Cifuentes, Xue et al. 2010).

1.3.5 The RISC complex

Following Dicer cleavage, the ~22 nt RNA duplex remains associated with RISC as a ribonucleoprotein effector complex. In human cells, Dicer, TRBP (and/or PACT) and Ago2 (and perhaps other Ago proteins) initiate the assembly of miRISC (miRNA-induced silencing complex) by forming a RISC loading complex (RLC) (Gregory, Chendrimada et al. 2005; Maniataki and Mourelatos 2005). The miRNA strand with relatively lower stability at its 5’ end remains in Ago as a mature miRNA, whereas the other strand (miRNA*) is released and degraded (Aza-Blanc, Cooper et al. 2003; Khvorova, Reynolds et al. 2003; Schwarz, Hutvagner et al. 2003) (Figure 1.8). However, miRNA* strand is not always by-product of miRNA biogenesis and in many cases it can also be loaded into miRISC to function as miRNA (Czech, Zhou et al. 2009; Okamura, Liu et al. 2009; Chiang, Schoenfeld et al. 2010; Ghildiyal, Xu et al. 2010). As part of miRISC, miRNA binds to the 3’ UTR of target mRNAs via base-pairing at nucleotides 2 to 8 (the seed region), and then leads to their translational repression, and/or mRNA destabilization.

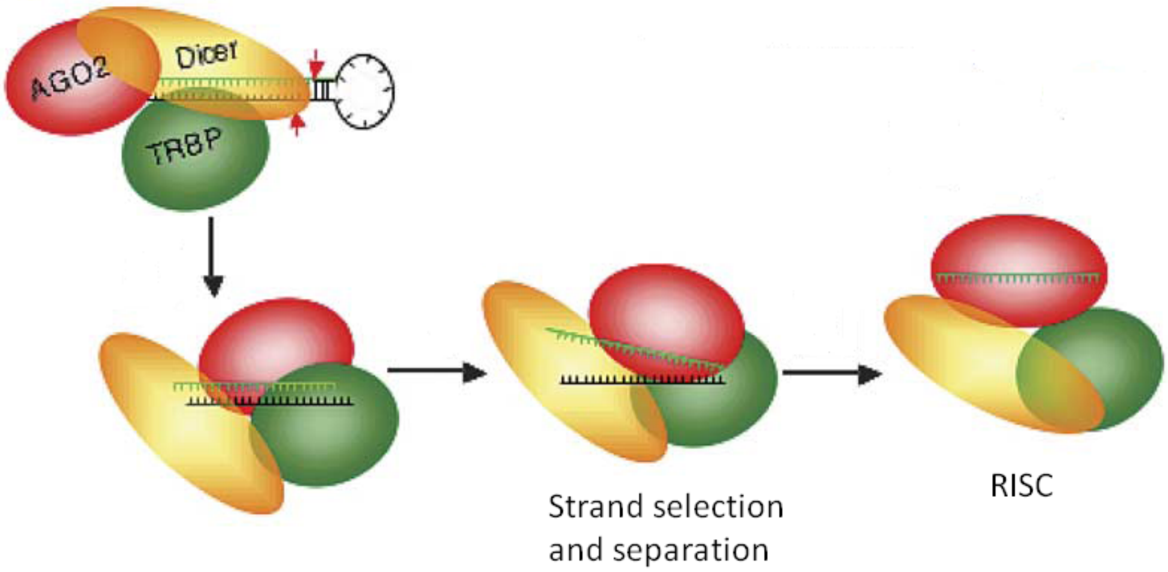


Figure 1.8 The model of RISC assembly. Pre-miRNAs are transported to the cytoplasm, where they are recognized by the RISC containing Dicer-TRBP-Ago2, and cleaved by Dicer to generate ~22 nt duplex miRNA (2 nt 3' overhangs). One strand of duplex is selected and the other strand is degraded. The guide strand of the miRNA remains associated with Ago2 in the active RISC complex. Adapted from (Gregory, Chendrimada et al. 2005).

Argonaute 2 (Ago2) The most important and well studied component of miRISC are proteins of the Argonaute family (Peters and Meister 2007; Tolia and Joshua-Tor 2007). Based on amino acid sequences alignments, the Argonaute family has been divided into two subfamilies: the Ago and Piwi family (Carmell, Xuan et al. 2002). Argonaute proteins contain a PAZ (Cerutti, Mian et al. 2000) and a PIWI (P-element induced wimpy testis) domain (Cerutti, Mian et al. 2000; Carmell, Xuan et al. 2002). In mammals, four Ago proteins, Ago1 to Ago4, function in the miRNA repression (Liu, Carmell et al. 2004; Meister, Landthaler et al. 2004; Pillai, Artus et al. 2004). Ago2 is the only human Ago protein with endonuclease activity because of its intact RNaseH-like PIWI domain (DDE motif) (Song, Smith et al. 2004).

1.4 Regulation of miRNA gene expression

1.4.1 Transcriptional regulation of miRNA gene expression

The cellular miRNA abundance could be regulated at the transcriptional level. Many transcription factors can control either positively or negatively the miRNA expression. For example, MYC activates expression of the mir-17-92 oncogenic clusters in lymphoma cells

(He, Thomson et al. 2005; O'Donnell, Wentzel et al. 2005) and mir-9 in nueroblastoma cells (Ma, Young et al. 2010), but represses a number of tumor suppressor miRNAs (Chang, Yu et al. 2008). The tumor suppressor p53 enhances the expression of mir-34 and mir-107 families, which are involved in cell cycle and apoptosis (He, He et al. 2007). Moreover, miRNAs can regulate their own transcription through negative or positive feedback loops with specific transcription factors. Examples of this type of regulation include Runx1-mir-27a in megakaryopoiesis (Ben-Ami, Pencovich et al. 2009), Pitx3-mir-133b in midbrain dopaminergic neurons (Kim, Inoue et al. 2007) and c-Myb-mir-15a in hematopoiesis (Zhao, Kalota et al. 2009). Also epigenetic control could contribute to the transcriptional regulation of miRNA gene, e.g. many miRNA gene loci including mir-137, -193a, -203, -342, -34b/c and -9-1, are found to be hypermethylated in several human cancers (Lujambio, Calin et al. 2008; Lujambio and Esteller 2009).

1.4.2 Post-transcprtional regulation of miRNA gene expression

The regulaton of miRNA expression could also be achieved after the pri-miRNA transcription.

Regulation of Drosha, Dicer and their double-stranded RBP partners

RNase III enzyme Drosha and Dicer form processing complexes together with double-stranded RBP partners, such as DGCR8 and TRBP, respectively. Both the levels and activities of all of these proteins are regulated in various ways. For example, Drosha together with DGCR8 forms a regulatory feedback loop. DGCR8 stabilizes Drosha through an interaction between its conserved carboxyl-terminal domain and the middle domain of Drosha (Yeom, Lee et al. 2006), whereas Drosha decreases DGCR8 level by cleaving two hairpin structures in the 5' UTR and the coding region of the *Dgcr8* mRNA (Han, Pedersen et al. 2009). A delicate balance between the Drosha and DGCR8 is important for miRNA biogenesis, as a threefold excess of DGCR8 dramatically inhibits Drosha processing activity *in vitro* (Gregory, Yan et al. 2004). As another example, accumulation of Dicer is dependent on its partner TRBP, and a decrease in TRBP leads to Dicer destabilization and pre-miRNA processing defects (Chendrimada, Gregory et al. 2005; Melo, Ropero et al. 2009; Paroo, Ye et al. 2009). In human carcinomas, the presence of TARBP2 frameshift mutations causes diminished TRBP protein expression and a defect in the processing of miRNAs (Melo, Ropero et al. 2009). Moreover, Dicer is also controlled by its own product, *let-7*, which binds to the 3' UTR and coding region of the Dicer mRNA (Forman, Legesse-Miller et al. 2008;

Tokumaru, Suzuki et al. 2008). This negative feedback loop suggests the potential of *let-7* to broadly influence miRNA biogenesis (Krol, Loedige et al. 2010).

Role of accessory protein Many proteins are involved in the regulating miRNA processing either by interacting with Drosha or Dicer, or by binding to miRNA precursors (Winter, Jung et al. 2009; Krol, Loedige et al. 2010).

The best-studied negative regulator of miRNA biogenesis is Lin-28, which is a stem-cell-specific regulator of let-7 processing via multiple mechanisms (Heo, Joo et al. 2008; Newman, Thomson et al. 2008; Piskounova, Viswanathan et al. 2008; Rybak, Fuchs et al. 2008; Viswanathan, Daley et al. 2008). Mature let-7 increases during embryonic stem cell differentiation but the pri-miRNA level remains the same, indicating post-transcriptional regulation of *let-7* expression. It has been shown that by binding to the terminal loop of pri-let-7, Lin-28 interferes with cleavage by Drosha (Newman, Thomson et al. 2008; Viswanathan, Daley et al. 2008). In addition, Lin-28 could also bind to pre-let-7 to block its processing by Dicer (Heo, Joo et al. 2008; Rybak, Fuchs et al. 2008). Finally, Lin-28 associates with cytoplasmic pre-let-7 and induces its 3'-terminal polyuridylation by the TUT4 terminal uridylyl transferase. Such uridylation prevents Dicer processing and accelerates the decays of pre-let-7 (Heo, Joo et al. 2009).

The p68 and p72 RNA helicases are identified as components of the Drosha Microprocessor complex and might act as specific factors to stimulate processing of one-third of pri-miRNAs (Fukuda, Yamagata et al. 2007). In both *p68*$^{-/-}$ and *p72*$^{-/-}$ embryos, expression levels of a set of, but not all, miRNAs and 5.8S rRNA are significantly reduced.

Drosha-mediated cleavage can also be regulated for individual miRNAs. For instance, without any impact on other miRNAs that are located in the same genomic locus, the heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) binds specifically to pri-mir-18a and facilitates its processing (Guil and Caceres 2007). Transforming growth factor-β (TGF-β) and bone morphogenetic factors (BMPs) induce the maturation of mir-21 by regulating the Drosha-mediated miRNA processing (Davis, Hilyard et al. 2008). It was proposed that TGF-β- and BMP-specific SMAD signal transducers are recruited to pri-mir-21 in a complex with p68 and Drosha. As a consequence, Drosha-mediated processing of pri-mir-21 is strongly enhanced and the abundance of mature mir-21 increases.

1.4.3 RNA editing of miRNA

RNA editing is a post-transcriptional change of RNA sequences. A-to-I RNA editing, conversion of adenosine (A) to inosine (I), is mediated by adenosine deaminase acting on RNA (ADAR) enzymes. Many pri-miRNAs are targeted by ADARs. Since the base-pairing properties of inosine are similar to those of guanosine (G), RNA editing could either by influence miRNA processing or changing their target specificity. For example, A-to-I editing in pri-mi-142 inhibits Drosha cleavage and results in its degradation by the Tudor-SN, which is a ribonuclease specific to inosine (Kawahara, Megraw et al. 2008). Editing can also influence further processing step: editing in pre-mir-151 prevents Dicer cleavage in cytoplasm (Kawahara, Zinshteyn et al. 2007). miRNA editing within seed sequences has an impact on miRNA target specificity. For example, tissue-specific A-to-I editing of mir-376 cluster transcript causes a single change in the seed sequence of mir-376, resulting in redirection of silencing targets, such as phosphoribosyl pyrophosphate synthetase 1 (PRPS1) which is not expressed by unedited miRNA (Kawahara, Zinshteyn et al. 2007). So far, 16 editing site in mature miRNAs are identified in mouse brain, 8 of them in the seed region (Chiang, Schoenfeld et al. 2010), indicating that editing events can expand the set of miRNA targets.

1.4.4 Regulation of miRNA turnover

miRNA decay

In comparison to our knowledge about the biogenesis of miRNAs, little is known about its turnover. It is generally thought that mature miRNAs are more stable than average mRNAs; an analysis of microRNA turnover in mammalian cells following Dicer1 ablation estimated an average miRNA half-life of 119 hours (i.e.~5 days) (Gantier, McCoy et al. 2011). However, certain miRNAs might degrade much more rapidly under specific conditions to serve their roles in developmental or metabolic transitions. For examples, mir-29b decays faster in cycling mammalian cells than in cells arrested in mitosis, which depends on the 3’ terminal motif of mir-29b but not for mir-29a (Hwang, Wentzel et al. 2007). In neurons, the neuron specific mir-183/96/182 cluster, mir-204 and mir-211, are downregulated during dark adaptation and upregulated in light, with both rapid miRNA decay and increased transcription being responsible for the changes. A similar high turnover of miRNAs (mir-124, mir-128, mir-134 and mir-138) also occurs in primary rodent neurons and neurons differentiated from mouse ESCs (Krol, Busskamp et al. 2010). Recently, several nucleases have been identified to degrade small RNAs. In *A.thaliana*, degradation of mature miRNAs is mediated by a family of 3’ to 5’ exoribonucleases, small RNA degrading nuclease 1 (SDN1), SDN2 and SDN3 (Ramachandran and Chen 2008). In *C.elegans*, XRN-2 (a 5’ to 3’ exonuclease) is involved in the degradation of mature miRNAs (Chatterjee and Grosshans 2009). In mammals, a general nuclease for miRNAs has yet to be identified.

3’ End Modification

The 3’ ends of mature miRNAs are often post-transcriptionally modified by adding 1-3 non-genome-encoded nucleotides. Large-scale cDNA analyses have identified several miRNAs from mammals (Landgraf, Rusu et al. 2007; Azuma-Mukai, Oguri et al. 2008; Burroughs, Ando et al. 2010; Chiang, Schoenfeld et al. 2010), worms (Ruby, Jan et al. 2006) and flies (Berezikov, Robine et al. 2010) that have 3’ adenylation or 3’ uridylation. These untemplated nucleotides addition might affect miRNA stability and abundance (Figure 1.9). For example, mir-122 is adenylated by cytoplasmic poly(A) polymerase GLD-2 and this 3’-terminal adenylation is required for the selective stabilization of mir-122 in the liver (Katoh, Sakaguchi et al. 2009). mir-26a is uridylated by Zcchc11 nucleotidyltransferase, and this modification abrogates its repressive function (Jones, Quinton et al. 2009). As described in 1.4.2, Lin-28 promotes uridylation of pre-let-7 by recruiting the TUT4 (known as Zcchc11 or PUP-2 in worms). Polyuridylation of pre-let-7 prevents Dicer processing and induces its degradation. Recently, kinetic analysis suggests a mechanism that target interaction can lead to uridylation of the miRNA which accelerates its decay (Baccarini, Chauhan et al. 2011).

1.5 How do miRNAs regulate gene expression?

miRNAs are key post-transcriptional regulators of gene expression that play important roles in a wide range of biological processes, including development, cellular differentiation, proliferation, apoptosis and metabolism (Bartel 2009; Carthew and Sontheimer 2009; Voinnet 2009; Huntzinger and Izaurralde 2011), as well as many human pathologies, such as cancer and metabolic disorders (Esquela-Kerscher and Slack 2006; Kloosterman and Plasterk 2006; Krutzfeldt and Stoffel 2006; Bushati and Cohen 2007; Chang and Mendell 2007). Computational predictions and genome-wide identification of miRNA targets estimate that mammalian miRNAs can regulate ~30% of all protein-coding genes (Krek, Grun et al. 2005; Lewis, Burge et al. 2005). In animals, the initial evidence suggested that miRNAs repress their targets largely at the translational level, without promoting the degradation of their target mRNAs. However, recent studies, on the genome-wide scale, reported that degradation of miRNA target is a widespread effect of miRNA-mediated regulation in gene expression (Baek, Villen et al. 2008; Selbach, Schwanhausser et al. 2008; Hendrickson, Hogan et al. 2009; Guo, Ingolia et al. 2010).

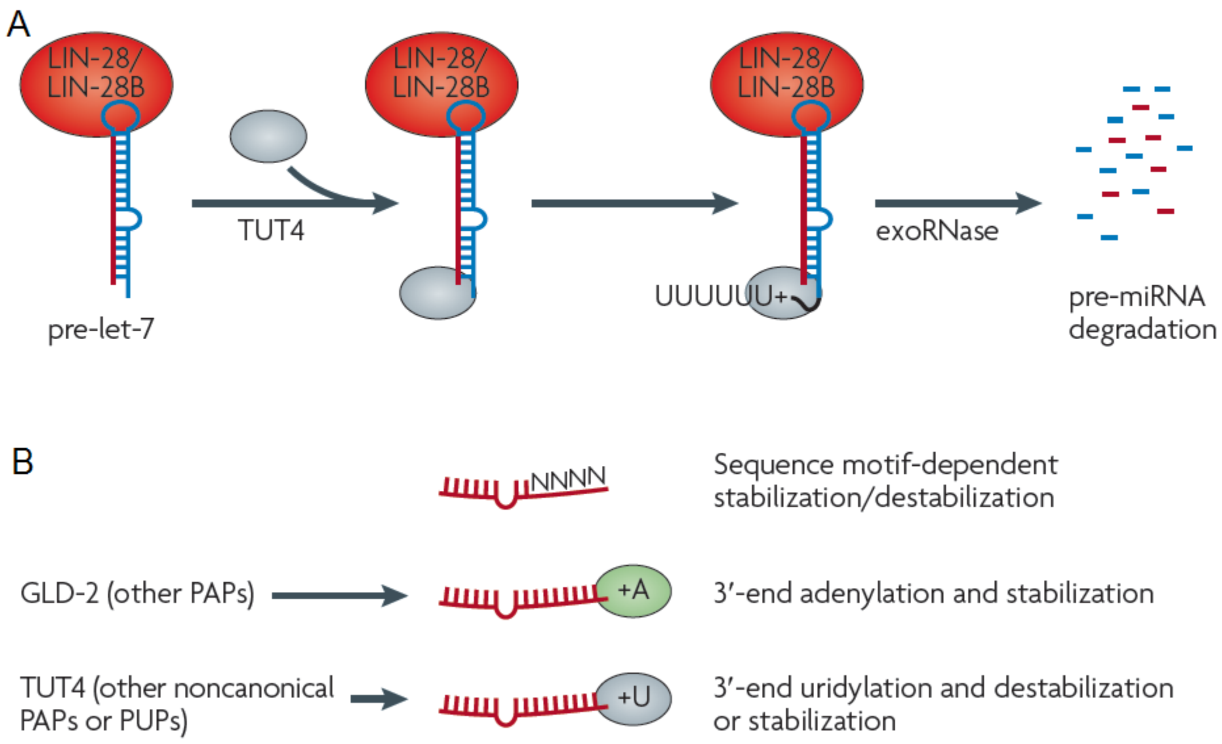

A
LIN-28/
LIN-28B
TUT4
UUUUUU+
exoRNase
pre-let-7
pre-miRNA
degradation
B
NNNN
Sequence motif-dependent stabilization/destabilization
GLD-2 (other PAPs)
+A
3′-end adenylation and stabilization
TUT4 (other noncanonical PAPs or PUPs)
+U
3′-end uridylation and destabilization or stabilization

Figure 1.9 Modification at the 3' end of miRNA. The post-transcriptional non-genome-encoded additions to the 3' end of either pre-miRNA or mature miRNA affects miRNA stability or abundance. A. The RNA-binding protein LIN-28 increases uridylation of pre-let-7 in *C.elegans* and mammalian cells by recruiting the poly(U) polymerase (PUP) TUT4 (also known as Zcchc11 or PUP-2 in worms), adding multiple uracil resides to the 3' end of RNA substrates. Polyuridylation of pre-let-7 impedes Dicer processing and causes precursor degradation by an unknown nuclease. B. RNA stability is affected by the 3' end motif or modifications (adenylation by poly(A) polymerase (PAP), uridylation by PUP or methylation), which label miRNAs for degradation or protection against exonucleolytic activity, depending on the specific miRNAs and the tissue. Adapted from (Krol, Loedige et al. 2010).

1.5.1 miRNAs and translational repression

It is well established that both the 5'-cap structure and the poly(A) tail of mRNA promote translation. In the cytoplasm, the cap structure is recognized by the eIF4F complex, which consists of the cytoplasmic cap-binding protein eIF4E, the scaffolding protein eIF4G and the

RNA helicase eIF4A. The poly(A) tail is bound by the cytoplasmic poly(A) binding protein PABPC. PABPC interacts with eIF4G and brings the two ends of the mRNA in close proximity (Figure 1.10). This circular mRNA could be efficiently translated and protected from degradation (Derry, Yanagiya et al. 2006). There is evidence to suggest that miRNAs interfere with the function of the eIF4F complex and PABPC during translation and/or mRNA stabilization.

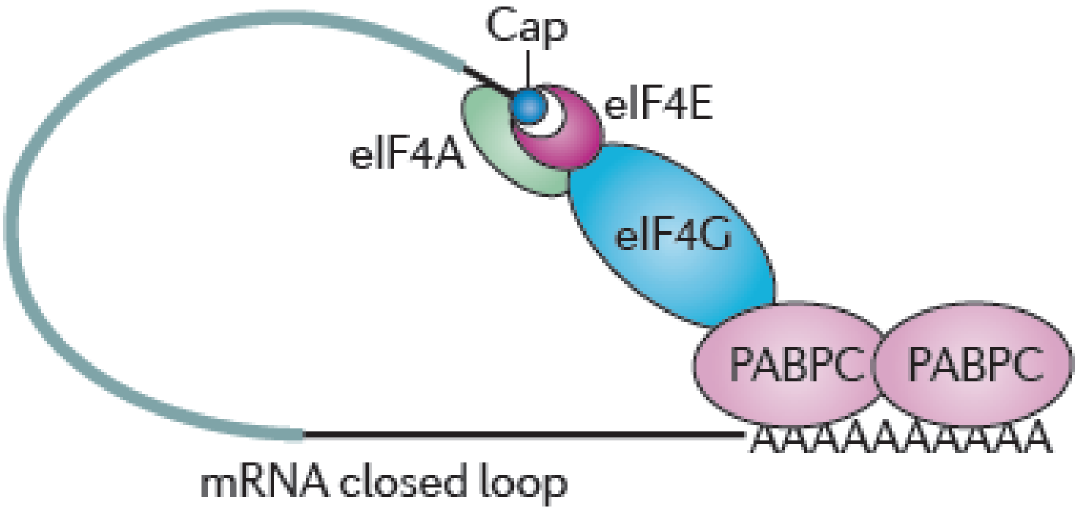


Figure 1.10 Both the 5'-cap structure and the poly(A) tail of mRNA promote translation. In the cytoplasm, the cap structure is recognized by the eIF4F complex, which consists of the cytoplasmic cap-binding protein eIF4E, the scaffolding protein eIF4G and the RNA helicase eIF4A. The poly(A) tail is bound by the cytoplasmic poly(A) binding protein PABPC. PABPC interacts with eIF4G and brings the two ends of the mRNA in close proximity. Adapted from (Huntzinger and Izaurralde 2011).

Evidence for repression at the post-initiation stage

Early studies in *C.elegans* showed that *lin-14* and *lin-28* mRNAs, which are targets of *lin-4* miRNA, are stably associated with polysomes despite a strong reduction in their protein products at a specific stage of larval development (Olsen and Ambros 1999; Seggerson, Tang et al. 2002). The subsequent studies in mammalian cell cultures present a common observation: in sucrose sedimentation gradients, miRNAs and their targets are associated with polysomes (Maroney, Yu et al. 2006; Nottrott, Simard et al. 2006; Petersen, Bordeleau et al. 2006). The sedimentation of miRNAs is shown to be sensitive to a variety of conditions that affected the initation of protein synthesis, indicating that the miRNAs are associated with actively translating mRNAs. All above studies present evidence that miRNAs repress protein synthesis after translation is initiated. To explain these findings, Nottrott *et al.* (Nottrott, Simard et al. 2006) proposed that the nascent polypeptide chain might be degraded cotranslationally. Whereas, Petersen *et al.* (Petersen, Bordeleau et al. 2006) suggested that miRNAs could cause ribosomes to dissociate prematurely (ribosome drop-off).

Evidence for repression at initiation

Other groups have presented evidence that miRNA inhibits translation at translation initiation. For example, the investigations revealed that the translation of m^{7}G-capped mRNAs, but not of mRNAs containing an IRES or a non-functional ApppN cap, is repressed by miRNAs (Humphreys, Westman et al. 2005; Pillai, Bhattacharyya et al. 2005). Polysome gradient analysis independently supports the effect on the initiation step: in the presence of cognate miRNAs, mRNA targets do not co-sediment with the polysomal fraction in sucrose gradients, but shift toward the top of the gradient containing fewer ribosomes or free messenger ribonucleoproteins (mRNPs) (Pillai, Bhattacharyya et al. 2005).

Furthermore, Kiriakidou *et al.* (Kiriakidou, Tan et al. 2007) reported that the central domain of Argonaute proteins contains sequence similar to the cytoplasmic cap-binding protein eIF4E, which is crucial for cap binding. They proposed that Ago2 represses the initiation of mRNA translation by competing with eIF4E for binding the m^{7}G cap of mRNA targets.

Evidence suggesting that miRNAs inhibit an early translation step was also reported by Chendrimada *et al.* (Chendrimada, Finn et al. 2007). The eIF6 and 60S ribosomal subunit proteins could be coimmunoprecipitated with the Ago2-Dicer-TRBP complex. Using human cells they showed that partial depletion of eIF6 rescues mRNA targets from miRNA inhibition. This suggests that Ago2 could recruit eIF6, then the large and small ribosomal subunits might not be able to associate, causing translational repression.

1.5.2 miRNAs and targets degradation

More recent work has demonstrated that the repression of many miRNA targets is frequently associated with their degradation. For example, it has been shown that ectopically overexpressed miRNAs inhibit the abundance of target transcripts (Lim, Lau et al. 2005), and if a miRNA is depleted, the transcripts containing binding sites for this miRNA become more abundant (Krutzfeldt, Rajewsky et al. 2005).

Recently, Huntzinger and Izaurralde proposed a stepwise model for miRNA-mediated target silencing (Huntzinger and Izaurralde 2011). In this model, the AGO interacts with GW182 (trinucleotide-repeat-containing protein), which, in turn, interacts with PABPC bound to the mRNA poly(A) tail. This AGO-GW182 complex directs the mRNA to deadenylation by the CAF1-CCR4-NOT deadenylase complex. The detailed mechanism remains to be determined. Depending on the cell type and/or specific target, deadenylated mRNAs can be stored in a translationally repressed state. However, in animal cell cultures, deadenylated mRNAs are generally decapped and rapidly degraded by the major 5'-to-3' exonuclease XRN1 (Figure 1.11).

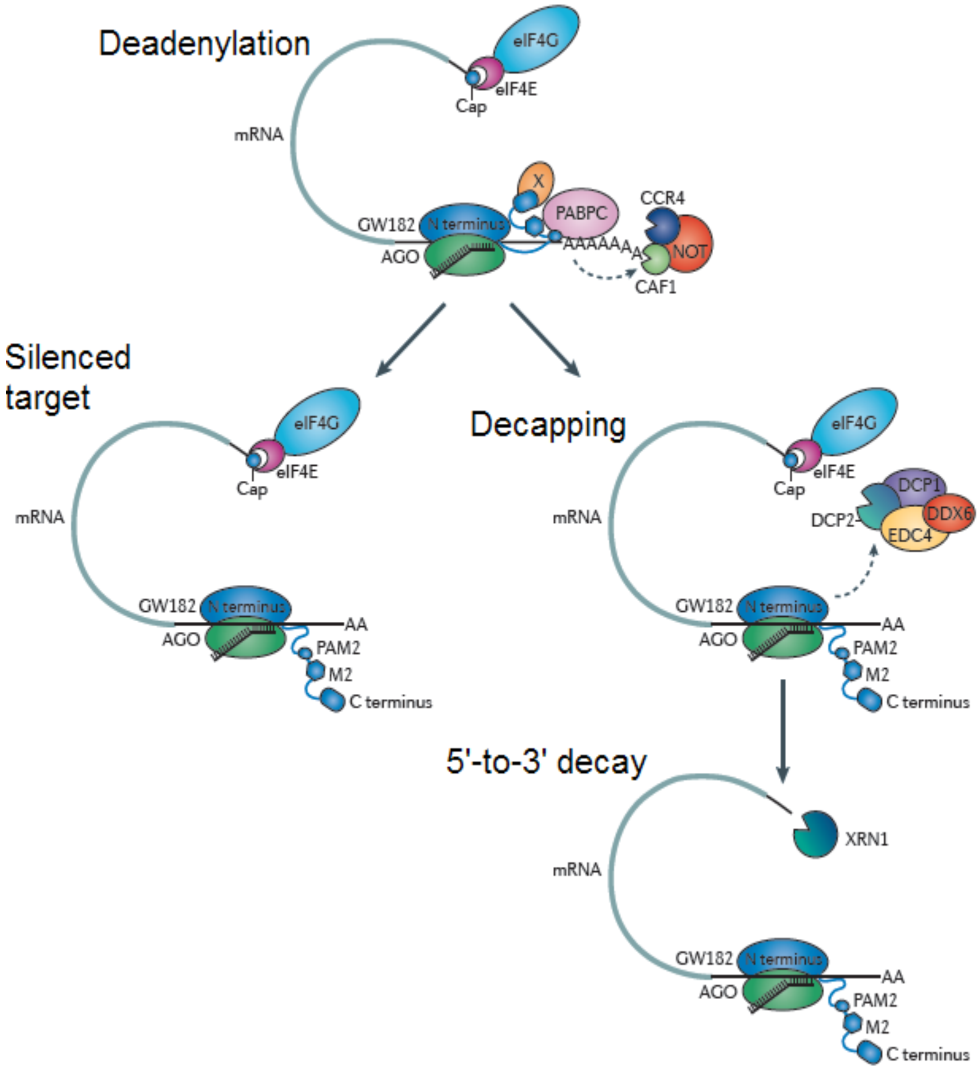


Figure 1.11 The model for miRNA-mediated target silencing. The AGO-GW182 complex directs the mRNA to deadenylation by the CAF1-CCR4-NOT deadenylase complex. Depending on the cell type and/or specific target, deadenylated mRNAs can remain in a translationally repressed state. However, in animal cell cultures, deadenylated mRNAs are

generally decapped and rapidly degraded majorly by the 5’-to-3’ exonuclease XRN1. Adapted from (Huntzinger and Izaurralde 2011).

1.6 Experimental and computational methods for miRNA discovery and expression profiling

The early discovery of miRNA genes was achieved by Sanger sequencing of cDNAs cloned from small RNAs and thereafter validated by Northern blotting (Lagos-Quintana, Rauhut et al. 2001; Lau, Lim et al. 2001; Lee and Ambros 2001; Aravin, Lagos-Quintana et al. 2003; Chen, Manninga et al. 2005; Berezikov, Cuppen et al. 2006; Ruby, Jan et al. 2006). Due to the limitation of affordable sequencing depth, only abundant miRNAs can be efficiently identified while those with low expression or present only in a certain development stage or specific cell populations were difficult to detect. This problem can be overcome, at least in part, by computational approaches with furthur experimental validation (Berezikov, Cuppen et al. 2006). For example, miRNA discovery tools have been developed based on the conservation of miRNAs and their target sequences (Adai, Johnson et al. 2005; Chan, Elemento et al. 2005; Xie, Lu et al. 2005). Briefly, short conserved motifs in the 3’ UTRs of protein-coding genes are first selected. Then, genomic searches for conserved sequences complementary to these motifs are used to find potential miRNA genes. When the matches are found, the flanking regions are examined for their ability to form stable hairpin structures. A disadvantage of such conservation-based strategy is the inability to identify lineage-specific miRNAs.

Recently, with the availability of next-generation sequencing technology, millions of small RNAs can be sequenced in parallel. This allows the detecting and profiling of known and novel miRNAs at unprecedented level of sensitivity. In 2008, the Rajewsky group developed miRDeep, the first publicly available software package for the discovery of miRNAs using deep-sequencing data (Friedlander, Chen et al. 2008). The general strategy is to search genomic sequences for the evidence of hairpin structures, and then determine if sequencing reads aligned to these mimic pre-miRNAs. By using Bayesian statistics, the sequenced RNAs are scored to determine whether they fit to the biological model of miRNA biogenesis. Till now, there are several other public tools for miRNA analysis from deep-sequencing data based on genome-reference information. For example, miRanalyzer is a web server tool that can classify miRNA transcripts from non-miRNA transcripts by using a support vector machine (SVM) trained on miRNA features (Hackenberg, Sturm et al. 2009). miRTrap identifies gene loci where many sequenced RNAs map to a few defined positions (Hendrix, Levine et al. 2010). As our knowledge of miRNA biogenesis improves and sequencing power increases, it is likely that more miRNAs could be continually annotated, particularly in poorly annotated genomes.

To determine the expression level of identified mIRNAs, several methods have been developed, such as miRNA microarrays (Miska, Alvarez-Saavedra et al. 2004; Nelson, Baldwin et al. 2004; Thomson, Parker et al. 2004; Baskerville and Bartel 2005; Bentwich, Avniel et al. 2005), quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) methods and more recently massive parallel sequencing technology (Calabrese, Seila et al. 2007; Babiarz, Ruby et al. 2008; Kuchenbauer, Morin et al. 2008).

1.7 Aim of this thesis

miRNAs, as the key component of small regulatory RNAs, are derived from distinct hairpin precursors. Compared with numerous investigations on mature miRNAs at genomic level of different organisms, global study on pre-miRNAs is rather limited.

In this project, in order to gain a deeper understanding of mammalian miRNAs, miRNAs and pre-miRNAs derived from ten different tissues of adult mice were sequenced in parallel via next-generation sequencing.

With the sequence information of both, especially the pre-miRNAs which to our knowledge have not before been sequenced in genome-wide manner, several new aspects of processing and modification of known mouse miRNAs were revealed.

Including:

1. Ago2-cleaved pre-miRNAs in mouse.

2. New instances of miRNA editing events.

3. Untemplated nucleotide additions at the 3’ end of both miRNAs and the hairpin precursors.

4. Exclusively 5’ tailed mirtrons in mouse.

Furthermore, we developed a computational pipeline for seaching genuine miRNA genes solely based on the sequencing dataset, without using genome sequences.

We believe our method could be widely used in the study of miRNAs, not only in the organisms whose genome has not yet been sequenced, but also in samples where the genome differs significantly from the reference sequences, such as cancer.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Chemicals for experiments were purchased from the following companies, unless indicated otherwise: Ambion, Gibco, Sigma, Invitrogen, Fermentas, Roth, Roche, Merck, Pierce, Promega, Evrogen, Beckman Coulter, Illumina, Applied Biosystems, Finnzymes.

Enzymes and Kits:

- Agencourt AMPure XP Kit (Beckman Coulter)
- Agilent RNA 6000 Nano Kit (Agilent Technologies)
- Agilent small RNA Kit (Agilent Technologies)
- Dacade marker (Ambion)
- DNA*free* Kit (Ambion)
- Duplex specific nuclease (Evrogen)
- flashPAGE Pre-cast Gels and Buffer Kit (Ambion)
- Lipofectamine RNAiMAX reagent (Invitrogen)
- Phusion High-Fidelity DNA polymerase (Finnzymes)
- QIAquick Gel Extraction Kit (Qiagen)
- Qubit dsDNA HS Assay (Invitrogen)
- SuperScript II reverse transcriptase (Invitrogen)
- SuperScript III reverse transcriptase (Invitrogen)
- SYBR Green PCR Master Mix (Applied Biosystems)
- T4 Polynucleotide Kinase (Fermentas)

- T4 RNA ligase (New England Biolabs)
- T4 RNA ligase2, truncated (New England Biolabs)
- T7 Transcription Kit (Fermentas)
- TaqMan 2 × Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems)
- TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems)
- TruSeq SBS Kit (Illumina)
- TruSeq Small RNA library preparation Kit (Illumina)
- TruSeq SR cluster Kit (Illumina)

2.1.2 Equipments

- ABI PRISM^{TM} 7500 Sequence Detection Systems (Applied Biosystems)
- ABI StepOnePlus^{TM} Real-Time PCR Systems (Applied Biosystems)
- Agilent 2100 Bioanalyzer (Agilent Technologies)
- Cell incubator (Hereaus Instruments)
- Centrifuges: 5417R, 5804R (Eppendorf); Biofuge pico (Heraeus)
- Dounce homogenizer Sonopuls GM70 (Bandelin)
- FLA 7000 imager (GE healthcare)
- flashPAGE fractionator (Ambion)
- Gel electrophoreses and blotting equipment (BioRad)
- Hybridization oven MWG (Biotech)
- Illumina cBot instrument (Illumina)
- Illumina Cluster Station (Illumina)
- Illumina Genome Analyzer IIx (Illumina)

- Illumina HiSeq 2000 (Illumina)
- NanoDrop ND-1000 (NanoDrop Technologies)
- PCR Block (MJ Research)
- pH Meter 537 (WTW)
- Power pac 200/300 (BioRad)
- Qubit Fluorometer (Invitrogen)
- XCell SureLock™ Mini-Cell Electrophoresis System (Invitrogen)
- UV crosslinker Stratalinker™ 2400 (Stratagene)

2.1.3 Software and database

- ABI StepOne™/StepOnePlus™ Software (Applied Biosystems)
- Adobe Acrobat Reader 8.0
- Adobe Photoshop CS3
- BLAST (http://www.ncbi.nlm.nih.gov/BLAST/)
- dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/)
- Ensembl genome browser (http://www.ensembl.org/index.html)
- miRBase (http://www.mirbase.org/)
- NCBI database (http://www.ncbi.nlm.nih.gov/)
- Primer3 program (http://frodo.wi.mit.edu/primer3/)
- RNAfold program (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)
- Randfold software (http://bioinformatics.psb.ugent.be/)
- SOAP sequencing reads alignment tool (http://soap.genomics.org.cn/)
- UCSC genome browser (http://genome.ucsc.edu/)

2.1.4 Cells, mouse strain and *C.elegans*

- Neuro-2a cells (N2a cells)-mouse Neuroblastoma cell line (provided by Dr. Matthew Poy, MDC)
- C57Bl/6J Mouse Line (provided by Dr. Yu Shi, MDC)
- Unsynchronous *C.elegans* sample containing worms of all stages (provided by Nadine Thierfelder from Rajewsky's lab, MDC)

2.2 Experimental methods

2.2.1 N2a cell culture

N2a cell, a cultured mouse neuronblastoma cell, was used for transfection. Cell culture was started from a frozen stock by seeding cells into a 100-mm tissue culture dish. The N2a cells were cultured in Dulbecco‘s modified Eagle‘s medium (DMEM) containing 4.5 g/L D-Glucose, sodium pyruvate, 25 mM HEPES and phenol red (Gibco Invitrogen) supplemented with 10% (v/v) fetal calf serum (Biochrom), 2 mM L-glutamine (Gibco Invitrogen) and 50 µg/ml Penicillin/streptomycin (Gibco Invitrogen). Cells were handled on a HERAsafe clean bench (Heraeus) and cultured in HERAcell CO_2-incubators (Heraeus) at 37°C in humid condition with 5% CO_2 concentration.

N2a cells were subcultured twice a week at a density about 2×10^5 cells per tissue culture dish (150 × 20 mm) or 150-cm^2 flask. To passage, cells were rinsed with PBS and detached from the tissue culture dish by incubation with a 0.25% Trypsin/EDTA solution (Gibco) for 5 min at 37°C. To stop the trypsinization, the same amount of DMEM was added to the detached cells and mixed by gently pipetting up and down for several times. Then cells were transferred into a 15-ml conical tube and pelleted by centrifugation at 1000 rpm for 5 min. After removing the supernatant, the cells were resuspended in 5 ml of fresh medium and seeded. To determine cell concentration, 5 µl of cell solutions were diluted to 1:4 with Trypan Blue 0.4% (Sigma-Aldrich) and the cells were counted using a Hemocytometer (Neubauer counting slide) under microscope.

For long-term storage, cells were pelleted and resuspended in freezing solution composed of 90% FCS and 10% dimethyl sulfoxide (DMSO) (Sigma). 1.5 ml of cells was transferred to 2 ml cryovials (GREINERbio-one). And the vials were pre-cooled at -80°C for 24 hours. Then cells were frozen in liquid nitrogen (-196°C).

2.2.2 Cell transfection with siRNA

For RNAi experiment, N2a cells were transfected with Lipofectamine RNAiMAX (Invitrogen), using the manufacturer’s reverse transfection protocol. In general, the transfection experiment was performed in a 6-well plate. First, 40 nM of RNAi duplex was diluted in 250 µl Opti-MEM® I Medium (Invitrogen) without serum in the well of the tissue culture plate and mixed well. Then, 5 µl of Lipofectamine™ RNAiMAX was added to each well containing the diluted RNAi molecules, and incubated for 20 min at room temperature. In the meanwhile, the cultured cells were diluted in complete growth medium without antibiotics so that 2 ml medium contains appropriate 2 × 10^5 of cells to give 30-50% confluence 24 hours after plating. To each well with RNAi duplex-Lipofectamine™ RNAiMAX complexes, 2 ml of the diluted cells was added. Cells were harvested 3 days after transfection and then total RNA was extracted and used for RT–PCR or deep-sequencing. RNAi duplex targeting the Dicer-1 ORF was obtained from Ambion (Ambion s101208). The RNA transfection control (BLOCK-iT™ Fluorescent Oligo) was obtained from Invitrogen.

2.2.3 Mouse tissues collection

Mouse tissues from cerebellum, cortex, heart, kidney, liver, lung, ovary, skeletal muscle, spleen and testes were collected from adult C57Bl/6J Mouse with the help from Dr. Yu Shi and Haitao Wang. The tissues were frozen in liquid nitrogen once they were transferred into the nuclease-free tubes (Eppendorf).

2.2.4 RNA isolation

Total RNA was isolated from mouse tissues, *C.elegans* and N2a cells with/without siDicer treatment using TRIZOL reagent (Invitrogen) in accordance with the manufacturer's instructions. Briefly, cells were directly lysed by adding 1 ml of TRIZOL reagent per well of 6-well plate. Mouse tissue samples were homogenized in 1 ml of TRIZOL reagent per 100 mg of tissue. Following homogenization, an additional isolation step was performed for liver and skeletal muscle, which with high content of proteins and fat. The insoluble materials were removed from the homogenate by centrifugation at 12,000 × g for 10 min at 4°C. Then after adding 0.2 ml of chloroform, the lysate was shaken by hand for 15 sec, incubated at room temperature for 2-3 min and centrifuged at 12,000 × g for 15 min at 4°C to separate the phases. The upper, aqueous phase containing RNA was collected in a new tube. RNA was precipitated by adding 0.5 ml of isopropyl alcohol per 1 ml of TRIZOL reagent and followed by centrifugation at 12,000 × g for 15 min at 4°C. Finally, pellets were washed with 70% ethanol and dissolved in nuclease-free water. Contaminating genomic DNA was removed from total RNA by DNaseI treatment using DNA*free* (Ambion) following the manufacturer's instructions. RNA concentrations were measured using NanoDrop ND-1000 (NanoDrop Technologies). The integrity of isolated RNA was assessed with the Agilent RNA Nano 6000 kit in combination with the Agilent 2100 Bioanalyzer (Agilent Technologies), using the Eukaryote Total RNA Nano assay according to the manufacturer's instructions. The total

RNAs from 10 mouse tissues were pooled in equal amount for pre-miRNA sequencing (see below).

Small RNA fraction with a size range of 10-40 nt from 10 mouse tissues, *C.elegans* and mouse N2a cells as well as small RNAs of size range of 50-100 nt from the mixture of 10 mouse tissues and *C.elegans* were separated from total RNA by using flashPAGE Fractionator (Ambion) (Figure 2.1) according to the manufacturer's instructions. In brief, 10 µg of total RNA for each sample was mixed with equal volumes of flashPAGE Gel Loading Buffer A40 and denatured at 95°C for 2 min. Then the sample was loaded onto the upper surface of flashPAGE Pre-cast Gel which contained 250 µl of flashPAGE Upper Running Buffer. After running the gel for ~12 min at 70V constant voltage until the blue dye began to exit the gel, the lower running buffer which contained the small RNAs $\leq$ 40 nt was collected. The lower running buffer was replaced and the gel running continued for 3 min at 75V constant voltage to collect the small RNA with size range of 40-50 nt and discard it. Then, by replacing the lower running buffer with the fresh one and continuing the addition of gel running at 75V for 15 min, the small RNAs with size range of 50-100 nt were collected. Finally, the small RNAs were precipitated overnight with sodium acetate and ethanol. The integrity of isolated small RNA was assessed with the Agilent small RNA kit in combination with the Agilent 2100 Bioanalyzer, using the small RNA assay according to the manufacturer's instructions.

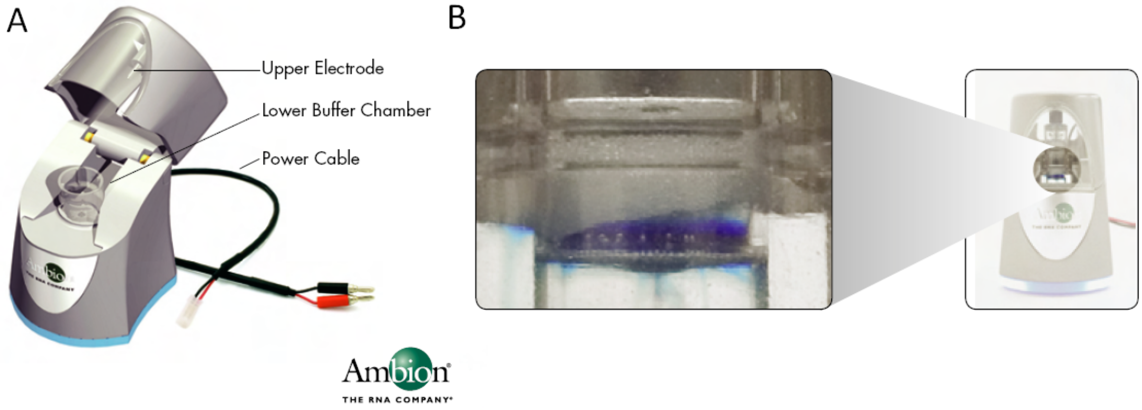


Figure 2.1 Small RNA isolation with flashPAGE Fractionator. A. The equipment of flashPAGE Fractionator. B. The blue dye in the loading buffer migrates with the 40 nucleic acids. Adapted from Ambion website (http://www.ambion.com/).

2.2.5 Small RNA sequencing library preparation

Apart from the initial purification of small RNA fractions as described above, small RNA sequencing libraries were prepared using Illumina small RNA library preparation kits (Figure 2.2). The general workflow is shown in Figure 2.3.

First of all, the purified small RNA fractions were ligated at the 3' end with synthetic 3' RNA adapter by incubation at 22°C for 1 hour using T4 RNA ligase 2 truncated (New England Biolabs). After that, 3' adapter ligated small RNA fractions were subsequently linked at the 5' end with 5' RNA adapter by incubation at 20°C for 1 hour using T4 RNA ligase (New England Biolabs). Then, the RT-PCR amplification was performed to the adapter ligated small RNA fractions using Illumina small RNA RT and PCR primers under optimal conditions. For small RNA fractions with the size range from 10-40 nt, reverse transcription was performed at 44°C for 1 hour using SuperScript II reverse transcriptase (Invitrogen) and subsequently twelve cycles (98°C for 10 sec, 60°C for 30 sec, and 72°C for 15 sec) of PCR amplification were performed using Phusion DNA polymerase (Finnzymes). For small RNA fraction with the size range from 50-100 nt, the first strand cDNA synthesis was performed at 65°C for 50 min using SuperScript III reverse transcriptase (Invitrogen) and subsequently fifteen cycles (95°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec) of PCR amplification were performed using Phusion DNA polymerase (Finnzymes). Finally, the amplified libraries were subsequently purified by PAGE according to the expected product size. The library concentration was determined with Qubit dsDNA HS assay on Qubit Fluorometer (Invitrogen) according to the manufacturer's instructions. The size distribution was assessed with the Agilent DNA 1000 kit in combination with the Agilent 2100 Bioanalyzer, using the DNA assay according to the manufacturer's instructions.

The oligo sequences for small RNA library preparation:

3' RNA adapter:

5'-/5rApp/ATCTCGTATGCCGTCTTCTGCTTG/3ddC/

5' RNA adapter:

5'-GUUCAGAGUUCUACAGUCCGACGAUC

RT Primer:

5'-CAAGCAGAAGACGGCATACGA

Small RNA PCR Primer 1:

5'-CAAGCAGAAGACGGCATACGA

Small RNA PCR Primer 2:

5'-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA

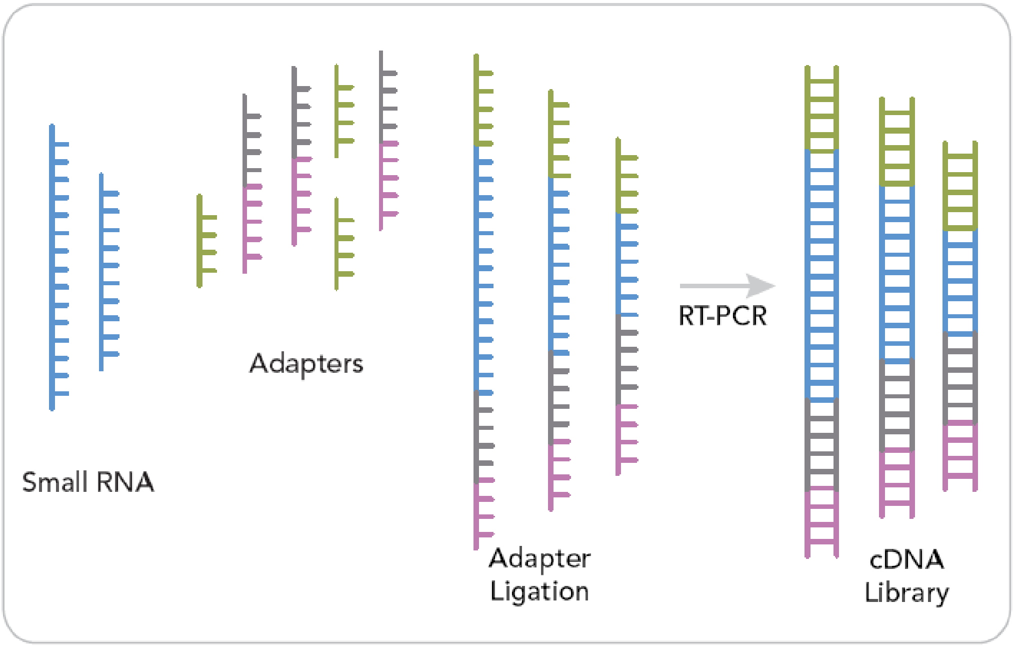


Figure 2.2 Fragments after small RNA sample preparation, adapted from Illumina website (http://www.illumina.com/).

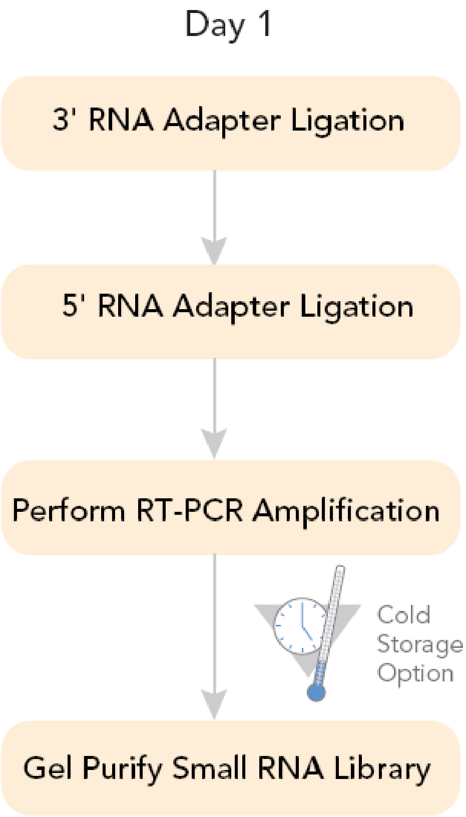


Figure 2.3 Small RNA sample preparation workflow. Adapted from Illumina website (http://www.illumina.com/).

2.2.6 Normalization of small RNA (50-100 nt fraction) sequencing library

The small RNA (50-100 nt fraction) sequencing libraries were normalized by using Duplex specific Nuclease (DSN, Evrogen) according to the manufacturer's instructions. Briefly, an aliquot (100 ng, 4 µl) of amplified small RNA (50-100 nt fraction) library was mixed with 1 µl of 4 × hybridization buffer (200 mM HEPES pH 7.5, 2 M NaCl), overlaid with mineral oil, denatured at 98°C for 3 min and allowed to renature at 68°C for 5 hours. After 5 hours of incubation, 5 µl of 2 × DSN master buffer (100 mM Tris–HCl pH 8.0, 10 mM $MgCl_2$, and 2 mM dithiothreitol) preheated to 68°C was added to the reaction mixture and then incubated for 10 min. Next, 0.5 units of DSN enzyme were added to the reaction and the incubation were continued for 25 min. DSN was subsequently inactivated by the addition of 10 µl of DSN stop solution (5 mM EDTA). After DSN inactivation, DNA was purified using the SPRI beads (Agencourt AMPure) and eluted in a final volume of 20 µl. An aliquot (5 µl) was used for PCR (95°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec, 12 cycles) with Illumina PCR primers, followed by purification using the SPRI beads. The DNA concentration was determined with Qubit dsDNA HS assay on Qubit Fluorometer according to the manufacturer’s instructions. The length distribution was assessed with the Agilent DNA 1000 kit in combination with the Agilent 2100 Bioanalyzer, using the DNA assay according to the manufacturer's instructions.

2.2.7 Illumina Deep-sequencing

First of all, Adaptor-ligated DNA library was hybridized to the surface of flowcells, and DNA clusters were generated using the Illumina cluster station or cBot instrument (Illumina), in accordance with the Illumina cluster amplification protocols (Figure 2.4). Briefly, 10 nM diluted library was first denatured in 2 N NaOH and subsequently diluted in the hybridization buffer to the final concentration of 8 pM. The denatured templates were hybridized to the nucleotides immobilized on the flowcell surface. The templates were copied from the hybridized primers by 3’ extension using high-fidelity DNA polymerase. Then the original templates were denatured by formamide, leaving the single-stranded templates anchored to the flowcell surface. The DNA template copies were amplified by isothermal bridge amplification for 30 cycles. Resulting with, each DNA template formed a dense clonal cluster containing ~2,000 molecules. Each cluster of dsDNA bridges was linearized with linearization mix. The 3’ ends of the DNA strand and flowcell-bound oligonucleotides were blocked with blocking mix containing ddNTP and terminal transferase. Finally, the sequencing primer was hybridized to the unbound ends of the templates in the clusters. The flowcell was ready for sequencing.

Following cluster generation, Small RNA (10-40 nt fraction) libraries from mouse tissues (cerebellum, cortex, heart, kidney, liver, lung, ovary, skeletal muscle, spleen and testes), *C.elegans* and N2a cells with/without siDicer treatment were sequenced for 36 cycles, each on a separate lane using Illumina Genome Analyzer IIx. Normalized small RNA (50-100 nt) libraries from the mixture of the above 10 mouse tissues and *C.elegans* were sequenced for 100 cycles, each on a seperate lane using Illumina HiSeq 2000, according to the manufacturer's instructions.

Small RNA Sequencing Primer: 5' CGACAGGTTCAGAGTTCTACAGTCCGACGATC

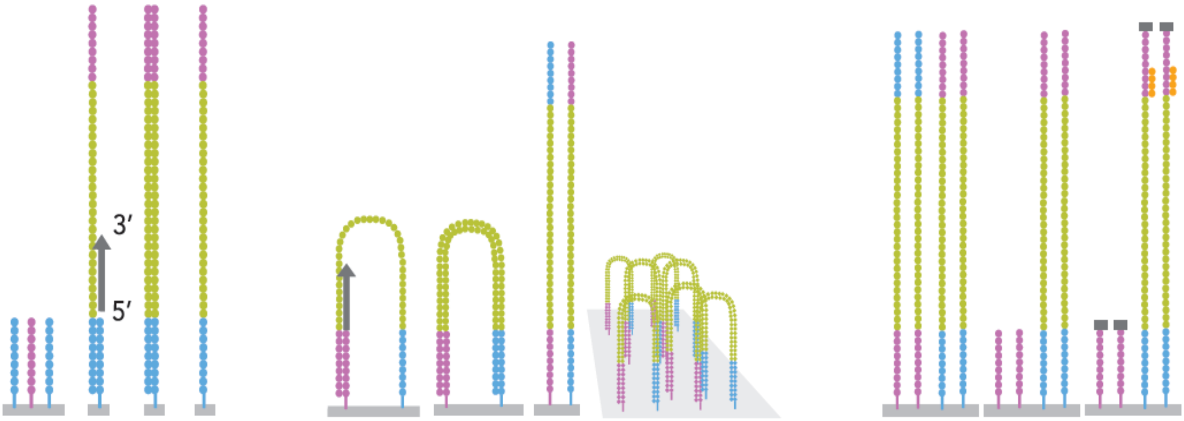


Figure 2.4 Cluster generation by isothermal bridge amplification. Cluster generation from single-molecule DNA templates occurs within the Illumina flowcell on the cluster station or cBot instrument, and involves immobilization and 3' extension, bridge amplification, linearization, and hybridization. Adapted from Illumina website (http://www.illumina.com/).

2.2.8 RT-qPCR

To quantify the expression of Dicer after siRNA knockdown, cDNA was synthesized from total RNA using SuperScript II according to the manufacturer's instructions (Invitrogen). In brief, 100 ng of total RNA, 1 µl of oligo(dT) 18 (500 µg/ml), 1µl of 10 mM dNTP in 12µl volume were heated to 65°C for 5 min and quick chilled on ice. 4 µl of 5 × first-strand buffer, 2 µl of 0.1 M DTT, 1 µl of RNase Inhibitor (Invitrogen) were added and incubated at 42°C for 2 min followed by adding 1 µl (200 units) SuperScript II enzyme. Reactions were incubated at 42°C for 50 min and then deactivated at 70°C for 10 min. Real-time PCR was carried out with SYBER Green PCR Master mix (Applied Biosystems) on the StepOnePlus^{TM} system (Applied Biosystems). All primer pairs were designed using Primer3 program (primer sequence see table 2.1). The PCR mix for each reaction contained 0.5 µl cDNA, 10 µl of 2 × SYBER Green PCR Master mix (Applied Biosystems), and 0.5 µM of each primer in a total volume of 20 µl. Standard reactions were performed using the following cycle parameters: AmpliTaq activation 95°C for 10 min; PCR: denaturation 95°C for 15 sec and annealing/extension 60°C for 1 min (repeated 40 times). All experiments were carried out in triplicate. The expression of Dicer was normalized to endogenous GAPDH mRNA levels using the ΔΔCT method (Livak and Schmittgen 2001).

The following primer pairs were used for quantification of Dicer knockdown:

mGAPDH_F: 5'-AACTTTGGCATTGTGGAAGG

mGAPDH_R: 5'-GGATGCAGGGATGATGTTCT

mDICER1_F: 5'-ACCAAGTGATCCGTTTACGC

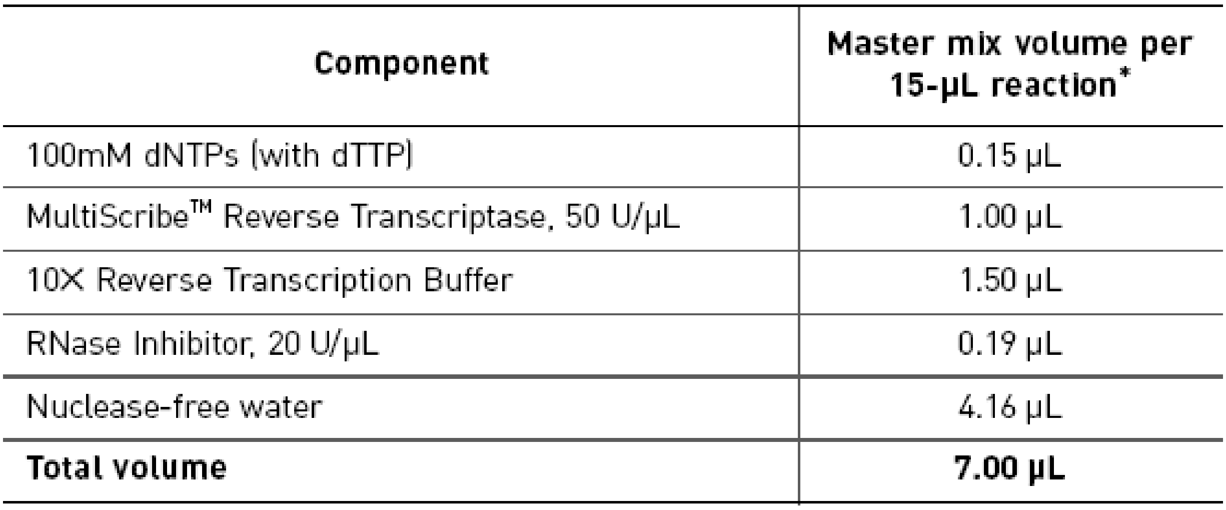
mDICER1_R: 5'-CAACCGTACACTGTCCATCG

To validate the efficiency of DSN normalization for small RNA library, Real-time PCR was carried out with SYBER Green PCR Master mix on the StepOnePlus^{TM} system. Since all the libraries contain the universal primers in both 5’ and 3’ end, one universal primer with the complementary sequences to the 3’ end of the libraries was served as reverse primer. The forward primers were designed as the combination of the sequences from the 5’ end of the libraries and the 5’ end of one of three snoRNAs and three pre-miRNAs (primer sequences, see table 2.1). The PCR mix for each reaction contained 5 µl/10 ng of non-normalized or normalized DNA template and the rests same as described above.

Table 2.1 primer sequences were used to validate the DSN normalization efficiency by Real-time PCR.

Name	Sequence
sno_064457_F	5’-TCCGACGATCGCGGATGATGA
sno_064450_F	5’-TCCGACGATCTGGAATGATGACA
sno_077221_F	5’-TCCGACGATC GAGAGTGAT
pre-mir-142_F	5’-AGTCCGACGATC CATAAAGTAGA
pre-mir-27b_F	5’-GTCCGACGATC AGAGCTTAGCT
pre-mir-19b_F	5’-GTCCGACGATC AGTTTTGCAGG
universal primer_R	5’-CAAGCAGAAGACGGCATACGA

To validate the expression of the two novel miRNAs derived from Snora7a and Snora41 loci, TaqMan® MicroRNA Assays from Applied Biosystems were custom designed and used according to the manufacturer’s instructions. The Reverse transcription was carried with TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems). Ten ng of total RNA was used for each reaction. In brief, 7 µl of the following master mix was mixed with 3 µl of 5 × RT primer, and 5 µl of RNA sample in a total volume of 15 µl. Then the reaction was performed under condition of 16°C 30 min, 42°C 30 min and 85°C 5 min.



Component	Master mix volume per 15-µL reaction*
100mM dNTPs (with dTTP)	0.15 µL
MultiScribe™ Reverse Transcriptase, 50 U/µL	1.00 µL
10X Reverse Transcription Buffer	1.50 µL
RNase Inhibitor, 20 U/µL	0.19 µL
Nuclease-free water	4.16 µL
Total volume	**7.00 µL**

The PCR mix for each reaction contained 1.33 µl of product from RT reaction, 10 µl of TaqMan 2 × Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems), and 1µl of 20 × TaqMan small RNA assay in a total volume of 20 µl. Standard reactions were performed using the following cycle parameters: AmpliTaq activation 95°C for 10 min; PCR: denaturation 95°C for 15 sec and annealing/extension 60°C for 1 min (repeated 40 times). All experiments were carried out in triplicate.

Target Sequence,

Snora7a: 5'-UGAUUGGAAGACACUCUGCAAC

Snora41: 5'-GGAGGAUUAUGUGUGACAGACA

2.2.9 Northern blotting

LNA probes (Exiqon) were labelled with [γ-^{32}P] ATP using T4 PNK (Fermantas), followed by purification with MicroSpin G-25 columns (Amersham Biosciences). 50 µg of total RNA prepared from mouse tissues was separated on 15% denaturing PAGE gel and transferred to a

Hybond-N+ membrane (Amersham Pharmacia) by semidry electroblotting. Then the membrane was UV cross-linked under the condition of 1200 J, ca. 30 sec, twice, on Stratalinker UV crosslinker (Stratagenes). After that, the membrane was pre-hybridized in hybridization buffer (5 × SSC, 20 mM Na_2HPO_4 (pH 7), 1% SDS, 100 µg/ml salmon sperm DNA, 1 × Denhardt's solution) for 1 hour, and subsequently hybridized overnight with hybridization buffer containing 30 pmol of [γ-^{32}P] ATP end labeled LNA probes at 20°-22°C below the calculated probe Tm. After hybridization, the membrane was washed twice 10 min with 5 × SSC, 1% SDS and once 10 min with 1 × SSC, 1% SDS at hybridization temperature. All radioisotopic images were recorded using phosphorimaging screen on FLA 7000 imager (GE healthcare).

The sequence of the LNA probes:

Snora41: 5'-TGTCTGTCACACATAATCCT

Snora7a: 5'-TTGCAGAGTGTCTTCCAATCA

2.2.10 *in vitro* miRNA processing

RNAs of size 140 nt and 70 nt, corresponding to the full length snoRNAs and the pre-miRNA candidates based on our prediction, were obtained by *in vitro* transcription, in which the templates were amplified from mouse genomic DNA using specific primer pairs linked with T7 promoter sequences (primer sequences see table 2.2) and purified from a 2.5% agarose gel using the QIAquick Gel Extraction kit (Qiagen). One µg of each template was transcribed by T7 RNA polymerase (Promega) at 37°C for 2 hours in the presence of [α-^{32}P] UTP (25 µCi at concentration of 10 µCi/µl). *In vitro* transcribed RNAs were purified by phenol/chloroform extraction and precipitated with 100% isopropanol in the presence of 0.3 M ammonium acetate. Pellets were dissolved in RNase free water at the concentration of 20.000 cpm/µl. For the processing assay, 20.000 cpm RNA was incubated with recombinant Dicer (Invitrogen) (0.1 U/reaction) in 75 mM NaCl, 20 mM Tris-HCl, pH 7.5, 3 mM $MgCl_2$ and 0.1 U/µl RNAse inhibitor at 37°C for 15, 30, 60 and 120 min. Processing products were resolved at 10% denaturing PAGE gel, which was then exposed for 1 hour to a phosphorimaging screen and visualized on FLA 7000 imager (GE healthcare).

Table 2.2 Primer sequences for *in vitro* transcription.

Name	Sequence
T7-mmu-SNORA41-full-F	5´-GAAATTAATACGACTCACTATATTCCACAGCTACTGGTCT
T7-mmu-SNORA41-full-R	5'- TTGTGTCTGTCACACATAATCCT
T7-mmu-SNORA41-candiA-F	5´-GAAATTAATACGACTCACTATAACTGTTACACAATTTAATGC
T7-mmu-SNORA41-candiA-R	5'- CCTTGTGTCTGTCACACAT
T7-mmu-SNORA7-full-F	5´-GAAATTAATACGACTCACTATAGACCTCTTGGGATCGCGT
T7-mmu-SNORA7-full-R	5'- TAATGTTGCAGAGTGTCTTC
T7-mmu-SNORA7-candiA-F	5´-GAAATTAATACGACTCACTATACTAGCAGAGGTACCCATTC
T7-mmu-SNORA7-candiA-R	5'- TAATGTTGCAGAGTGTCTTC

2.3 Data analysis methods

2.3.1 Small RNA sequence reads mapping

First, 3' adapter sequences were removed from the sequencing reads using an in-house Perl script. The reads of length between 17 and 30 nt from small RNA 10-40 nt fraction were retained and mapped to genome reference sequences (UCSC genome browser mm9) and known mouse pre-miRNA sequences deposited in miRBase (v16.0) (Kozomara and Griffiths-Jones 2011) using soap1 and soap.short (Li, Li et al. 2008), respectively, and no mismatch allowed. To be considered as a known miRNA, the 5' and 3' ends of sequencing read should be within 1 nt and 3 nt from the 5' and 3' ends of the miRNA in miRBase v16.0, respectively. The 5' or 3' ends of 13 miRNAs in miRBase v16.0 were manually corrected because: 1) 5' ends of >= 90% of our sequencing reads mapped to the miRNA loci were at least 2 nt away from the annotated 5' end. 2) The corrected annotation of 5' or 3' ends can fit better with characteristics of miRNA biogenesis (Appendix table 1). From the small RNA 50-100 nt fraction, the sequencing reads of length between 40 and 94 nt were retained. After removing the last 5 nt at the 3' end, we aligned them to mouse genome reference sequences (UCSC genome browser mm9) using soap2 (Li, Yu et al. 2009). To determine the reads derived from full-length pre-miRNAs, we mapped the first 40 nt to the mouse pre-miRNAs deposited in miRBase (v16.0) allowing 2 mismatches using soap2 and then further extended the alignment to the 3’ end. The 5’ and 3’ ends of pre-miRNAs in miRBase were manually annotated based on the secondary structure if miRNA* has not been identified. Reads to be considered as full-length pre-miRNAs should satisfy the following criteria: 1) The 5’ and 3’ end of the alignment should be within 2 nt and 5 nt from 5’ and 3’ end of the reference pre-miRNA, respectively. 2) No more than five mismatches were found in the alignment.

In order to predict miRNAs based on the sequencing reads obtained from the two small RNA fractions corresponding to potential mature miRNAs and pre-miRNAs, we mapped the sequencing reads of length between 17 and 30 nt on the sequencing reads of length between 40 and 94 nt using soap.short allowing no mismatch.

2.3.2 Identification of tissue-specific miRNA

To quantify the specificity of the expression of mature miRNAs in 10 mouse tissues, we adopted the analysis of miRNA Tissue Specificity method from (Landgraf, Rusu et al. 2007). First of all, we normalized the miRNA reads within each samples and calculated a “normalized tissue enrichment” for each miRNA as the ratio of the normalized reads of one tissue to the sum of all tissues. Then, we calculated the information content of the “normalized tissue enrichment” distribution across tissues.

We called Em,t the number of miRNA reads m in the tissue type t. To allow for comparison between tissue types, we first normalized the counts in each tissue type t:

$F_{m,t} = E_{m,t} / \sum_{m'} E_{m',t}$

If a miRNA m is expressed with high specificity in tissue type t, then the value $F_{m,t}$ is large (close to 1) not only relative to other miRNAs m' ≠ m in the same tissue type t, but also relative to the same miRNA m in other tissue types t' ≠ t. Then “normalized tissue enrichment” for each miRNA as the ratio of the normalized reads of one tissue to the sum of all tissues as:

$G_{m,t} = F_{m,t} / \sum_{t'} F_{m,t'}$

Each line m of G contains a distribution of normalized frequencies of miRNA m across samples.

When $G_{m,t*}$ is close to 1 in tissue t*, we may infer that miRNA m is specifically expressed in tissue type t*, meaning that a large fraction of clones in t* and much smaller fractions of clones in other t ≠ t* correspond to miRNA m. We used the information-theoretic concept of "information content" to quantify how strongly biased the distribution of $G_{m,t}$ is for a given m as the specificity score:

s_m = log2(number of tissue types) + $\sum_t G_{m,t} \log2(G_{m,t})$

To achieve the sufficient expression, we considered, only miRNAs with a normalized total read number ($\sum_{t'} F_{m,t'}$) above 50.

2.3.3 Identification of Ago2-cleaved pre-miRNA (ac-pre-miRNA)

To identify potential ac-pre-miRNAs, we mapped the first 40 nt of the 100 nt reads after adapter trimming to the mouse pre-miRNAs deposited in miRBase (v16.0) allowing 2 mismatches using soap2 and then further extended the alignment to the 3' end. Following filters were applied to extract the reads derived from potential ac-pre-miRNAs: 1) Compared with annotated ends of pre-miRNAs, one end of the alignment is within a distance of 2 nt, the other end is truncated by 9-12 nt. 2) The truncated part should consist of 8-10 nt that can form base pairs with the nucleotides on the other arm. 3) No bulge locates within 4 nt from the potential cleavage site. 4) Ac-pre-miRNA candidates should be supported by at least two reads.

2.3.4 Analysis of untemplated nucleotide addition

To examine the untemplated nucleotide addition at 3' end of mature miRNA, we searched for the non-genome-mapping reads from small RNA (10-40 nt) libraries that can be mapped to the known miRNA genomic loci with poly (N) mismatches at 3‘ end. The 5' end of reads must be at most 1 nt away from the 5' end of miRNAs. The frequency of untemplated addition of specific nucleotide was calculated as the ratio of number of reads containing untemplated addition of that nucleotide to the sum of all reads derived from the same known miRNA locus. Sequencing reads identical to mir-1 apart from the 3' residues was excluded from calculation because such reads corresponded to mir-1b, mir-1c or mir-1d (http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000139). A similar approach was applied to examine the untemplated nucleotide addition of pre-miRNAs. Pre-mir-677 was excluded from the analysis to avoid potential bias because 61.7% of pre-miRNA reads were mapped to this pre-miRNA.

2.3.5 Identification of RNA editing events

To examine the miRNA editing events, we mapped the non-genome-mapping reads from small RNA (10-40 nt) libraries to reference miRNA sequences with one mismatch. The uniquely mappable reads, which at least 1 nt away from the 3’ end and requiring the 5’ end at most 1 nt away from the 5’ end of known miRNAs, were selected. For each of the mismatches identified in these reads, we calculated the fraction of the mismatch as the number of reads bearing that mismatch divided by the number of all reads containing a mismatch at the same position. We obtained a set of highly confident A-to-I editing sites by searching for A-G changes that could pass the following filters: 1) The fraction was higher than 90%. 2) The change was found in at least 10 reads. 3) The same change was found in at least one pre-miRNA reads, and the sequencing quality score of that base was higher than 30. 4) The change was not annotated as a SNP in dbSNP (build 128).

2.3.6 Extraction of potential pre-miRNA sequences

We selected the 40-94 nt (long) reads as potential pre-miRNAs on which the mapping pattern of 17-30 nt (short) reads were compatible with Dicer processing in the following four steps.

1) On one long read, a cluster of mapped short reads was defined as all short reads with overlapped mapping positions and the maximal distance between the start position of any two reads within one cluster did not exceed 14 nt. If the long and short reads were originated from genuine precursor and mature miRNAs, the short reads should form less than three clusters at the 5’ end, 3’ end and the middle of the long read, corresponding to the miRNA/miRNA* and the loop, respectively. Also, we would expect that the 5’ and 3’ end clusters contain much more short reads than the middle cluster. Finally, given the length distribution of canonical mature miRNAs, the majority of short reads from 5’ and 3’ end clusters should be of length between 17 and 25 nt. Therefore, based on these rules, we discarded the long reads if they had any of the following mapped pattern:

a) Number of clusters > 3.

b) Minimal distance between any two reads in different clusters <= 5 nt.

c) Number of reads in the middle cluster exceeds that in the 5’ and 3’ end cluster.

d) Less than 66% of distinct/non-redundant mapped short reads are of length between 17 nt and 25 nt, or less than 90% of all mapped short reads are of length between 17 nt and 25 nt.

2) After filtering out the obvious non-Dicer compatible reads, we further selected the potential pre-miRNA reads. For each remaining long read, we first identified the most abundant distinct/non-redundant short reads from the 5' and 3' end clusters. The long reads were retained only if the most abundant short reads start or end less than 5 nt away from the 5' or 3' end of the long read respectively. We then counted the number of short reads that start <= 3 nt away from the 5' end of the most abundant reads in the 5' and 3' end clusters. The term "Sharpness" denoted the percentage of these reads out of all short reads mapped on the same long read. Because most short reads that mapped to a genuine pre-miRNA should origin from miRNA/miRNA*, we selected long reads with a sharpness value above the threshold of 0.75. Then they were clustered if they contained the same most abundant short reads, differed less than 5 nt in length and had a sequence similarity above 90%. One representative read with the highest abundance from each cluster was selected.

3) We predicted the secondary structures of the selected long reads using RNAfold (parameters: -p –d 2 –noLP) (Hofacker and Stadler 2006) and randfold (parameter: -d 199) (Bonnet, Wuyts et al. 2004) , respectively. Only the long reads which can fold into unbifurcated hairpin structures were retained.

4) The remaining long reads satisfying the following criteria were selected as potential pre-miRNA candidates. The rests were used as "background" in the probabilistic scoring of potential pre-miRNA candidates.

a) The randfold p-value was smaller than 0.2.

b) More than 60% of the nucleotides in the "mature" part (the most abundant distinct/non-redundant short reads from 5' or 3' end clusters) were base paired.

2.3.7 Probabilistic scoring of potential pre-miRNA candidates

We scored the potential pre-miRNA candidates using a Naïve Bayesian classifier with six features:

f1: Minimal folding free energy calculated by RNAfold divided by the sequence length

f2: Randfold p-value

f3: Number of unpaired nucleotides at 5' end

f4: Length of 3' overhang (number of unpaired nucleotides at 3' end minus that at 5' end) f5: Average length of the most abundant distinct/non-redundant short reads from the 5’ and 3’ end cluster that correspond to potential miRNA/miRNA*

f6: Length of candidate pre-miRNA

The “positive training dataset” was pre-miRNAs from miRBase (v 16.0). We calculated the probability of a given potential pre-miRNA candidate to be a genuine pre-miRNA using the following formula:

Pr(pre|data) = P(data|pre)*P(pre)/[P(data|pre)*P(pre) + P(data|non)*P(non)]

where P(data|pre) = P(f1|pre)*P(f2|pre)*P(f3|pre)*P(f4|pre)*P(f5|pre)*P(f6|pre)

and P(data|non) = P(f1|non)*P(f2|non)*P(f3|non)*P(f4|non)*P(f5|non)*P(f6|non)

P(pre) is the prior probability that a long read is actually a miRNA precursor.

P(non) is the prior probability that a long read is non-miRNA background stem-loop and equal to 1-P(pre). Both P(pre) and P(non) are set to 0.5 by default, but can be changed based on the expected pre-miRNA sequences in the deep-sequencing samples.

P(f1|pre) to P(f6|pre) is the probability that a miRBase pre-miRNA has the value of f1 to f6.

P(f1|non) to P(f6|non) is the probability that a background stem-loop sequence has the value of f1 to f6.

3 Results

3.1 Global profilling of mouse miRNAs and their precursors

3.1.1 Small RNA library construction

The scheme of working procedures for small RNA library construction is illustrated in Figure 2.3. In brief, the small RNA fractions were ligated sequentially at the 3' and 5' end with synthetic RNA adapter, reverse transcribed and amplified using Illumina PCR primers. Given the difficulty resulted from the stable pre-miRNA secondary structure, we chose to use SuperScript III reverse transcriptase from Invitrogen, which has a high thermostability, in first strand cDNA synthesis for cloning pre-miRNAs. The cDNA libraries resulted from 10-40 nt and 50-100 nt RNA fractions were analyzed by Bioanalyzer (Figure 3.1).

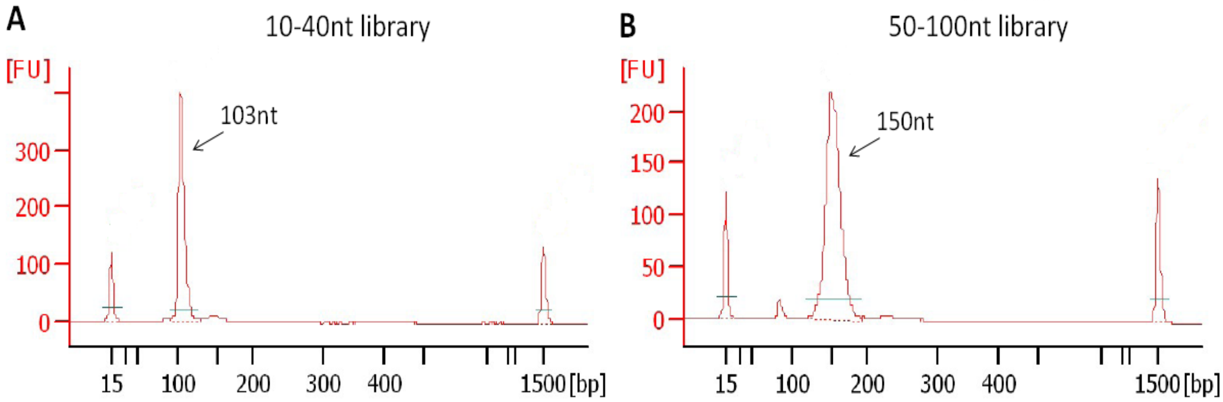


Figure 3.1 1 Small RNA sequencing libraries were analyzed on Bioanalyzer (A) 10-40 nt small RNA librariy was ~103 nt in size. (B) 50-100 nt small RNA library was ~150 nt in size.

3.1.2 Normalization of small RNA (50-100 nt fraction) sequencing library

It is known that 50-100 nt RNA fraction contains a variety of other small RNAs such as C/D box snoRNAs which are much more abundant than pre-miRNAs and have both 5' (monophosphate) and 3' (hydroxyl) ends (Underwood, Uzilov et al. 2010) which are compatible with our small RNA cloning procedure. To reduce the representation of the potentially extremely abundant non-pre-miRNA transcripts, the normalization was performed to the cDNA library before sequencing.

The normalization was based on the use of the Duplex Specific thermostable nuclease (DSN) enzyme, purified from Kamchatka crab hepatopancreas and manufactured by Evrogen (www.evrogen.com). DSN normalization was performed after small RNA sequencing library preparation. The normalization procedure is illustrated in Figure 3.2. Following DNA denaturation and re-association, the double-stranded DNA of repetitive sequences was degraded using DSN, and the intact ssDNA fraction enriched for low-copy sequences was amplified by PCR. As a result, the level of abundant transcipts, which more likely form double strand after re-association, got decreased.

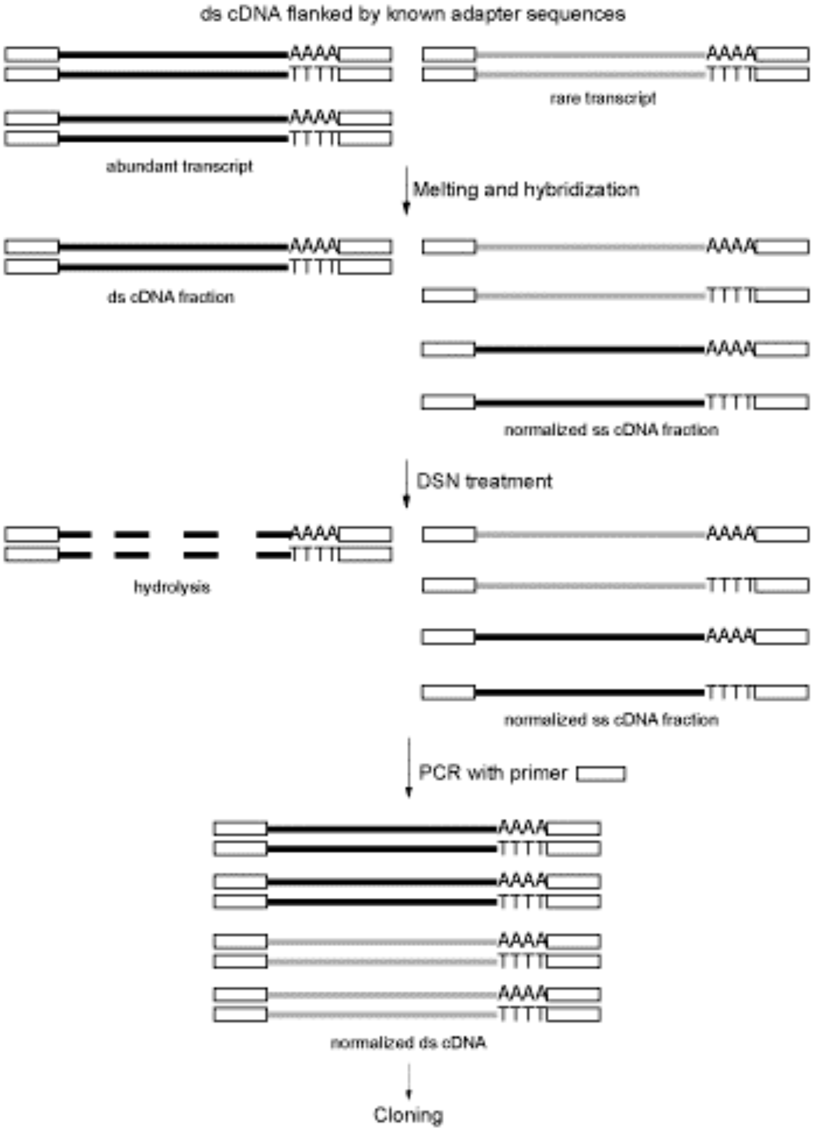

ds cDNA flanked by known adapter sequences
AAAA
TTTT
rare transcript
abundant transcript
Melting and hybridization
ds cDNA fraction
normalized ss cDNA fraction
DSN treatment
hydrolysis
normalized ss cDNA fraction
PCR with primer
normalized ds cDNA
Cloning

Figure 3.2 Schematic outline of DSN-normalization. Black lines represent abundant transcripts, grey line – rare transcripts. Rectangle represents adapter sequence and its complement. Adapted from Evrogen (www.evrogen.com).

To evaluate the normalization efficiency, the quantitative PCR was performed by using the primers targeted to three snoRNAs and three pre-miRNAs, which showed the different expression level according to the preliminary deep-sequencing results. The detailed primer design can be found in the Methods. By DSN normalization, the expression level of the highly abundant snoRNAs, sno_064457 and sno_064450 decreased, whereas, that of the lowly expressed snoRNA as well as three pre-miRNAs was increased upon DSN normalization. The differences in PCR cycle number between the most abundant one and the least abundant one reduced from 16 to 7 PCR cycles (Figure 3.3).

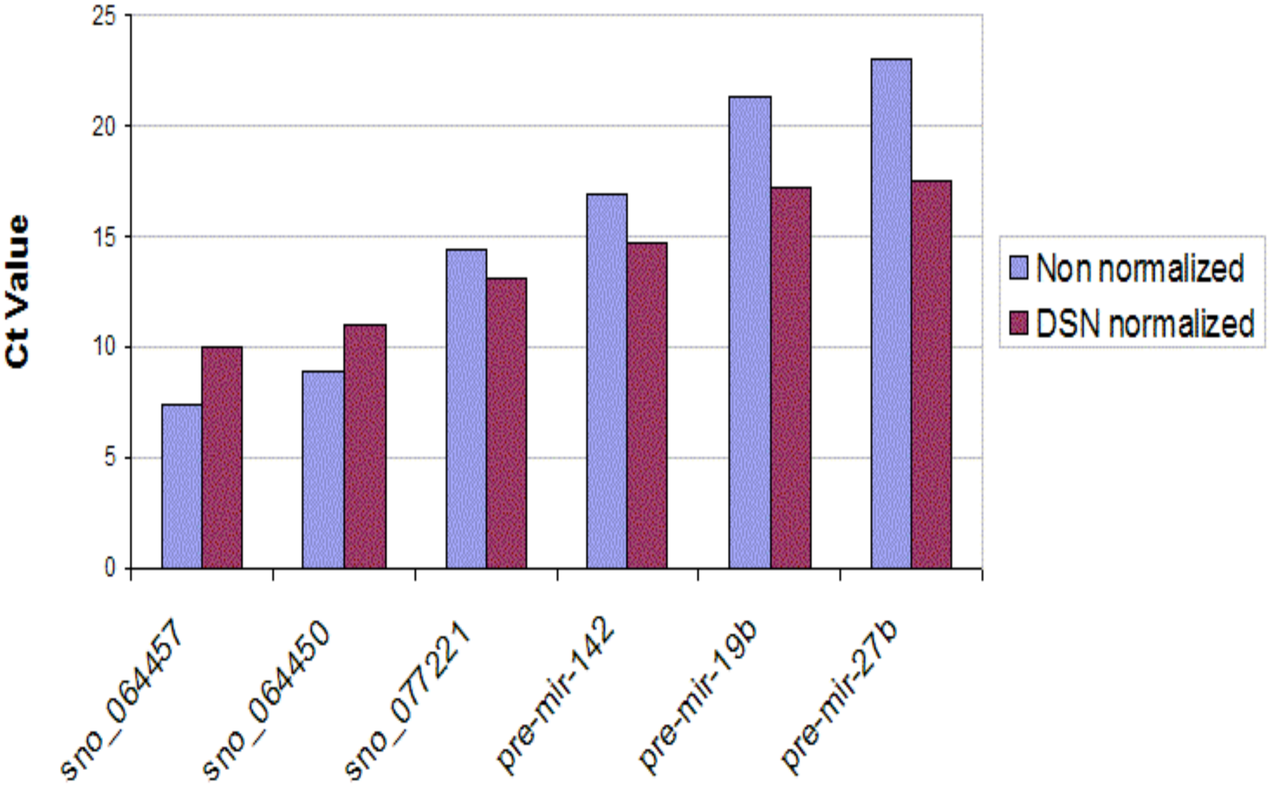


Figure 3.3 The efficiency of DSN normalization was evaluated by quantitative PCR. After DSN normalization, the level of two highly abundant snoRNAs decreased and the expression level of the lowly expressed snoRNA as well as three pre-miRNAs increased. The y-axis is the Ct value obtained from qPCR results, the blue bars stand for non-normalized library, the red bars represent for DSN normalized library. sno_064457, sno_064450 and sno_077221 represent the Ensembl Gene ENSMUSG00000064457, ENSMUSG00000064450 and ENSMUSG00000077221, respectively.

3.1.3 Deep-sequencing of miRNA and pre-miRNA

To profile mouse mature miRNAs, we sequenced small RNA (10-40 nt fraction) libraries from 10 different mouse tissues (cerebellum, cortex, heart, kidney, liver, lung, ovary, skeletal muscle, spleen and testes) and obtained 167 million reads between 17 nt and 30 nt in length ("short reads") (Table 3.1). Of these, 75.2% could be aligned to the mouse genome without any mismatch and 52.8% were derived from known mouse miRNA loci (Figure 3.4) (see methods).

In parallel, we also sequenced pre-miRNAs. To characterize as many pre-miRNAs as possible, the total RNAs from the 10 mouse tissues, the same as above, were pooled and small RNAs between 50 nt and 100 nt in length were extracted for deep-sequencing. To reduce the representation of the potentially extremely abundant non-pre-miRNA transcripts, we normalized the cDNA library before sequencing. With one lane of Illumina HiSeq 2000 sequencing run, we obtained over 57 million reads between 40 nt and 94 nt in length ("long reads"), of which 86.7% could be mapped to the mouse genome. In contrast to mature miRNA sequencing, only 0.80% of the long reads were derived from known pre-miRNA loci whereas the vast majority was from snoRNAs (Figure 3.4).

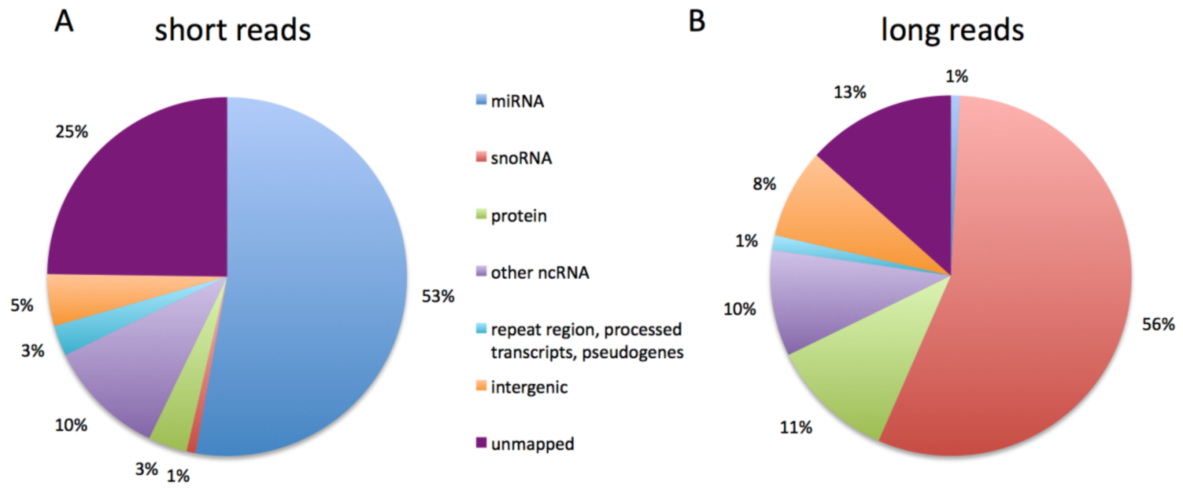


Figure 3.4 Annotation of sequencing reads based on Ensembl Genes (59) (www.biomart.org). 10-40 nt short reads (A) and 50-100 nt long reads (B). "Other ncRNA" includes rRNA, tRNA, scRNA, snRNA and srpRNA.

Table 3.1 Summary of sequencing results.

	short reads (10-40 nt fraction)	long reads (50-100 nt fraction)
No. of total reads	225,382,734	63,504,260
No. of trimmed reads	167,484,979	57,572,046
No. of trimmed reads mapped to mouse genome	125,995,570	49,889,031
No. of reads mapped to known miRNA	88,457,557	462,082
No. of miRNA/pre-miRNA reads	87,369,704	252,003

3.1.4 Identification of known miRNA and pre-miRNA

3.1.4.1 Expression profiling of mouse miRNA from 10 tissues

To be considered as a known miRNA, the 5’ and 3’ ends of sequencing read from 10-40 nt small RNA libraries should be within 1 nt and 3 nt from the 5’ and 3’ ends of the miRNA in miRBase v16.0, respectively. In total, 87,369,704 short reads matched 887 known mouse miRNAs and miRNAs* from 568 pre-miRNAs. 687 miRNA/miRNA* from 481 pre-miRNAs were expressed (defined as RPM >=1; RPM: Reads Per Million total miRNA reads) in at least one tissue (Table 3.1). To compare expression levels of each miRNA in different sequenced samples, we constructed the overall miRNA expression profiles (Figure 3.5). Most miRNAs had substantially stronger expression in some tissues than in others, in agreement with previous observation (Chiang, Schoenfeld et al. 2010).

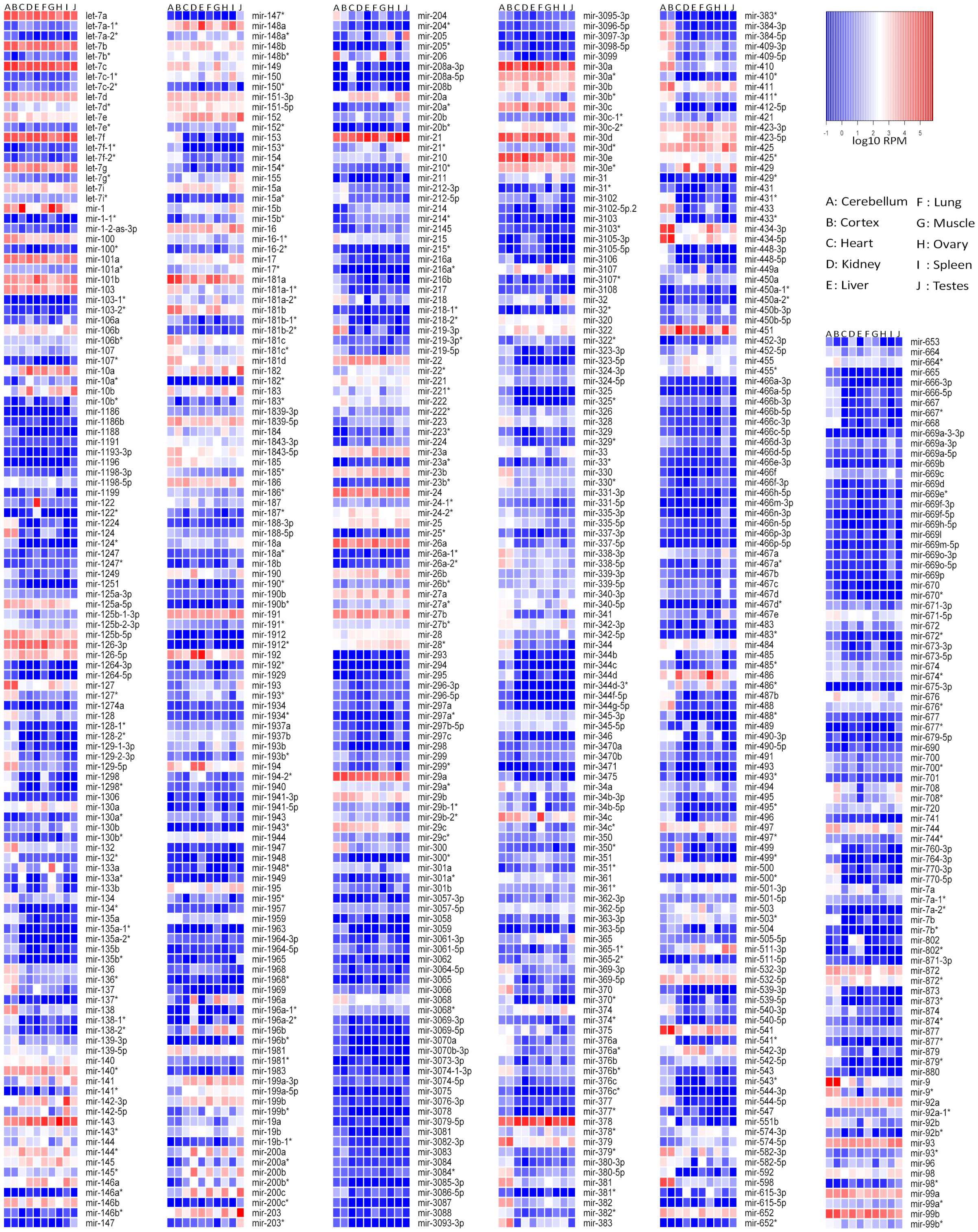


Figure 3.5 The expression profiles of the mouse miRNAs across the 10 tissues. For each tissue, miRNA reads are normalized to the total reads derived from miRNA hairpins in that tissue. miRNAs with normalized expression larger than 1 are plotted. Blue-white-red color intensity indicates an increasing log10 ratio of normalized reads derived from that tissue.

Understanding the tissue specificity of miRNA expression can often provide a hint of their function. Here, we defined tissue specificity using information content of the distribution of the relative sequencing frequencies acorss different tissues (see methods). The miRNA expression varied from highly tissue-specific to ubiquitous. Figure 3.6A shows the 50 miRNAs with the highest tissue specificity, of which, 36 were preferentially expressed in brain (cortex and cerebellum); mir-499, mir-499*, mir-208a and mir-208a* in heart; mir-133a and mir-206 in muscle; mir-196a-1* in kidney; mir-142-5p and mir-142-3p in spleen; mir-122 and mir-122* in liver; mir-203, mir-203* and mir-205 in testes. However, we also noticed that the abundant miRNAs were ubiquitously expressed across the tissues and showed little tissue specificity, the results are shown in Figure 3.6B.

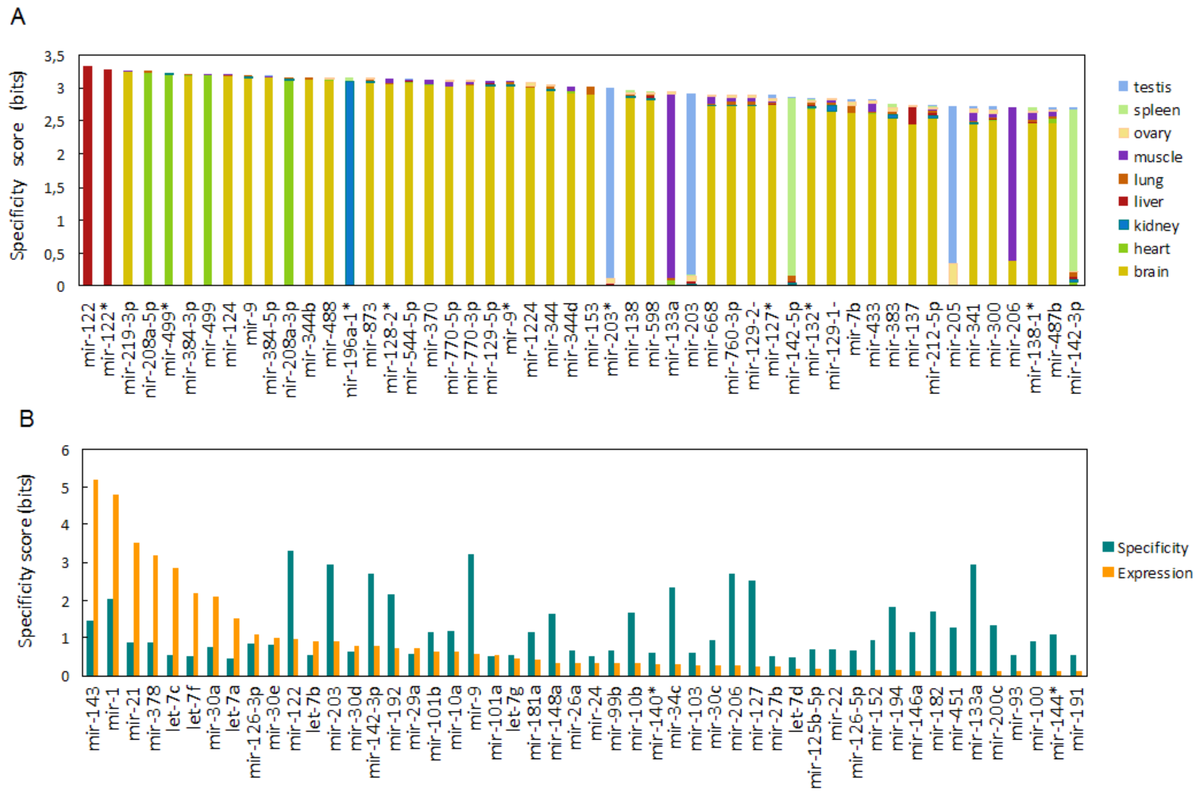


Figure 3.6 Tissue Specificity of miRNA Expression. A. The 50 most specific miRNAs are depicted. The total height of each bar represents the information content reflecting tissue specificity, while the relative heights for each of the tissues are proportional to the reads number of a miRNA in a given tissue type relative to all tissue types. B. The 50 most expressed miRNAs are depicted, each orange bar represents the pseudocount of each miRNA by calculating total normalized reads devided by 25000 (orange bar), and the tissue specificity (blue bar) are indicated. The most abundant miRNAs are ubiquitously expressed and show little tissue specificity.

3.1.4.2 Expression profiling of mouse pre-miRNA from the tissue mixture

In comparison with short reads, only 252,003 long reads generated from the tissue mixture sample matched 281 known pre-miRNAs, i.e., 5' and 3' end within a distance of 2 nt and 5 nt away from the corresponding ends of reference pre-miRNAs in miRBase v16.0 (Table 3.1). The number of sequence reads derived from 281 pre-miRNAs is in wide ranges, 40% of them are detected with 1-5 sequence reads (Figure 3.7). The distribution of length and the RNA secondary structure of these 281 pre-miRNAs do not differ from that of all mouse pre-miRNAs deposited in miRBase, indicating that our detection of pre-miRNAs is not biased towards particular subsets of pre-miRNAs (Figure 3.8).

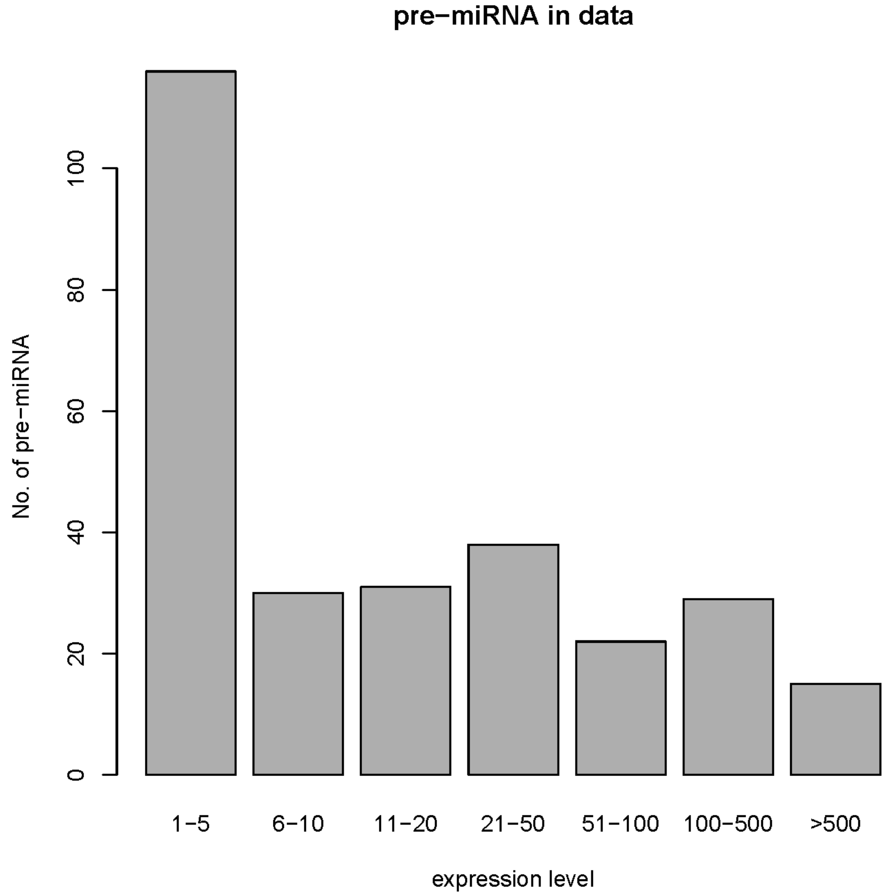


Figure 3.7 Expression level of detected pre-miRNA. Nearly half of the detected pre-miRNAs were sequenced with only a few reads, while a few had over 500 reads. Pre-miRNAs were grouped into seven bins according to the number of sequencing reads (x-axis), the height (y-axis) denotes the number of pre-miRNAs with corresponding read counts.

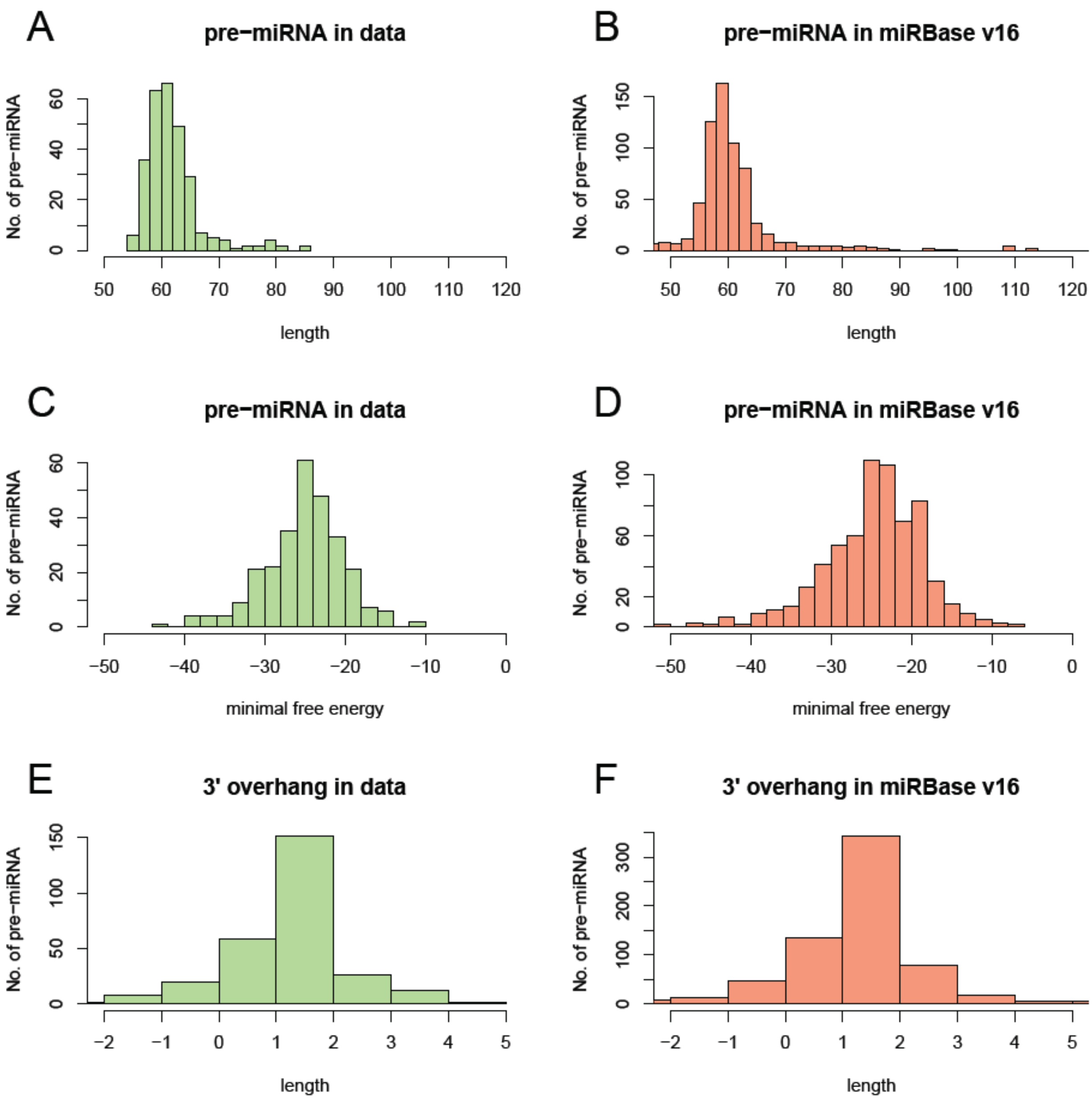


Figure 3.8 Length distribution of detected pre-miRNAs (A) and pre-miRNAs in miRBase v16.0 (B). Minimal free energy distribution of detected pre-miRNAs (C) and pre-miRNAs in miRBase v16.0 (D). Length distribution of 3’ overhang for detected pre-miRNAs (E) and pre-miRNAs in miRBase v16.0 (F).

3.1.4.3 Correlation between miRNA and pre-miRNA level

278 out of 281 detected pre-miRNAs have the corresponding miRNA/miRNA* present in at least one tissue. As shown in Figure 3.9, miRNAs with their pre-miRNAs detected are expressed at a significant higher level than those without (two-sided Wilcox rank-sum test, P < 2.2e-16), whereas the correlation between the abundance of miRNA and that of its precursor is low (R^2=0.1501). Such low correlation is most likely due to the fact that the majority of the detected pre-miRNAs have only few mapped reads and the different biases in cloning miRNAs versus the pre-miRNAs, although it might also be explained by the regulation of pre-miRNA processing via Dicer and its cofactors, especially for those miRNA loci in which there are many pre-miRNA reads, but few mature miRNA reads (Table 3.2).

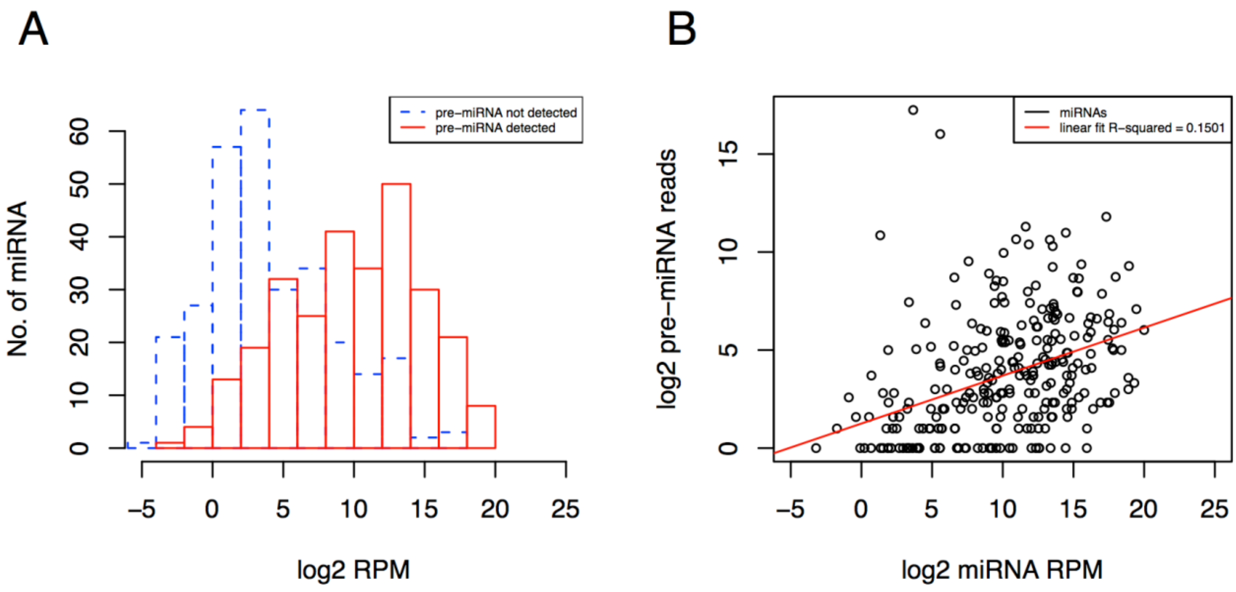


Figure 3.9 Comparison of the miRNA and pre-miRNA abundance. The miRNAs with precursor detected (red solid) are expressed at a higher lever than those without (blue dotted) (A). Correlation of the abundance between miRNAs (X-axis) and the pre-miRNAs (Y-axis) (B).

Table 3.2 List of highly expressed pre-miRNAs. miRNAs with abundant precursor, but low level of mature form are in red. The reads count of pre-miRNA is listed in second column, the abundance (RPM) of their mature miRNA in 10 tissues is listed in 3rd to12th columns.

ID	pre-miRNA	cerebellum	cortex	heart	kidney	liver	lung	muscle	ovary	spleen	testes
mmu-mir-677	155452	1	2	0	1	1	0	0	0	0	6
mmu-mir-3096	66471	1	1	0	2	1	2	2	2	18	19
mmu-mir-142	3569	275	235	3643	1646	1679	5418	41	382	149570	1157
mmu-mir-330	2515	1108	1702	18	62	16	83	33	43	32	52
mmu-mir-200c	2024	17	12	276	3730	75	4091	263	3196	149	10730
mmu-mir-3078	1850	1	1	0	0	0	0	0	0	0	0
mmu-mir-139	1609	117	922	155	112	111	29	236	103	116	69
mmu-mir-186	1583	1332	1329	2339	601	459	886	465	287	2352	224
mmu-mir-370	1340	671	2946	4	1	3	1	34	3	0	1
mmu-mir-133a-2	1261	15	2	419	4	3	134	11116	131	1	5
mmu-mir-193	994	5	4	10	80	95	83	105	27	6	653
mmu-mir-208a	738	0	0	188	1	0	2	0	0	0	0
mmu-mir-27b	664	5071	6324	2124	7238	3918	7831	2132	5305	1082	7308
mmu-mir-30a	626	63672	59082	12955	119706	47303	123267	18416	18741	3481	31303
mmu-mir-133a-1	605	15	2	419	4	3	134	11116	131	1	5

3.2 Novel aspects of miRNA processing and modification

3.2.1 Detection of Ago2-cleaved pre-miRNAs in mouse

Almost half of the long reads mapped to known miRNA loci did not cover full-length pre-miRNAs. Whereas most of these reads possibly represented degradation product from pre-miRNAs or pri-miRNAs, some were derived from miRNA processing intermediates. For example, the long reads that missed one arm of the hairpin but contained the other arm and full loop sequences might be abnormally processed by Dicer. Another kind of long reads matching pre-miRNAs at one end but truncated from the other end resembled an endogenous intermediate resulted from Ago2-mediated endonucleolytic cleavage within one arm of the hairpin precursor, which has been identified before in human cells and termed as Ago2-cleaved pre-miRNAs (ac-pre-miRNA) (Diederichs and Haber 2007).

After carefully screening for the slicing signature of Ago2 (methods), we found eight potential ac-pre-miRNAs (Table 3.3), out of which seven were from let-7 families. The cleavage sites were found always at the 3' arms that also represent passenger strands (Table 3.4). This finding is the same as reported before and is consistent with the proposed function of Ago2 cleavage at hairpin precursor, i.e. to facilitate removal of the nicked passenger strand from RISC after maturation. Interestingly, for all the ac-pre-miRNAs, we observed uridylation at 3' end (Table 3.3). Importantly, this provides further evidence that our ac-pre-miRNA candidates likely represent the real processing intermediates rather than degradation products.

Table 3.3. Number of sequencing reads derived from full-length pre-miRNA, ac-pre-miRNAs, (poly-)uridylated ac-pre-miRNAs and number of their mature miRNA reads derived from 5’ and 3’ arms.

miRNA	full-length Pre- miRNA	ac-pre-miRNA	uridylated ac-pre-miRNA	5’ arm miRNA	3’ arm miRNA
mmu-let-7a-1	46	25	10	1682015	420
mmu-let-7b	34	178	100	2222105	299
mmu-let-7c-2	84	22	8	3347997	419
mmu-let-7d	6	2	2	389631	7748
mmu-let-7f-1	29	297	62	1940949	72
mmu-let-7i	93	49	2	111553	1533
mmu-mir-98	24	4	3	18033	73
mmu-mir-30b	144	2	2	88407	3040

Table 3.4 The cleavage site of the detected ac-pre-miRNAs.

mmu-let-7a-1

```
       u                     uuagggucacac
        gagguaguagguuguauaguu            c
        |||||||||||||||||||||            c
     cu uucugucaucuaacauaucaa            a
       -                     uagagggucacc
                 ↑
              cleavage
```

mmu-let-7b

```
   u                     ----  ---a      u
    gagguaguagguugugugguu    uc    gggcag g
    |||||||||||||||||||||    ||    |||||| a
 cc uuccgucauccaacauaucaa    ag    cccguu u
   -                     uaga  ccuc      g
            ↑
        cleavage
```

mmu-let-7c-2

```
         u                     uugggcucugc
          gagguaguagguuguaugguu           c
          |||||||||||||||||||||
       cc uucugucaucuaacauaucaa           c
         u                     uggcgucucgc
                   ↑
                cleavage
```

mmu-let-7d

```
      a              c      -------uua
       gagguaguagguug auaguu          gggcagag
       |||||||||||||| ||||||          ||||||||a
    cu uuccgucguccagc uaucaa          cccguuuu
      -              a      uugaggaaca
               ↑
           cleavage
```

mmu-let-7f-1

```
      u                       ---------       u
       gagguaguagauuguauaguugu         gggguag g
       |||||||||||||||||||||||         ||||||| a
    cc uuccguuaucuaacauaucaaua         ucccauu u
      c                       gaggauuug       u
                 ↑
              cleavage
```

mmu-let-7i

```
       u                 u   --------  u     ugu
        gagguaguaguuugugc guu        gg cgggu   g
        ||||||||||||||||| |||        || |||||   a
     cg uuccgucaucgaacgcg caa        uc gcccg   c
       -                 u   uagaggug  -     uua
                ↑
            cleavage
```

mmu-mir-98

```
      u                 u    ---------      aggg
       gagguaguaaguuguau guug         uggggu    a
       ||||||||||||||||| ||||         ||||||    u
    cc uuucaucauucaacaua caau         accccg    u
      -                 u    agaagaaug      gauu
               ↑
           cleavage
```

mmu-mir-30b

```
                       u   -        u aua
            uguaaacaucc aca cucagcug c   c
            ||||||||||| ||| |||||||| |
            ugcauuuguagg ugu gggucggu g   a
           c            -   a        u cgu
                    ↑
                cleavage
```

3.2.2 Untemplated modification of miRNAs and pre-miRNAs

Addition of untemplated nucleotides at the 3’ end of miRNAs has been observed in mammals, worms and fruit flies (Ruby, Jan et al. 2006; Landgraf, Rusu et al. 2007; Ruby, Stark et al. 2007). It has been shown that uridylation by TUT-4 can block Dicer processing and cause decay of pre-let-7 (Heo, Joo et al. 2009), while a single adenine added to the 3’ end of mir-122 by GLD-2 can lead to its stabilization (Katoh, Sakaguchi et al. 2009).

To further understand the untemplated modification of mouse miRNAs, we characterized such events for both mature and precursor miRNAs using the sequence information obtained in this study. Nucleotides most frequently added to the 3’ end of miRNAs were U and A, whereas additions of C and G were not or only slightly higher than that at tRNA fragments (Table 3.5) (see methods). As observed before, uridylation happened much more frequently at the end of miRNAs derived from the pre-miRNA 3’ arm. This bias has been attributed to uridylation of the 3’ end of pre-miRNA prior to Dicer cleavage which would only remain at the 3’ arm but not the 5’ arm of mature miRNAs (Chiang, Schoenfeld et al. 2010).

Table 3.5 Untemplated nucleotide addition frequency for miRNA/miRNA* reads from the indicated arm, as well as pre-miRNAs. As a control, tRNA degradation fragments were analyzed similarly. Numbers of genes added with the specific nucleotide are indicated in the parentheses.

	A	C	G	U
5’ arm	5.28% (416)	0.17% (264)	0.24% (303)	1.89% (396)
3’ arm	5.44% (318)	0.45% (270)	0.28% (277)	14.52% (399)
pre-miRNA	0.09% (42)	0.02% (17)	0.05% (12)	0.39% (77)
tRNA	0.87% (450)	0.65% (450)	0.11% (450)	0.26% (450)

For pre-miRNAs, uridylation was also the most frequent modification (Table 3.5). Out of 77 uridylated pre-miRNAs, 52 had U addition at the 3’ arm miRNAs more frequent than that of the 5’ arm miRNAs. Our results indicated that the uridylation, the dominant modification acted on the pre-miRNA hairpins, can to some extent lead to the higher frequency of U addition to 3’ arm miRNAs.

We compared the addition pattern of individual mature miRNA across the 10 tissues, and observed that mir-143, mir-23b, mir-1957, mir-3103*, mir-328* mir-699e* and mir-879* were consistently uridylated, while let-7f, mir-369-5p and mir-802* were consistently adenylated.

3.2.3 miRNA editing

A-to-I editing has been reported in mammalian pre-miRNAs and such events can affect miRNA biogenesis as well as target selection (Nishikura 2010). Given the prominent expression of adenosine deaminases (ADARs) in brain, we focused our analysis on two libraries generated from cerebellum and cortex. Since A-to-I editing represents as A-to-G change in sequence, we searched for A-to-G change in the sequencing data.

First, the non-genome-mapping reads from small RNA (10-40 nt) libraries were mapped to reference miRNA sequences with one mismatch and the reads aligned to unique position were selected for miRNA editing analysis. After excluding reads with mismatches at the 3' end, we obtained 292,573 (cortex) and 324,408 (cerebellum) reads, corresponding to 2.18% (cortex) and 2.38% (cerebellum) of the reads that were mapped to miRBase without mismatch. Of these reads, the fraction of A-G change was 51% (cortex) and 54% (cerebellum). Most of these changes could be background noise caused by errors introduced during sequencing library preparation or sequencing process.

To distinguish the true editing events from noise, the fraction of a change, as the number of reads bearing that change divided by the number of all reads containing all mismatches at the same position, was calculated. Then, to be a good candidate, its fraction should exceed a threshold of 90% and the change should be found in at least 10 reads. Of the 165,274 (cortex) and 202,227 (cerebellum) reads bearing these significant changes, A-G changes made up of 82% (cortex) and 82% (cerebellum).

Finally, we took advantage of our pre-miRNA sequencing data and considered only the changes that were also found in at least one pre-miRNA sequencing reads. Because most of the changes found in the pre-miRNA were present in the one or two reads, we also required that the sequencing quality at the position to be high (quality score >= 30). Only 36,280 (cortex) and 64,173 (cerebellum) reads containing A-G changes were found for 13 (cortex) and 11 (cerebellum) pre-miRNAs, which accounts for 99.4% (cortex) and 99.8% (cerebellum) of the reads after final filtering (Table 3.6).

Table 3.6 Number of A-G changes in cerebellum and cortex during each step of filtering. The A-G change percentage (in parentheses) increases to nearly 100% after using both miRNA and pre-miRNA sequencing data.

Tissue	without mismatch	with 1 mismatch	with A-G change	With 1 mismatch after 1st filtering	with A-G changes after 1st filtering	with 1 mismatch after precursor filtering	with A-G change after precursor filtering
Cerebellum	13625198	324408	174886 (0.54)	202227	165436 (0.82)	64292	64173 (0.998)
Cortex	13403852	292573	148592 (0.51)	165274	135984 (0.82)	36493	36280 (0.994)

As listed in Table 3.7, in total, 15 editing sites were found in brain tissues. Of them, 9 editing sites were shared between cortex and cerebellum, but the editing frequency was still different in the 2 different brain regions. 11 of 15 sites located in the seed regions of miRNAs, which might affect selection of mRNA targets, as previously described in the mir-376 cluster(Kawahara, Zinshteyn et al. 2007). Of the remaining 4 sites outside of seed regions, the A-to-I editing at pre-mir-497 could affect its secondary structure by forming 'I-C' base pair with the cytosine on the opposite arm, thereby impacting the processing by Dicer and even loading into RISC.

Table 3.7 Inferred A-to-I editing sites in miRNAs.

miRNA	Position	No. of pre-miRNA reads	Edited fraction (cortex)	Edited fraction (cerebellum)
mmu-mir-137--3p	11	2	0.0053	0
mmu-mir-151--3p	3	5	0.1179	0.0266
mmu-mir-186--5p	3	1	0.0006	0
mmu-mir-27a--5p	7	14	0.1229	0
mmu-mir-376a--5p	4	2	0.0831	0.1115
mmu-mir-376b-5p	6	2	0	0.3781
mmu-mir-376b--3p	6	7	0.4612	0.4390
mmu-mir-376c--3p	6	3	0.2342	0.2603
mmu-mir-378--3p	16	1	0.0606	0.1814
mmu-mir-381--3p	4	1	0.3940	0.2278
mmu-mir-384--5p	4	1	0.0115	0.0096
mmu-mir-497--3p	20	104	0.8846	0.9769
mmu-mir-540--5p	3	1	0.2661	0.4810
mmu-mir-770-5p	4	1	0	0.0206
mmu-mir-770--3p	11	2	0.0023	0

If as previously reported, editing of some pri- or pre-miRNAs can inhibit or even block the miRNA biogenesis, we will underestimate the editing frequency in the analysis. To check for the possible edited pre-miRNA that cannot be processed by Dicer, we searched for the editing events identified in precursor reads not supported by matching mature reads. Of 94701 precursor reads with one or two nucleotide difference, A-to-I changes accounted for only 5% of reads and thus not significant. Further manual checking also did not reveal any to be true editing candidates.

It is well known that RNA (ADAR) enzyme converts adenosines to inosines (A-to-I editing) specifically in double-stranded RNA (dsRNA) substrates (Kawahara, Megraw et al. 2008). Here, we examined the effects of base pairings on editing site as well as at the -1 and +1 position in the 15 editing sites we identified. 10 out of 15 editing sites were uridine in the opposite strand which contributed 91.77% and 83.30% of editing events in cortex and cerebellum, respectively, followed by Cytosine (4/15) and Adenosine (1/15) (Table 3.8).

Table 3.8 Frequency of the nucleotide opposite to the editing sites.

The complementary nucleotide					
		U	C	A	G
Edited / total		10/15	4/15	1/15	0/15
Editing frequency (%)	Cortex	91.77	0.82	0	0
	Cerebellum	83.30	16.04	0.66	0

We also found that both the -1 and +1 positions were stabilized by Watson-Crick base pairings at 12 out of 15 editing sites (Table 3.9), which indicated that two neighbouring nucleotides of the editing sites needed to be stably base paired in order to be edited efficiently.

Table 3.9 The effect of the base pairing at –1 and +1 position of editing sites.

		W-C group	Mismatch group
Edited /total		12	3
Editing frequency (%)	Cortex	56.4	43.6
	Cerebellum	53.2	46.8

The preference of the UAG triplet sequences for A-to-I editing was reported previously (Kawahara, Megraw et al. 2008), and we observed the same pattern in the 15 editing site we identified, that 33.9% (cortex) and 26.9% (cerebellum) of the editing reads contained UAG triplet. Besides UAG triplet, AAG, UAC, UAU, CAG were also frequently edited (Table 3.10). In contrast, guanosine as the 5' end of the editing of triplets was extremely rare, which was consistent with the previous finding of the preference of 5'-A and U, disfavor of G, as direct upstream nucleotide for A-to-I editing by ADAR1 and ADAR2 (Kawahara, Megraw et al. 2008).

Interestingly, following the same filtering procedure, we also observed the editing in other tissues (Table 3.11). mmu-mir-497, mmu-mir-378, mmu-mir-381, mmu-mir-151 and mmu-mir-27a were found to be edited in 6, 6, 3, 2 and 1 different tissues other than brain, respectively. This result indicates the A-to-I editing is predominant in brain tissues, but also presents in tissues other than brain.

Table 3.10 Editing frequency for triplet sequence in editing sites.

	Cortex	**Cerebellum**
aAg	10.26%	15.45%
uAg	33.9%	26.9%
uAc	14.9%	7.32%
uAa	0.43%	0.31%
uAu	0%	12.15%
cAg	40.39%	37.21%
cAc	0%	0.66%
gAc	0.09%	0%
aAa	0.02%	0%

Table 3.11 A-to-I editing events in other mouse tissues other than brain.

miRNA	Position	No. of pre-miRNA reads	Fraction edited (heart)	Fraction edited (kidney)	Fraction edited (liver)	Fraction edited (lung)	Fraction edited (muscle)	Fraction edited (ovary)	Fraction edited (spleen)	Fraction edited (testes)
mmu-mir-151--3p	3	5	0	0	0	0.0018	0.0009	0	0	0
mmu-mir-27a--5p	7	14	0	0	0.0719	0	0	0	0	0
mmu-mir-378--3p	16	1	0.0038	0.0207	0.0232	0.0492	0	0.0668	0	0.0078
mmu-mir-381--3p	4	1	0	0	0	0.0634	0.0344	0.0272	0	0
mmu-mir-497--3p	20	104	0.5429	0.6154	0.8235	0.9100	0	0.5769	0.6667	0

3.3 Novel miRNA prediction

3.3.1 Development of miRNA prediction strategy independent of genome reference

In order to identify potential miRNAs based on both short reads (corresponding to miRNA/miRNA*) and long reads (corresponding to pre-miRNA), we developed a computational pipeline. In this pipeline, after mapping the short reads to the long reads, potential pre-miRNA sequences were extracted by selecting the long reads that can form a hairpin and on which the mapping pattern of short reads were compatible with Dicer processing. We scored these pre-miRNA candidates using six known features and finally reported a list of highly confident pre-miRNAs together with the corresponding mature miRNAs.

In more detail, first of all, the positions of short reads mapped to long reads were investigated. If a long read and its mapped short reads represent a genuine pre-miRNA and its Dicer processing products, the short reads should clustered to a maximum of three positions, i.e. at 5’ end, 3’ end and the middle of the long read, corresponding to the miRNA/miRNA* and the loop, respectively. Given the rapid degradation of loop fragments, we would also expect that the number of short reads from the 5’ and 3’ end clusters is much bigger than that from the middle cluster. After discarding the long reads that did not fit with such criteria and most likely arose from other small non-coding RNAs or degradation products from longer RNA transcripts, we predicted the secondary structure of the remaining long reads and kept only those that can form stable hairpin structures as potential candidates. Finally, these candidates were scored based on the characteristics of known miRNAs and pre-miRNAs, including pre-miRNA thermodynamic stability, 5’ and 3’ end duplex overhang as well as the length of pre-miRNA and mature miRNA (Figure 3.10 and 3.11).

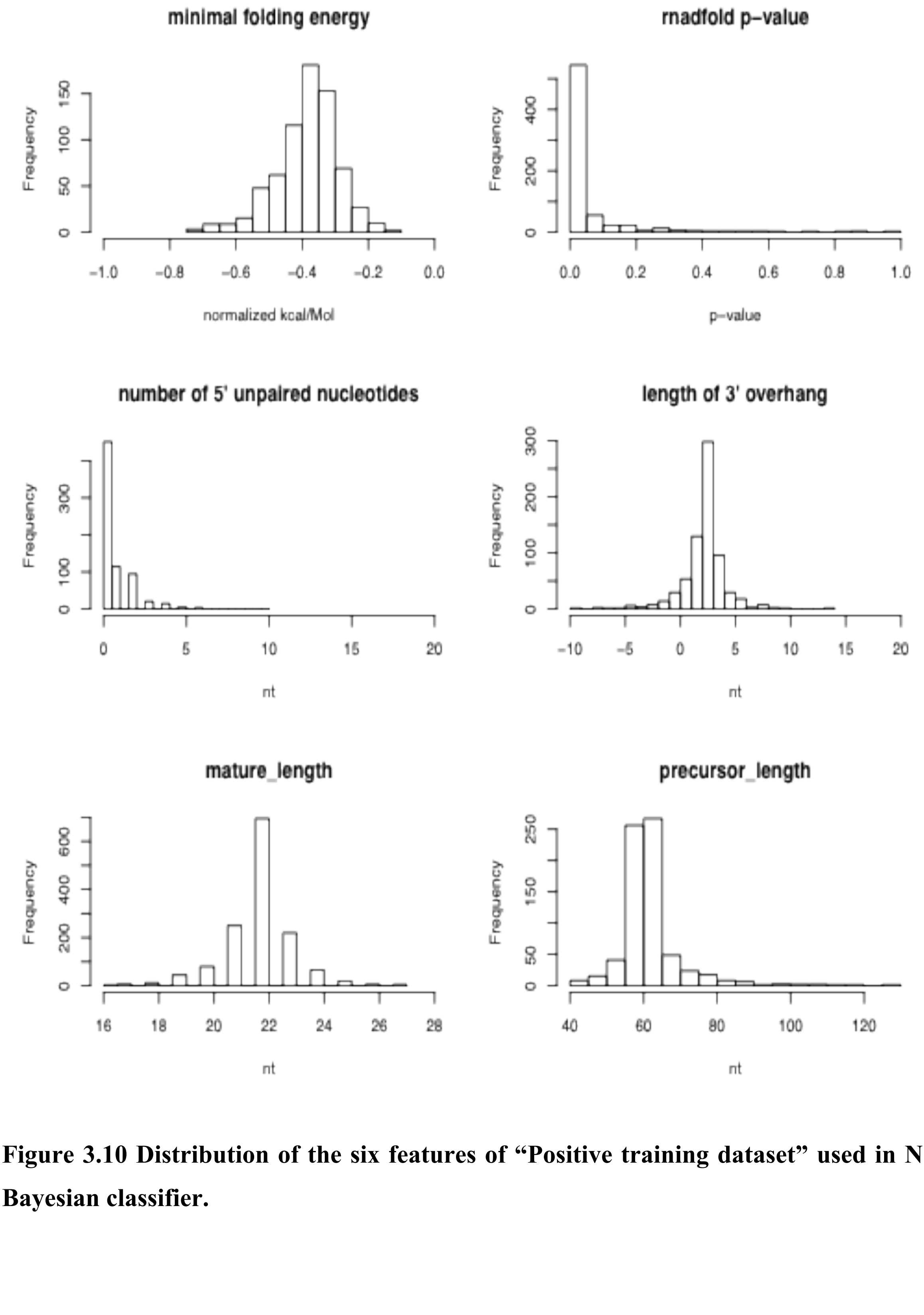


Figure 3.10 Distribution of the six features of "Positive training dataset" used in Naïve Bayesian classifier.

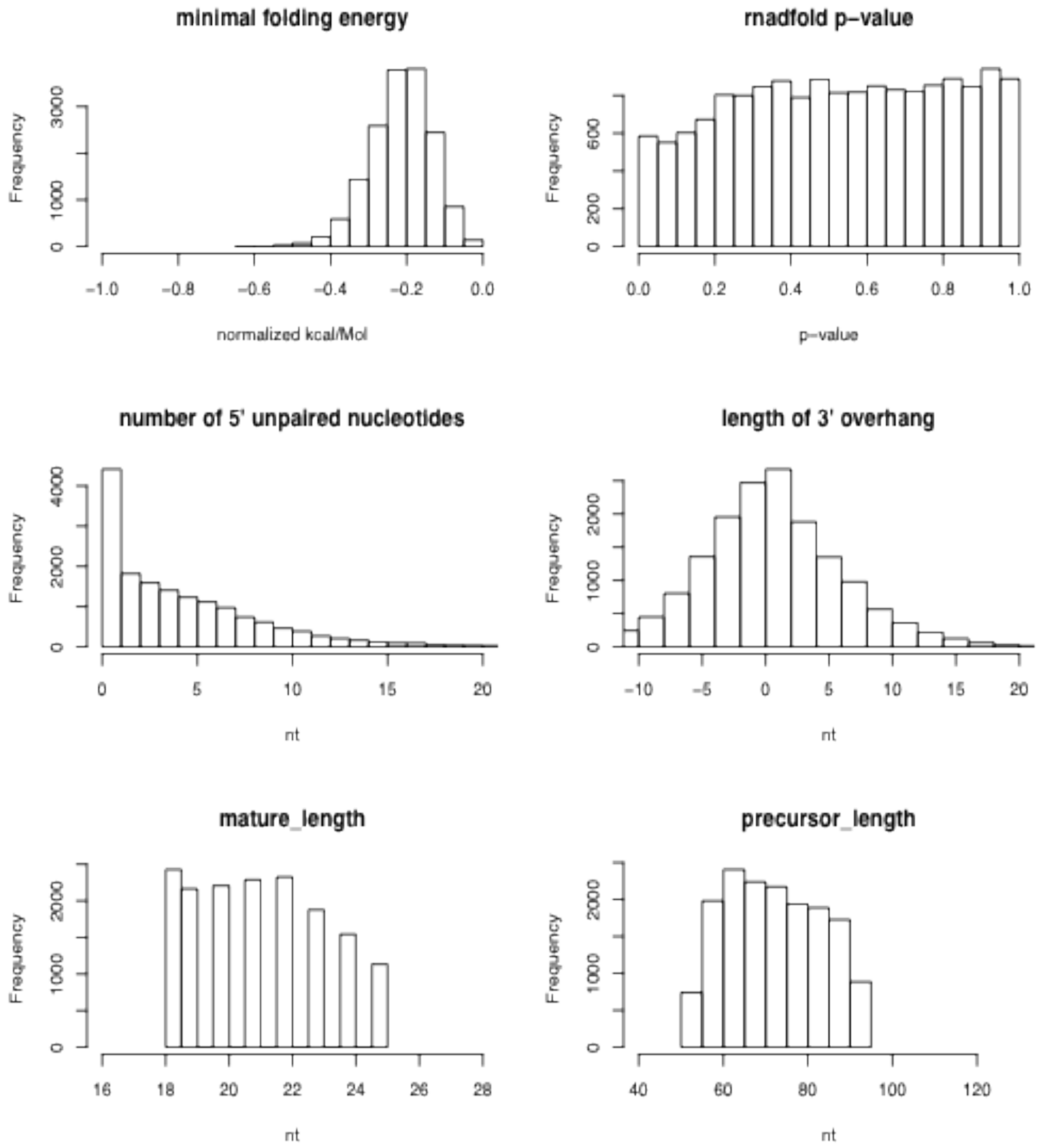


Figure 3.11 Distribution of the six features of "Negative training dataset" used in Naïve Bayesian classifier.

3.3.2 Application of miRNA prediction strategy independent of genome reference in mouse data

We applied the pipeline to predict mouse miRNAs based on our sequencing dataset. In total, 155,760,811 short reads were perfectly mapped to 5,789,406 distinct long reads. From these, 5,524,656 long reads were discarded because the mapping position of short reads did not fit with the model of Dicer processing. The remaining 264,750 long reads were then merged into 131,207 clusters and one representative read from each cluster was selected to predict the secondary structure (see methods). Out of 1,277 long reads that can form stable hairpin structures, 538 potential pre-miRNAs passed the cut-off of 0.95 after probabilistic scoring (Figure 3.12). 324 candidate pre-miRNAs, on which at least five short reads could be mapped, were retained.

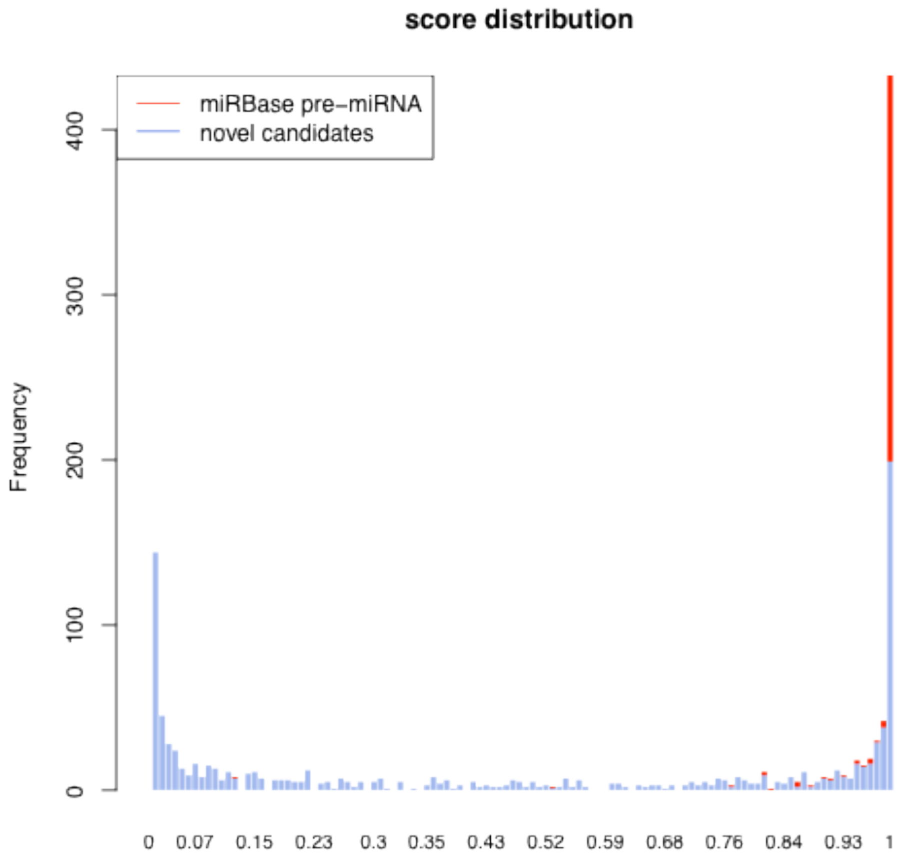


Figure 3.12 Probabilistic score distribution of pre-miRNA candidates. Known miRNAs are shown in red, novel candidates in blue.

Of these 324 pre-miRNA candidates, 247 corresponded to 238 known mouse pre-miRNAs. Manual inspection of the 77 novel candidates revealed that eight were derived from chimerical reads representing ligation artifacts. Of the remaining 69 pre-miRNA candidates, four were recently registered in miRBase (latest version 17). Five could be mapped to LINE, SINE or LTR repetitive loci, consistent with previous finding that miRNA genes could originate from transcribed transposons (Smalheiser and Torvik 2005; Piriyapongsa and Jordan 2007). Twelve candidates located to known snoRNA loci. The similar finding has been reported in the study of human cells (Ender, Krek et al. 2008). The remaining 48 predicted novel pre-miRNAs mapped to either intergenic (5), or exonic (8) and intronic (35) regions of protein coding genes. From 69 novel pre-miRNAs, 112 miRNA/miRNA* could be identified. Compared to known miRNAs, these novel miRNAs expressed at a much lower level (Figure 3.13). Only eight were expressed (RPM >= 1) in one tissue while six were expressed in at least 2 tissues. Detailed information about these novel candidates can be found in Appendix table 2.

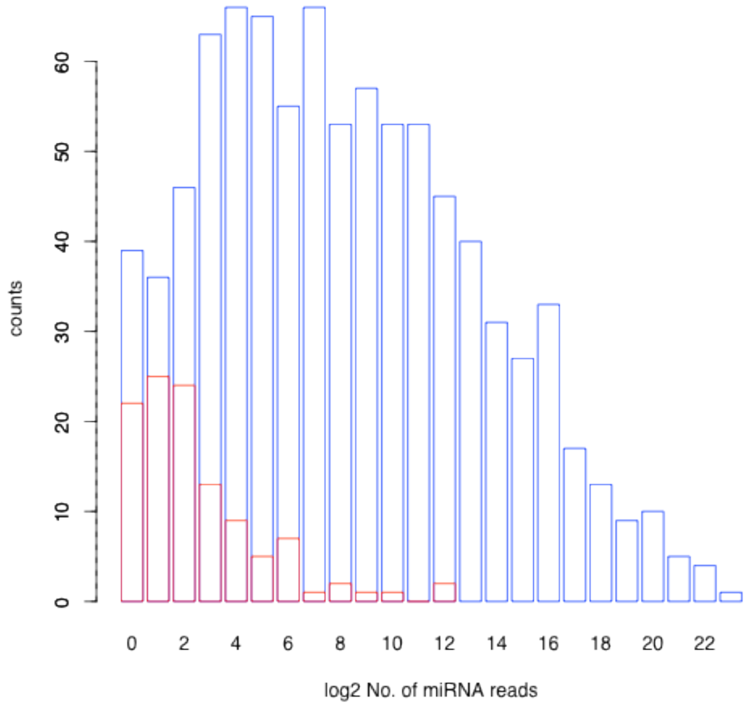


Figure 3.13 Distribution of the abundance (number of reads) of known miRNAs (blue) versus novel miRNAs (red).

3.3.3 Experimental validation of novel miRNAs

3.3.3.1 The predicted novel miRNAs are downregulated upon Dicer knockdown

To investigate whether the novel miRNAs we identified are indeed dependent on Dicer for expression, we used RNA interference to knock down Dicer in a mouse N2a cell line (see methods). RT-qPCR showed that in cells treated with siRNA, the level of Dicer mRNA transcripts was reduced to 15% of that in unperturbed cells (data not show). After sequencing the small RNAs (10-40 nt) from unperturbed and treated cells, we compared the abundance of different non-coding RNA derived transcripts between the two samples. More specifically, we counted the sequencing reads mapped to tRNA and rRNA, transcripts that are believed not to be processed by Dicer, to known miRNA loci deposited in miRBase, as well as to the novel pre-miRNAs that we have identified. As shown in Figure 3.14, whereas on average rRNAs showed no change and there was a median of 23% increase for tRNA transcripts after silencing Dicer, both known and novel miRNAs decreased in abundance with a median reduction of 32% (log2 fold-change of -0.55) and 46% (log2 fold-change of -0.89), respectively. These results demonstrate that the novel miRNAs predicted in this study are enriched in Dicer dependent small RNAs.

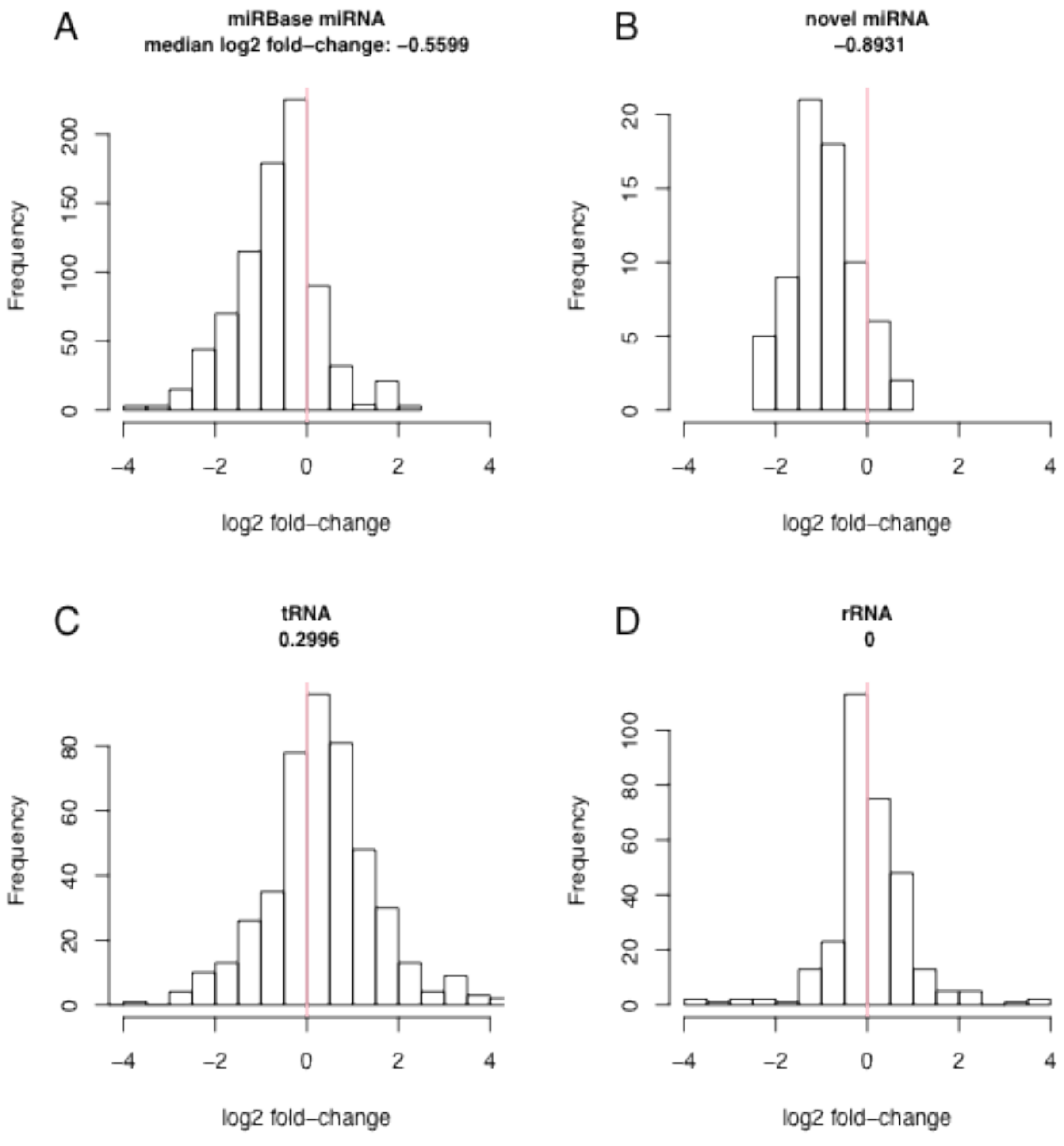


Figure 3.14 RNA interference was used to silence Dicer in mouse N2a cell line. Log2 fold-changes in small RNA expression are noted for known miRNAs (A), novel miRNAs (B), tRNAs (C), rRNAs (D). The median fold-change is indicated above each plot.

3.3.3.2 Two novel pre-miRNA candidates are derived from snoRNAs

Two novel pre-miRNA candidates that derived from Snora7a and Snora41 loci were selected for further experimental investigation. More specifically, we validated the processing of the two snoRNAs into mature miRNAs *in vivo* and *in vitro*. Using northern blotting, with a probe complementary to the potential mature miRNA, we could detect the full-length snoRNAs as well as two bands of the sizes of the pre-miRNAs and mature miRNAs that we predicted (Figure 3.15). The expression of mature miRNAs could also be validated by RT-qPCR using customized TaqMan assays (data not show).

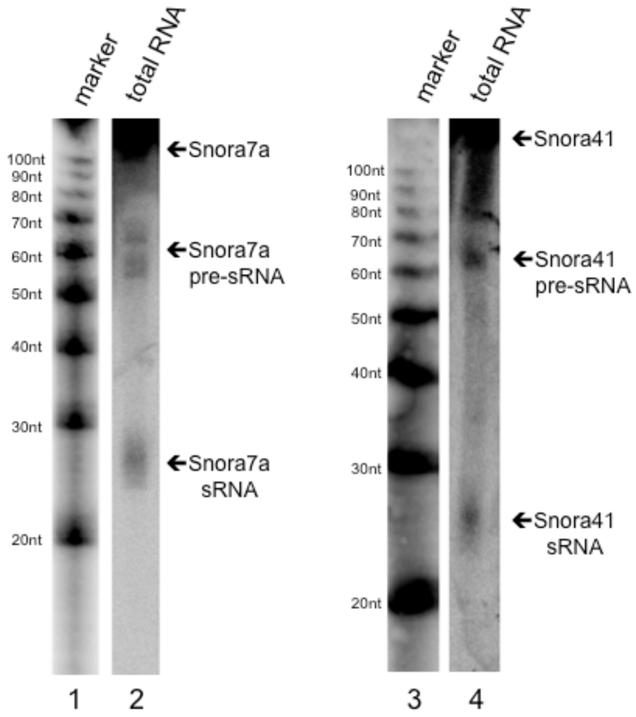


Figure 3.15 Detection of two novel miRNAs using northern blot. Total RNA from mouse tissue (50 μg) was blotted, and the membrane was incubated with probes specific for the novel miRNAs derived from Snora7a (lane 2) and Snora41 (lane 4). Lane 1 and 3 show a size marker.

To investigate whether the processing of the two snoRNAs is dependent on Dicer, we incubated the *in vitro* transcribed and ^{32}P-labeled potential pre-miRNAs as well as full-length snoRNAs with recombinant Dicer. Potential pre-miRNAs from both loci could be efficiently processed (Figure 3.16A). Full-length Snora7a, but not Snora41, could be processed by Dicer to the mature miRNA, with the potential pre-miRNA as an intermediate (Figure 3.16B). The mechanism behind the processing of Snora41 to the hairpin precursor remains unclear.

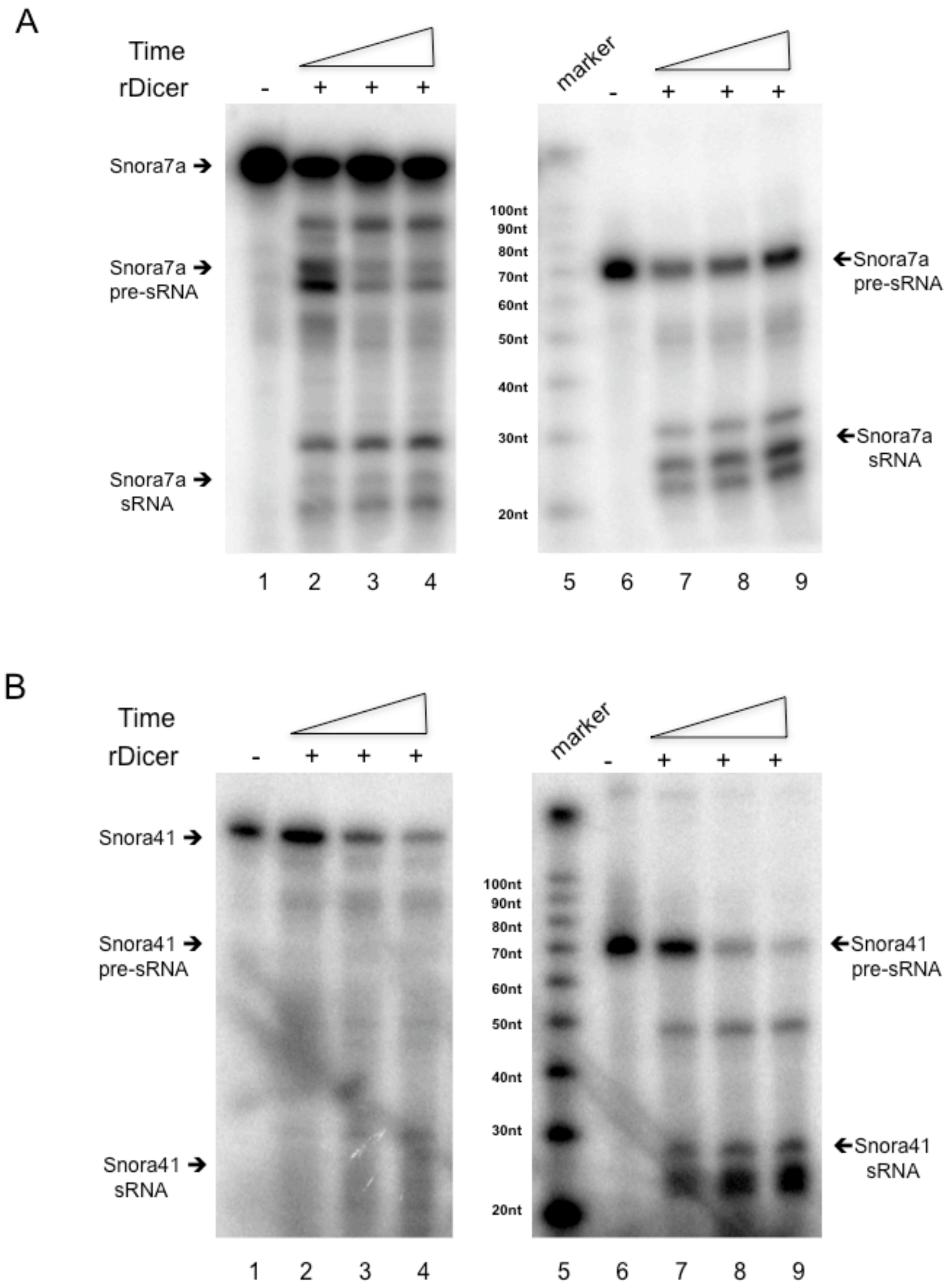


Figure 3.16 *In vitro* rDicer processing. A. ^{32}P-labeled Snora7a (lane 1-4) or Snora7a pre-sRNA (lane 6-9) were incubated with recombinant Dicer and an increasing incubation time. Lane 5 shows a size marker. B. ^{32}P-labeled Snora41 (lane 1-4) or Snora41 pre-sRNA (lane 6-9) were incubated with recombinant Dicer and an increasing incubation time. Lane 5 shows a size marker.

4 Discussion

4.1 The difficulties of pre-miRNA sequencing

In contrast to the analysis of mature miRNAs, attempts to profile pre-miRNAs are rather limited. To date, the expression patterns of known pre-miRNAs have been analyzed by using northern blot, *in situ* hybridization and qPCR. But due to the relatively laborious procedure, such experiments have seldom been performed at the global level. The precise sequences of most, if not all, pre-miRNA sequences were not directly determined by sequencing experiments. Instead, they were often inferred from the sequences of the corresponding miRNA and miRNA* and ambiguity could arise when the miRNA* was not identified.

Generation of pre-miRNAs from pri-miRNA transcripts and their further processing are critical steps in miRNA biogenesis. Several studies on individual miRNAs have demonstrated that these steps are under control of Drosha, Dicer and other accessory proteins, such as Lin-28. In principle, understanding of such regulation on a genome-wide level would greatly benefit from efficient parallel profiling of pre-miRNAs and mature miRNAs. In this study, we sequenced for the first time pre-miRNAs in an unbiased genome-wide manner. Out of over 50 millions reads we obtained, pre-miRNAs constitute only < 1% even after cDNA normalization. In total, 281 pre-miRNAs were identified, most of which had a limited number of mapped reads. This might be due to two non-mutually exclusive possibilities, i.e. 1) The abundance of most pre-miRNAs is rather low, because they serve as an intermediate during miRNA maturation and most of them might be rapidly processed by Dicer. 2) Their stable hairpin structure make the cloning extremely inefficient. With the limited sequencing depth, apparently we could not estimate the pre-miRNA expression level based on our pre-miRNA sequencing dataset. However, together with mature miRNA sequencing results, we used the sequencing information of pre-miRNA for not only revealing new aspects of miRNA processing and modification but also predicting novel miRNAs independent of the genome reference sequences.

4.2 The aspects of pre-miRNA processing

To date, only four ac-pre-miRNAs have been reported in a study of human cells, including three let-7 family members and mir-20a (Diederichs and Haber 2007). Although we have detected full-length pre-mir-20a, no reads mapped to the mouse mir-20a locus could be generated by Ago2 cleavage. To find out why mouse pre-mir-20a was not processed as its human homolog, we compared the sequence and secondary structure of those two. Whereas human pre-mir-20a contains no mismatch in the vicinity of Ago2 cleavage site, the potential cleavage site at mouse pre-mir-20a is flanked by one mismatch that might disrupt the endonuclease activity of Ago2. Only two out of nine potential ac-pre-miRNAs found in this and previous studies did not belong to let-7 families. Also, the frequency of the cleaved form relative to the full-length pre-miRNAs was much lower in these two miRNAs, compared with let-7 family members. Even though the pre-miRNA sequencing coverage achieved in our study is not high enough to draw a concrete conclusion, we are still prompted to consider pre-let-7s as major, if not the only substrates of Ago2 cleavage with functional significance.

By deep-sequencing of mouse miRNAs from ES cells, brain, ovary and testes, Chiang *et al.* (Chiang, Schoenfeld et al. 2010) observed 16 sites with an editing fraction higher than 5% in miRNAs from brain, seven of which were also detected in our study. For the remaining nine sites, we did not observe the A-G change at the precursor reads although we indeed found the editing for eight sites at the mature miRNA sequencing reads (except mmu-mir-219-2-3p). The eight new editing sites that we identified were not reported in previous studies most likely due to their low editing fractions (Figure 4.1), indicating the capability of our approach to identify the miRNAs edited with low frequency. In a complex tissue such as brain with extremely heterogeneous cell types, the miRNAs detected with low editing frequency in an anatomically distinct region could well be much more significantly edited in a specific cell type. Also, depending on the miRNA expression level, even at low frequency, the absolute number of edited molecules can still be significant. Therefore the new editing events that we have identified with low frequency were not necessary without functional consequence.

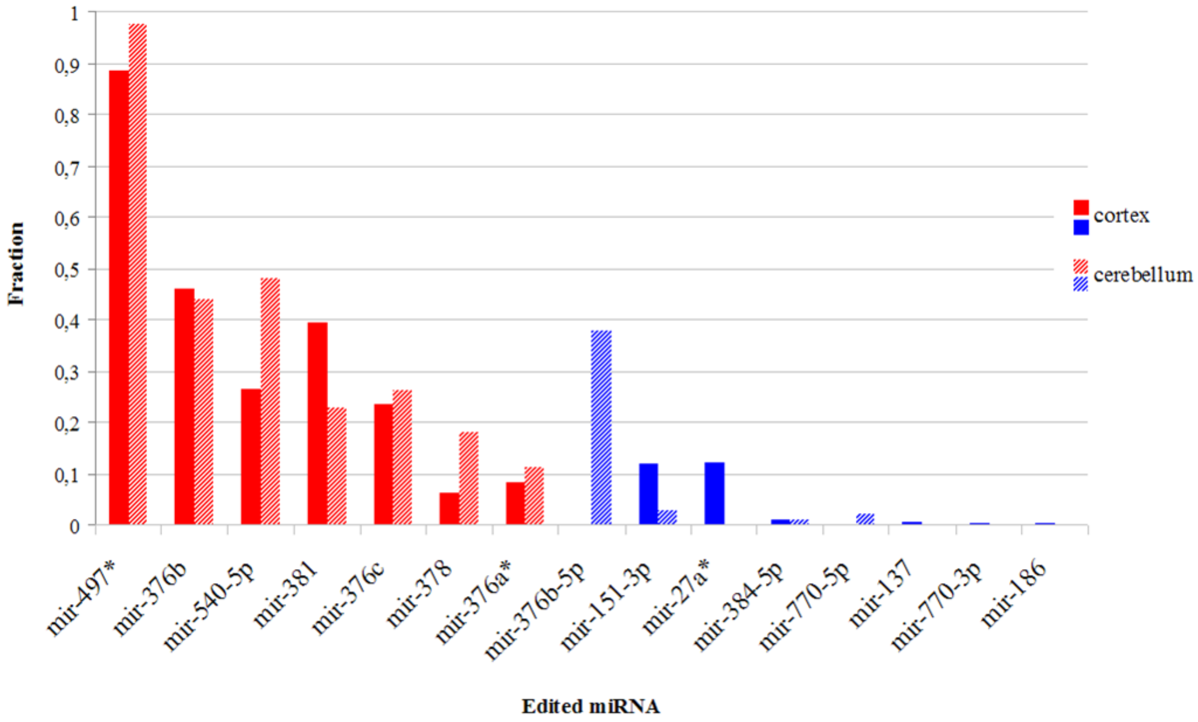


Figure 4.1 Editing frequency in cortex and cerebellum of previously known (red) editing events (Chiang, Schoenfeld et al. 2010) is generally much higher than editing frequency of the newly identified (blue) editing events.

4.3 The performance of miRNA prediction strategy independent of genome reference

To our knowledge, most miRNA discovery tools based on analyzing small RNA sequencing datasets rely on the alignment of sequencing reads to reference genome sequences. Obviously, these tools have limited usage in the study of organisms whose genome has not been sequenced. Some other tools that do not depend on genome sequences make use of evolutionary conservation to identify the miRNAs with orthologs already found in other organisms, but will obviously miss the miRNAs specific to certain lineages. In contrast to these available tools, our miRNA discovery strategy takes advantage of the parallel sequencing of potential mature and precursor miRNAs. Our approach successfully identified 238 pre-miRNAs detected in our sample, corresponding to 35% of all mouse pre-miRNAs deposited in miRBase. Of 438 known mature miRNAs recovered by our approach, 48 are not conserved in other species and could not be identified only by homology search. It suggests that our approach can discover not only the well-conserved miRNAs, but also lineage-specific ones.

In probabilistic scoring of pre-miRNA candidates, we used the known mouse miRNAs to estimate the model parameters. In this case, it could be argued that our model was over-trained for predicting mouse miRNAs. If the performance is dependent on a particular training dataset, it will be questionable that our approach could work in an organism in which nothing is known about its miRNAs. To investigate the potential bias, we trained our model again using known human, fruit fly and *C. elegans*, respectively. The predictions overlap with each other very well (Figure 4.2), indicating that the features included in our model represent the miRNA characteristics common to all the organisms.

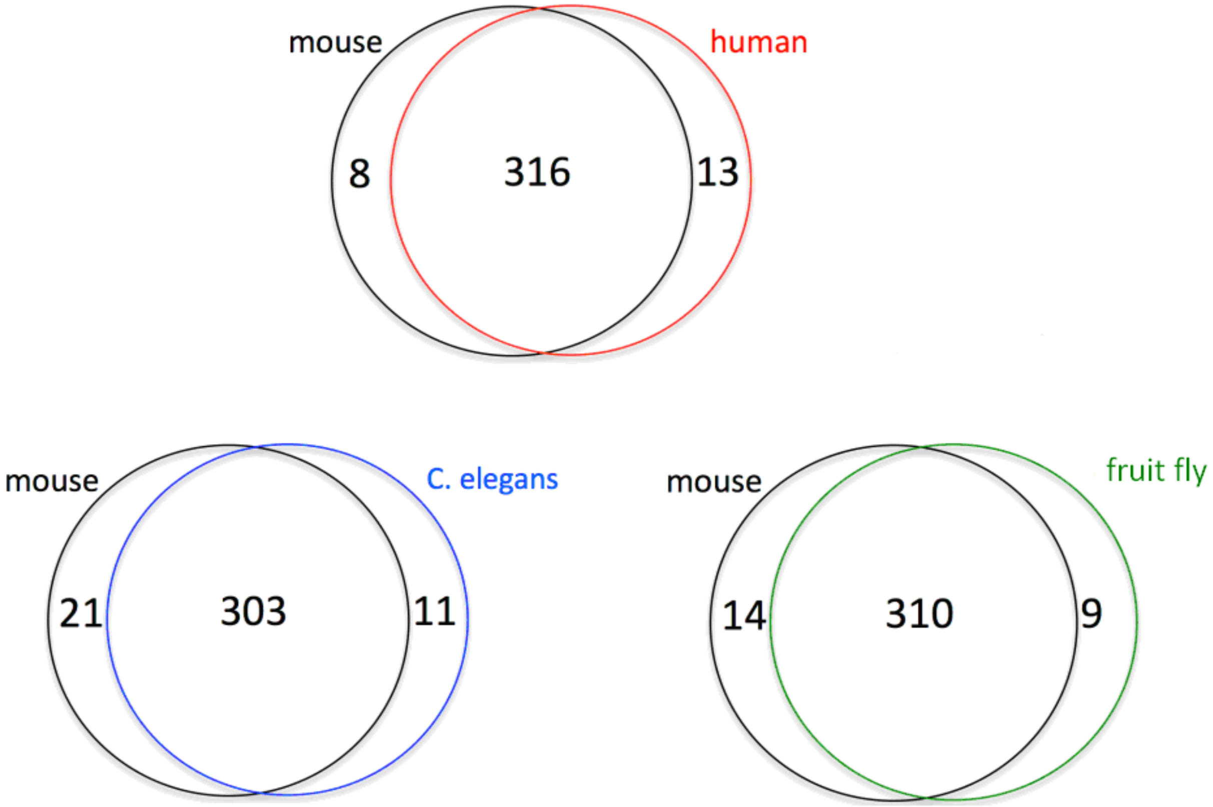


Figure 4.2 Prediction results were nearly identical using different training datasets from miRBase: mouse (black), human (red), *C.elegans* (blue) and fruit fly (green).

Furthermore, to assess whether our method could be applied in other metazoans, we sequenced mature and precursor miRNAs from *C.elegans*. We applied the pipeline to predict *C.elegans* miRNAs based on our sequencing dataset. In total, 91% (16,270,159) of short reads can be mapped to 97% (49,702,713) of long reads perfectly. After discarding the long reads in which the mapping position of short reads did not fit with the model of Dicer processing and clustering the long reads with the sequence similarity >90%, 47,052 long reads were selected to predict the secondary structure (see methods). Out of 6,007 long reads that can form stable hairpin structures, 187 potential pre-miRNAs passed the cut-off of 0.95 after probabilistic scoring. 126 candidate pre-miRNAs, on which at least five short reads could be mapped, were retained. Of these 126 pre-miRNA candidates, 99 corresponded to 88 known *C.elegans* pre-miRNAs. Manual inspection of the 27 novel candidates revealed that six were derived from chimerical reads representing ligation artifacts and one was derived from E.coli transcript representing potential contamination. Of the remaining 20 pre-miRNA candidates, one could be mapped to DNA transposable element, four candidates located to known rRNA transcripts. The remaining 15 predicted novel pre-miRNAs mapped to either intergenic (8), or exonic (4) and intronic (3) regions of protein coding genes. Using our computational pipeline, we could recover 80% of pre-miRNAs detected in our sample, corresponding to 50% of all known *C.elegans* miRNAs deposited in miRBase.

4.4 Exclusively 5' tailed miRtron

In total, 49 novel mouse pre-miRNAs predicted in this study located in introns of protein coding genes. Among these, 24 had both 5' and 3' end at least 10 nt away from the corresponding ends of the introns. Of the remaining 25 intron-containing pre-miRNAs, whereas six had the 'nearly' exact boundary as the hosting introns and thus resembled the canonical mirtrons, 19 had only one end generated by spliceosome while the other end likely matured through Drosha independent trimming. Interestingly, all the pre-miRNAs from the latter category were 5' tailed mirtron, which shared only their 3' ends with the hosting introns.

To investigate the tailing bias of tailed mirtrons in different organisims, we grouped the known human and mouse miRNA located in the intronic region into three categories according to the distance of 5' or 3' end of pre-miRNA to the neighboring exon-intron boundary, i.e. 1. Canonical mirtron, both 5' and 3' end of pre-miRNA are within 5 nt of exon-intron boundary. 2. miRNA in the middle of a large intron, both 5' and 3' end of pre-miRNA should be at least 5 nt far away from exon-intron boundary. 3. Tailed mirtron, only one end of pre-miRNA locates within 5 nt of exon-intron boundary. Table 4.1 shows the number of intronic miRNAs belonging to the three groups in mouse and human.

Interestingly, all mouse and most human mirtrons were 5' tailed (Table 4.2). This is consistent with our newly discovered tailed mirtrons. Indeed, we found the long reads possibly derived from the intermediate tailing products for several tailed mirtrons (Figure 4.3). In contrast to mouse, the tailed mirtrons identified so far in *Drosophila* are all from 3' end (Flynt, Greimann et al. 2010). It awaits further investigation whether the inconsistence between the two organisms is due to the difference in underlying processing mechanisms such as more efficient usage of 5’-3’ (mouse) versus 3’-5’ (fly) exoribonuclease after splicing.

Table 4.1. Intronic miRNAs in mouse and human. Type 1. Canonical mirtron, both 5’ and 3’ end of pre-miRNA are within 5 nt of exon-intron boundary. Type 2. miRNA in the middle of a large intron, both 5’ and 3’ end of pre-miRNA should be at least 5 nt far away from exon-intron boundary. Type 3. Tailed mirtron, only one end of pre-miRNA locates within 5 nt of exon-intron boundary.

	Mouse	Human
Type 1	4	7
Type 2	298	616
Type 3	21	35
Total	323	658

Table 4.2 The number of type 3 intronic miRNA either 5’ end or 3’ end of hairpin coincides the splicing-junction.

	Mouse	Human
5’ end	0	8
3’ end	21	27
Total	21	35

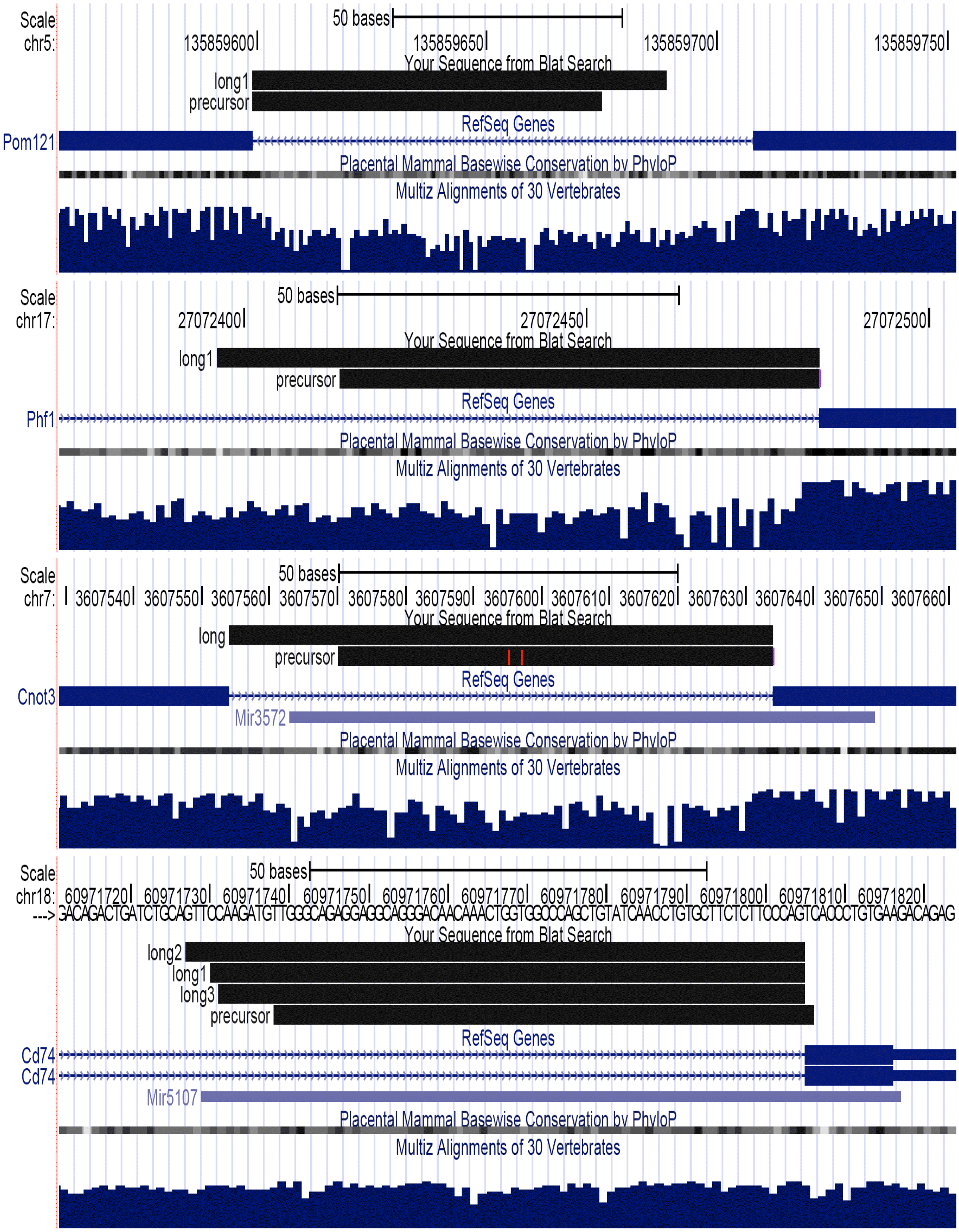

Scale
chr5:
50 bases
135859600
135859650
135859700
135859750
Your Sequence from Blat Search
long1
precursor
RefSeq Genes
Pom121
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates
Scale
chr17:
50 bases
27072400
27072450
27072500
Your Sequence from Blat Search
long1
precursor
RefSeq Genes
Phf1
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates
Scale
chr7:
50 bases
3607540 3607550 3607560 3607570 3607580 3607590 3607600 3607610 3607620 3607630 3607640 3607650 3607660
Your Sequence from Blat Search
long
precursor
RefSeq Genes
Cnot3
Mir3572
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates
Scale
chr18:
50 bases
60971720 60971730 60971740 60971750 60971760 60971770 60971780 60971790 60971800 60971810 60971820
---> GACAGACTGATCTGCAGTTCCAAGATGTTGGGCAGAGGAGGCAGGGACAACAAACTGGTGGCCCAGCTGTATCAACCTGTGCTTCTCTTCCCAGTCACCCTGTGAAGACAGAG
Your Sequence from Blat Search
long2
long1
long3
precursor
RefSeq Genes
Cd74
Cd74
Mir5107
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates

Figure 4.3 Intermediates of mirtron processing for several identified 5' tailed mirtions. "precursor" denotes the predicted pre-miRNA sequence, and "long" denotes full introns sequence or intron processing intermediates.

In summary, we have performed the first unbiased genome-wide parallel profiling of mature and precursor miRNAs. Compared with mature miRNA sequencing, our pre-miRNA sequencing has rather limited efficiency and awaits further technical improvement. However, even with the current dataset, we could improve the understanding of the miRNA processing and modification. More importantly, we developed a novel miRNA discovery strategy that does not rely on the available genome reference sequences. We believe our method could be widely used in the study of miRNAs not only in the organisms whose genome has not yet been sequenced, but also in samples where the genome differs significantly from the reference sequences, such as cancer.

5 REFERENCES

Abrahante, J. E., A. L. Daul, et al. (2003). "The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs." Dev Cell **4**(5): 625-37.

Adai, A., C. Johnson, et al. (2005). "Computational prediction of miRNAs in Arabidopsis thaliana." Genome Res **15**(1): 78-91.

Aravin, A. A., M. Lagos-Quintana, et al. (2003). "The small RNA profile during Drosophila melanogaster development." Dev Cell **5**(2): 337-50.

Aza-Blanc, P., C. L. Cooper, et al. (2003). "Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening." Mol Cell **12**(3): 627-37.

Azuma-Mukai, A., H. Oguri, et al. (2008). "Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing." Proc Natl Acad Sci U S A **105**(23): 7964-9.

Babiarz, J. E., J. G. Ruby, et al. (2008). "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs." Genes Dev **22**(20): 2773-85.

Baccarini, A., H. Chauhan, et al. (2011). "Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells." Curr Biol **21**(5): 369-76.

Baek, D., J. Villen, et al. (2008). "The impact of microRNAs on protein output." Nature **455**(7209): 64-71.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-33.

Baskerville, S. and D. P. Bartel (2005). "Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes." RNA **11**(3): 241-7.

Basyuk, E., F. Suavet, et al. (2003). "Human let-7 stem-loop precursors harbor features of RNase III cleavage products." Nucleic Acids Res **31**(22): 6593-7.

Ben-Ami, O., N. Pencovich, et al. (2009). "A regulatory interplay between miR-27a and Runx1 during megakaryopoiesis." Proc Natl Acad Sci U S A **106**(1): 238-43.

Bentwich, I., A. Avniel, et al. (2005). "Identification of hundreds of conserved and nonconserved human microRNAs." Nat Genet **37**(7): 766-70.

Berezikov, E., W. J. Chung, et al. (2007). "Mammalian mirtron genes." Mol Cell **28**(2): 328-36.

Berezikov, E., E. Cuppen, et al. (2006). "Approaches to microRNA discovery." Nat Genet **38 Suppl**: S2-7.

Berezikov, E., N. Robine, et al. (2010). "Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence." Genome Res **21**(2): 203-15.

Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-6.

Bonnet, E., J. Wuyts, et al. (2004). "Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences." Bioinformatics **20**(17): 2911-7.

Borchert, G. M., W. Lanier, et al. (2006). "RNA polymerase III transcribes human microRNAs." Nat Struct Mol Biol **13**(12): 1097-101.

Bracht, J., S. Hunter, et al. (2004). "Trans-splicing and polyadenylation of let-7 microRNA primary transcripts." RNA **10**(10): 1586-94.

Brameier, M., A. Herwig, et al. (2011). "Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs." Nucleic Acids Res **39**(2): 675-86.

Burroughs, A. M., Y. Ando, et al. (2010). "A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness." Genome Res **20**(10): 1398-410.

Bushati, N. and S. M. Cohen (2007). "microRNA functions." Annu Rev Cell Dev Biol **23**: 175-205.

Cai, X., C. H. Hagedorn, et al. (2004). "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." RNA **10**(12): 1957-66.

Calabrese, J. M., A. C. Seila, et al. (2007). "RNA sequence analysis defines Dicer's role in mouse embryonic stem cells." Proc Natl Acad Sci U S A **104**(46): 18097-102.

Carmell, M. A., Z. Xuan, et al. (2002). "The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis." Genes Dev **16**(21): 2733-42.

Carthew, R. W. and E. J. Sontheimer (2009). "Origins and Mechanisms of miRNAs and siRNAs." Cell **136**(4): 642-55.

Cerutti, L., N. Mian, et al. (2000). "Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain." Trends Biochem Sci **25**(10): 481-2.

Chalfie, M., H. R. Horvitz, et al. (1981). "Mutations that lead to reiterations in the cell lineages of C. elegans." Cell **24**(1): 59-69.

Chan, C. S., O. Elemento, et al. (2005). "Revealing posttranscriptional regulatory elements through network-level conservation." PLoS Comput Biol **1**(7): e69.

Chang, T. C. and J. T. Mendell (2007). "microRNAs in vertebrate physiology and human disease." Annu Rev Genomics Hum Genet **8**: 215-39.

Chang, T. C., D. Yu, et al. (2008). "Widespread microRNA repression by Myc contributes to tumorigenesis." Nat Genet **40**(1): 43-50.

Chatterjee, S. and H. Grosshans (2009). "Active turnover modulates mature microRNA activity in Caenorhabditis elegans." Nature **461**(7263): 546-9.

Cheloufi, S., C. O. Dos Santos, et al. (2010). "A dicer-independent miRNA biogenesis pathway that requires Ago catalysis." Nature **465**(7298): 584-9.

Chen, P. Y., H. Manninga, et al. (2005). "The developmental miRNA profiles of zebrafish as determined by small RNA cloning." Genes Dev **19**(11): 1288-93.

Chendrimada, T. P., K. J. Finn, et al. (2007). "MicroRNA silencing through RISC recruitment of eIF6." Nature **447**(7146): 823-8.

Chendrimada, T. P., R. I. Gregory, et al. (2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing." Nature **436**(7051): 740-4.

Chiang, H. R., L. W. Schoenfeld, et al. (2010). "Mammalian microRNAs: experimental evaluation of novel and previously annotated genes." Genes Dev **24**(10): 992-1009.

Cifuentes, D., H. Xue, et al. (2010). "A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity." Science **328**(5986): 1694-8.

Czech, B., R. Zhou, et al. (2009). "Hierarchical rules for Argonaute loading in Drosophila." Mol Cell **36**(3): 445-56.

Davis, B. N., A. C. Hilyard, et al. (2008). "SMAD proteins control DROSHA-mediated microRNA maturation." Nature **454**(7200): 56-61.

Denli, A. M., B. B. Tops, et al. (2004). "Processing of primary microRNAs by the Microprocessor complex." Nature **432**(7014): 231-5.

Derry, M. C., A. Yanagiya, et al. (2006). "Regulation of poly(A)-binding protein through PABP-interacting proteins." Cold Spring Harb Symp Quant Biol **71**: 537-43.

Diederichs, S. and D. A. Haber (2007). "Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression." Cell **131**(6): 1097-108.

Ender, C., A. Krek, et al. (2008). "A human snoRNA with microRNA-like functions." Mol Cell **32**(4): 519-28.

Esquela-Kerscher, A. and F. J. Slack (2006). "Oncomirs - microRNAs with a role in cancer." Nat Rev Cancer **6**(4): 259-69.

Flynt, A. S., J. C. Greimann, et al. (2010). "MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila." Mol Cell **38**(6): 900-7.

Forman, J. J., A. Legesse-Miller, et al. (2008). "A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence." Proc Natl Acad Sci U S A **105**(39): 14879-84.

Friedlander, M. R., W. Chen, et al. (2008). "Discovering microRNAs from deep sequencing data using miRDeep." Nat Biotechnol **26**(4): 407-15.

Fukuda, T., K. Yamagata, et al. (2007). "DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs." Nat Cell Biol **9**(5): 604-11.

Gantier, M. P., C. E. McCoy, et al. (2011). "Analysis of microRNA turnover in mammalian cells following Dicer1 ablation." Nucleic Acids Res **39**(13): 5692-703.

Ghildiyal, M., J. Xu, et al. (2010). "Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway." RNA **16**(1): 43-56.

Gregory, R. I., T. P. Chendrimada, et al. (2005). "Human RISC couples microRNA biogenesis and posttranscriptional gene silencing." Cell **123**(4): 631-40.

Gregory, R. I., K. P. Yan, et al. (2004). "The Microprocessor complex mediates the genesis of microRNAs." Nature **432**(7014): 235-40.

Grishok, A., A. E. Pasquinelli, et al. (2001). "Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing." Cell **106**(1): 23-34.

Guil, S. and J. F. Caceres (2007). "The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a." Nat Struct Mol Biol **14**(7): 591-6.

Guo, H., N. T. Ingolia, et al. (2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels." Nature **466**(7308): 835-40.

Gwizdek, C., B. Ossareh-Nazari, et al. (2003). "Exportin-5 mediates nuclear export of minihelix-containing RNAs." J Biol Chem **278**(8): 5505-8.

Haase, A. D., L. Jaskiewicz, et al. (2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing." EMBO Rep **6**(10): 961-7.

Hackenberg, M., M. Sturm, et al. (2009). "miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments." Nucleic Acids Res **37**(Web Server issue): W68-76.

Han, J., Y. Lee, et al. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes Dev **18**(24): 3016-27.

Han, J., Y. Lee, et al. (2006). "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex." Cell **125**(5): 887-901.

Han, J., J. S. Pedersen, et al. (2009). "Posttranscriptional crossregulation between Drosha and DGCR8." Cell **136**(1): 75-84.

He, L. and G. J. Hannon (2004). "MicroRNAs: small RNAs with a big role in gene regulation." Nat Rev Genet **5**(7): 522-31.

He, L., X. He, et al. (2007). "A microRNA component of the p53 tumour suppressor network." Nature **447**(7148): 1130-4.

He, L., J. M. Thomson, et al. (2005). "A microRNA polycistron as a potential human oncogene." Nature **435**(7043): 828-33.

Hendrickson, D. G., D. J. Hogan, et al. (2009). "Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA." PLoS Biol **7**(11): e1000238.

Hendrix, D., M. Levine, et al. (2010). "miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data." Genome Biol **11**(4): R39.

Heo, I., C. Joo, et al. (2008). "Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA." Mol Cell **32**(2): 276-84.

Heo, I., C. Joo, et al. (2009). "TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation." Cell **138**(4): 696-708.

Hofacker, I. L. and P. F. Stadler (2006). "Memory efficient folding algorithms for circular RNA secondary structures." Bioinformatics **22**(10): 1172-6.

Humphreys, D. T., B. J. Westman, et al. (2005). "MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function." Proc Natl Acad Sci U S A **102**(47): 16961-6.

Huntzinger, E. and E. Izaurralde (2011). "Gene silencing by microRNAs: contributions of translational repression and mRNA decay." Nat Rev Genet **12**(2): 99-110.

Hutvagner, G., J. McLachlan, et al. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." Science **293**(5531): 834-8.

Hwang, H. W., E. A. Wentzel, et al. (2007). "A hexanucleotide element directs microRNA nuclear import." Science **315**(5808): 97-100.

Jones, M. R., L. J. Quinton, et al. (2009). "Zcchc11-dependent uridylation of microRNA directs cytokine expression." Nat Cell Biol **11**(9): 1157-63.

Katoh, T., Y. Sakaguchi, et al. (2009). "Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2." Genes Dev **23**(4): 433-8.

Kawahara, Y., M. Megraw, et al. (2008). "Frequency and fate of microRNA editing in human brain." Nucleic Acids Res **36**(16): 5270-80.

Kawahara, Y., B. Zinshteyn, et al. (2007). "RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex." EMBO Rep **8**(8): 763-9.

Kawahara, Y., B. Zinshteyn, et al. (2007). "Redirection of silencing targets by adenosine-to-inosine editing of miRNAs." Science **315**(5815): 1137-40.

Ketting, R. F., S. E. Fischer, et al. (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." Genes Dev **15**(20): 2654-9.

Khvorova, A., A. Reynolds, et al. (2003). "Functional siRNAs and miRNAs exhibit strand bias." Cell **115**(2): 209-16.

Kim, J., K. Inoue, et al. (2007). "A MicroRNA feedback circuit in midbrain dopamine neurons." Science **317**(5842): 1220-4.

Kim, V. N., J. Han, et al. (2009). "Biogenesis of small RNAs in animals." Nat Rev Mol Cell Biol **10**(2): 126-39.

Kim, Y. K. and V. N. Kim (2007). "Processing of intronic microRNAs." EMBO J **26**(3): 775-83.

Kiriakidou, M., G. S. Tan, et al. (2007). "An mRNA m7G cap binding-like motif within human Ago2 represses translation." Cell **129**(6): 1141-51.

Kloosterman, W. P. and R. H. Plasterk (2006). "The diverse functions of microRNAs in animal development and disease." Dev Cell **11**(4): 441-50.

Knight, S. W. and B. L. Bass (2001). "A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans." Science **293**(5538): 2269-71.

Kok, K. H., M. H. Ng, et al. (2007). "Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA." J Biol Chem **282**(24): 17649-57.

Kozomara, A. and S. Griffiths-Jones (2011). "miRBase: integrating microRNA annotation and deep-sequencing data." Nucleic Acids Res **39**(Database issue): D152-7.

Krek, A., D. Grun, et al. (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.

Krol, J., V. Busskamp, et al. (2010). "Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs." Cell **141**(4): 618-31.

Krol, J., I. Loedige, et al. (2010). "The widespread regulation of microRNA biogenesis, function and decay." Nat Rev Genet **11**(9): 597-610.

Krutzfeldt, J., N. Rajewsky, et al. (2005). "Silencing of microRNAs in vivo with 'antagomirs'." Nature **438**(7068): 685-9.

Krutzfeldt, J. and M. Stoffel (2006). "MicroRNAs: a new class of regulatory genes affecting metabolism." Cell Metab **4**(1): 9-12.

Kuchenbauer, F., R. D. Morin, et al. (2008). "In-depth characterization of the microRNA transcriptome in a leukemia progression model." Genome Res **18**(11): 1787-97.

Lagos-Quintana, M., R. Rauhut, et al. (2001). "Identification of novel genes coding for small expressed RNAs." Science **294**(5543): 853-8.

Landgraf, P., M. Rusu, et al. (2007). "A mammalian microRNA expression atlas based on small RNA library sequencing." Cell **129**(7): 1401-14.

Landthaler, M., A. Yalcin, et al. (2004). "The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis." Curr Biol **14**(23): 2162-7.

Lau, N. C., L. P. Lim, et al. (2001). "An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans." Science **294**(5543): 858-62.

Lee, R. C. and V. Ambros (2001). "An extensive class of small RNAs in Caenorhabditis elegans." Science **294**(5543): 862-4.

Lee, R. C., R. L. Feinbaum, et al. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-54.

Lee, Y., C. Ahn, et al. (2003). "The nuclear RNase III Drosha initiates microRNA processing." Nature **425**(6956): 415-9.

Lee, Y., I. Hur, et al. (2006). "The role of PACT in the RNA silencing pathway." EMBO J **25**(3): 522-32.

Lee, Y., K. Jeon, et al. (2002). "MicroRNA maturation: stepwise processing and subcellular localization." EMBO J **21**(17): 4663-70.

Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." EMBO J **23**(20): 4051-60.

Lewis, B. P., C. B. Burge, et al. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell **120**(1): 15-20.

Li, R., Y. Li, et al. (2008). "SOAP: short oligonucleotide alignment program." Bioinformatics **24**(5): 713-4.

Li, R., C. Yu, et al. (2009). "SOAP2: an improved ultrafast tool for short read alignment." Bioinformatics **25**(15): 1966-7.

Lim, L. P., N. C. Lau, et al. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-73.

Liu, J., M. A. Carmell, et al. (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-41.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-8.

Lujambio, A., G. A. Calin, et al. (2008). "A microRNA DNA methylation signature for human cancer metastasis." Proc Natl Acad Sci U S A **105**(36): 13556-61.

Lujambio, A. and M. Esteller (2009). "How epigenetics can explain human metastasis: a new role for microRNAs." Cell Cycle **8**(3): 377-82.

Ma, L., J. Young, et al. (2010). "miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis." Nat Cell Biol **12**(3): 247-56.

MacRae, I. J., K. Zhou, et al. (2007). "Structural determinants of RNA recognition and cleavage by Dicer." Nat Struct Mol Biol **14**(10): 934-40.

Macrae, I. J., K. Zhou, et al. (2006). "Structural basis for double-stranded RNA processing by Dicer." Science **311**(5758): 195-8.

Maniataki, E. and Z. Mourelatos (2005). "A human, ATP-independent, RISC assembly machine fueled by pre-miRNA." Genes Dev **19**(24): 2979-90.

Maroney, P. A., Y. Yu, et al. (2006). "Evidence that microRNAs are associated with translating messenger RNAs in human cells." Nat Struct Mol Biol **13**(12): 1102-7.

Megraw, M., P. Sethupathy, et al. (2007). "miRGen: a database for the study of animal microRNA genomic organization and function." Nucleic Acids Res **35**(Database issue): D149-55.

Meister, G., M. Landthaler, et al. (2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Mol Cell **15**(2): 185-97.

Melo, S. A., S. Ropero, et al. (2009). "A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function." Nat Genet **41**(3): 365-70.

Miska, E. A., E. Alvarez-Saavedra, et al. (2004). "Microarray analysis of microRNA expression in the developing mammalian brain." Genome Biol **5**(9): R68.

Mourelatos, Z., J. Dostie, et al. (2002). "miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs." Genes Dev **16**(6): 720-8.

Nelson, P. T., D. A. Baldwin, et al. (2004). "Microarray-based, high-throughput gene expression profiling of microRNAs." Nat Methods **1**(2): 155-61.

Newman, M. A., J. M. Thomson, et al. (2008). "Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing." RNA **14**(8): 1539-49.

Nishikura, K. (2010). "Functions and regulation of RNA editing by ADAR deaminases." Annu Rev Biochem **79**: 321-49.

Nottrott, S., M. J. Simard, et al. (2006). "Human let-7a miRNA blocks protein production on actively translating polyribosomes." Nat Struct Mol Biol **13**(12): 1108-14.

O'Donnell, K. A., E. A. Wentzel, et al. (2005). "c-Myc-regulated microRNAs modulate E2F1 expression." Nature **435**(7043): 839-43.

Okada, C., E. Yamashita, et al. (2009). "A high-resolution structure of the pre-microRNA nuclear export machinery." Science **326**(5957): 1275-9.

Okamura, K., J. W. Hagen, et al. (2007). "The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila." Cell **130**(1): 89-100.

Okamura, K., N. Liu, et al. (2009). "Distinct mechanisms for microRNA strand selection by Drosophila Argonautes." Mol Cell **36**(3): 431-44.

Olsen, P. H. and V. Ambros (1999). "The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation." Dev Biol **216**(2): 671-80.

Ono, M., M. S. Scott, et al. (2011). "Identification of human miRNA precursors that resemble box C/D snoRNAs." Nucleic Acids Res **39**(9): 3879-91.

Paroo, Z., X. Ye, et al. (2009). "Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling." Cell **139**(1): 112-22.

Pasquinelli, A. E., B. J. Reinhart, et al. (2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA." Nature **408**(6808): 86-9.

Peters, L. and G. Meister (2007). "Argonaute proteins: mediators of RNA silencing." Mol Cell **26**(5): 611-23.

Petersen, C. P., M. E. Bordeleau, et al. (2006). "Short RNAs repress translation after initiation in mammalian cells." Mol Cell **21**(4): 533-42.

Pillai, R. S., C. G. Artus, et al. (2004). "Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis." RNA **10**(10): 1518-25.

Pillai, R. S., S. N. Bhattacharyya, et al. (2005). "Inhibition of translational initiation by Let-7 MicroRNA in human cells." Science **309**(5740): 1573-6.

Piriyapongsa, J. and I. K. Jordan (2007). "A family of human microRNA genes from miniature inverted-repeat transposable elements." PLoS One **2**(2): e203.

Piskounova, E., S. R. Viswanathan, et al. (2008). "Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28." J Biol Chem **283**(31): 21310-4.

Ramachandran, V. and X. Chen (2008). "Degradation of microRNAs by a family of exoribonucleases in Arabidopsis." Science **321**(5895): 1490-2.

Reinhart, B. J., F. J. Slack, et al. (2000). "The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans." Nature **403**(6772): 901-6.

Ruby, J. G., C. Jan, et al. (2006). "Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans." Cell **127**(6): 1193-207.

Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." Nature **448**(7149): 83-6.

Ruby, J. G., A. Stark, et al. (2007). "Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs." Genome Res **17**(12): 1850-64.

Rybak, A., H. Fuchs, et al. (2008). "A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment." Nat Cell Biol **10**(8): 987-93.

Saraiya, A. A. and C. C. Wang (2008). "snoRNA, a novel precursor of microRNA in Giardia lamblia." PLoS Pathog **4**(11): e1000224.

Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell **115**(2): 199-208.

Seggerson, K., L. Tang, et al. (2002). "Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation." Dev Biol **243**(2): 215-25.

Selbach, M., B. Schwanhausser, et al. (2008). "Widespread changes in protein synthesis induced by microRNAs." Nature **455**(7209): 58-63.

Shiohama, A., T. Sasaki, et al. (2003). "Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region." Biochem Biophys Res Commun **304**(1): 184-90.

Smalheiser, N. R. and V. I. Torvik (2005). "Mammalian microRNAs derived from genomic repeats." Trends Genet **21**(6): 322-6.

Song, J. J., S. K. Smith, et al. (2004). "Crystal structure of Argonaute and its implications for RISC slicer activity." Science **305**(5689): 1434-7.

Taft, R. J., E. A. Glazov, et al. (2009). "Small RNAs derived from snoRNAs." RNA **15**(7): 1233-40.

Thomson, J. M., J. Parker, et al. (2004). "A custom microarray platform for analysis of microRNA gene expression." Nat Methods **1**(1): 47-53.

Tokumaru, S., M. Suzuki, et al. (2008). "let-7 regulates Dicer expression and constitutes a negative feedback loop." Carcinogenesis **29**(11): 2073-7.

Tolia, N. H. and L. Joshua-Tor (2007). "Slicer and the argonautes." Nat Chem Biol **3**(1): 36-43.

Underwood, J. G., A. V. Uzilov, et al. (2010). "FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing." Nat Methods **7**(12): 995-1001.

Vella, M. C., E. Y. Choi, et al. (2004). "The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR." Genes Dev **18**(2): 132-7.

Viswanathan, S. R., G. Q. Daley, et al. (2008). "Selective blockade of microRNA processing by Lin28." Science **320**(5872): 97-100.

Voinnet, O. (2009). "Origin, biogenesis, and activity of plant microRNAs." Cell **136**(4): 669-87.

Winter, J., S. Jung, et al. (2009). "Many roads to maturity: microRNA biogenesis pathways and their regulation." Nat Cell Biol **11**(3): 228-34.

Xie, X., J. Lu, et al. (2005). "Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals." Nature **434**(7031): 338-45.

Yamagishi, H. and D. Srivastava (2003). "Unraveling the genetic and developmental mysteries of 22q11 deletion syndrome." Trends Mol Med **9**(9): 383-9.

Yeom, K. H., Y. Lee, et al. (2006). "Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing." Nucleic Acids Res **34**(16): 4622-9.

Zeng, Y. and B. R. Cullen (2004). "Structural requirements for pre-microRNA binding and nuclear export by Exportin 5." Nucleic Acids Res **32**(16): 4776-85.

Zhang, H., F. A. Kolb, et al. (2004). "Single processing center models for human Dicer and bacterial RNase III." Cell **118**(1): 57-68.

Zhao, H., A. Kalota, et al. (2009). "The c-myb proto-oncogene and microRNA-15a comprise an active autoregulatory feedback loop in human hematopoietic cells." Blood **113**(3): 505-16.

6 APPENDIX

Appendix table 1 The 5’ or 3’ end of 13 miRNAs in miRBase v16.0 were manually corrected

mmu-mir-181a-2

```
   u  a   u      cu       a    gggauuc
cca gg aca ucaacg  gucggug guuu       a
||| || ||| ||||||  ||||||| ||||       a
ggu cc ugu aguugc  cagccac caaa       a
   u  a   c      --       -    aaaacaa
```

↓

```
   u  a   u      cu       a    gggauuc
cca gg aca ucaacg  gucggug guuu       a
||| || ||| ||||||  ||||||| ||||       a
ggu cc ugu aguugc  cagccac caaa       a
   u  a   c      --       -    aaaacaa
```

mmu-mir-153

```
     -             gu       -a  aa
cggug ucauuuuugugac  ugcagcu  gu  u
||||| |||||||||||||  |||||||  ||  a
guuac agugaaaacacug  acguuga  cg  u
     u             au       cc  ag
```

↓

```
     -             gu       -a  aa
cggug ucauuuuugugac  ugcagcu  gu  u
||||| |||||||||||||  |||||||  ||  a
guuac agugaaaacacug  acguuga  cg  u
     u             au       cc  ag
```

mmu-mir-138-1

```
cucua   ug    u   -  a ag             uca       gc a
     gca  gugu gug gg c  cugguguugugaa   ggccguu  c a
     |||  |||| ||| || |  |||||||||||||   |||||||  | u
     cgu  cacg cac cc g  gaccacaacacuu   ucggcaa  g c
--gaa   --    u   a  - -g             -ca       ga a
```

↓

```
cucua   ug    u   -  a ag             uca       gc a
     gca  gugu gug gg c  cugguguugugaa   ggccguu  c a
     |||  |||| ||| || |  |||||||||||||   |||||||  | u
     cgu  cacg cac cc g  gaccacaacacuu   ucggcaa  g c
--gaa   --    u   a  - -g             -ca       ga a
```

mmu-mir-337

```
caguguagugagaa uu         -   c         c         u  ca
              g  ggggggugg gaa ggcgucaug aggaguuga ug  c
              |  ||||||||| ||| ||||||||| ||||||||| ||
              c  ucccccacu cuu ccguaguau uccucgacu ac  a
-------------a -u         u   u         a         u  cg
```

↓

```
caguguagugagaa uu         -   c         c         u  ca
              g  ggggggugg gaa ggcgucaug aggaguuga ug  c
              |  ||||||||| ||| ||||||||| ||||||||| ||
              c  ucccccacu cuu ccguaguau uccucgacu ac  a
-------------a -u         u   u         a         u  cg
```

mmu-mir-344g

```
-   c               a    u    cuc  a
 ugc agucaggcuccuggc ggag ccug   uc g
 ||| ||||||||||||||| |||| ||||   ||
 acg ucaguucggggaccg ucuc ggac   gg c
a   -               a    -    cuu  u
```

↓

```
-   c               a    u    cuc  a
 ugc agucaggcuccuggc ggag ccug   uc g
 ||| ||||||||||||||| |||| ||||   ||
 acg ucaguucggggaccg ucuc ggac   gg c
a   -               a    -    cuu  u
```

mmu-mir-374

```
gaa  aau  ---ua      ug au                     u
   ga   cc     cucggg  g  auaauacaaccugcuaagugu c
   ||   ||     ||||||  |  |||||||||||||||||||||
   cu   gg     gagccu  u  uauuauguuggacgauucacg u
cug  cuc  uauug      gu ac                     a
```

↓

```
gaa  aau  ---ua      ug au                     u
   ga   cc     cucggg  g  auaauacaaccugcuaagugu c
   ||   ||     ||||||  |  |||||||||||||||||||||
   cu   gg     gagccu  u  uauuauguuggacgauucacg u
cug  cuc  uauug      gu ac                     a
```

mmu-mir-382

```
    u       -a          ug      - uuu
uacu gaagaga  guuguucgugg  gauucg c   a
|||| |||||||  |||||||||||  |||||| |   c
auga cuuuuuu  caacaggcacu  cuaagc g   u
    -       ca          ua      a ugu
```

↓

```
    u       -a          ug      - uuu
uacu gaagaga  guuguucgugg  gauucg c   a
|||| |||||||  |||||||||||  |||||| |   c
auga cuuuuuu  caacaggcacu  cuaagc g   u
    -       ca          ua      a ugu
```

mmu-mir-452

```
  u     g    aa  g        g  a        uuua
gc aagca uuac  cu uuugcaga ga acugagac    u
|| ||||| ||||  || |||||||| || ||||||||    a
cg uucgu aaug  ga aaacgucu cu ugacucug    a
  u     g    --  g        a  c        uauc
```

↓

```
  u     g    aa  g        g  a        uuua
gc aagca uuac  cu uuugcaga ga acugagac    u
|| ||||| ||||  || |||||||| || ||||||||    a
cg uucgu aaug  ga aaacgucu cu ugacucug    a
  u     g    --  g        a  c        uauc
```

mmu-mir-1957

```
--    aaagu       u   aga       a
  cugg     agcucag ggu   gcauaug c
  ||||     ||||||| |||   ||||||| c
  gacc     uugaguc ccg   cguguac a
ac    ccuau       -   gaa       g
```

↓

```
--    aaagu       u   aga       a
  cugg     agcucag ggu   gcauaug c
  ||||     ||||||| |||   ||||||| c
  gacc     uugaguc ccg   cguguac a
ac    ccuau       -   gaa       g
```

mmu-mir-496

```
     u    u        c     gu         uua
agugu cgaa ggagguug ccaug  guguucauu   u
||||| |||| |||||||| |||||  |||||||||
ucacg gcuu ccucuaac gguac  uaugaguag   u
     -    u        c     au         uau
```

↓

```
     u    u        c     gu         uua
agugu cgaa ggagguug ccaug  guguucauu   u
||||| |||| |||||||| |||||  |||||||||
ucacg gcuu ccucuaac gguac  uaugaguag   u
     -    u        c     au         uau
```

mmu-mir-582

```
-----         aua            ca     uaaucu
     acucuuugg   caguuguucaac  guuac      a
     |||||||||   ||||||||||||  |||||
     ugggaaacc   gucaacaaguug  caaug      a
aaacg         caa            uc     uuaauc
```

↓

```
-----         aua            ca     uaaucu
     acucuuugg   caguuguucaac  guuac      a
     |||||||||   ||||||||||||  |||||
     ugggaaacc   gucaacaaguug  caaug      a
aaacg         caa            uc     uuaauc
```

mmu-mir-653

```
--       u              u     c     caa   c
  cauucuu caguguugaaacaa cucua ugaac   gcu c
  ||||||| |||||||||||||| ||||| |||||   |||
  gugagga guuaugacuuuguu gaggu acuug   cga a
uc       c              u     c     -ag   a

                     ↓

--       u              u     c     caa   c
  cauucuu caguguugaaacaa cucua ugaac   gcu c
  ||||||| |||||||||||||| ||||| |||||   |||
  gugagga guuaugacuuuguu gaggu acuug   cga a
uc       c              u     c     -ag   a
```

mmu-mir-669m-2

```
--     --       g      gu        c     c        au
  auauu  ugcaugu uguaua  uuugugug augug augugugu  a
  |||||  ||||||| ||||||  |||||||| ||||| ||||||||
  uauag  acguacg acauau  aaacacac uacau uacauaua  u
ca     ac       a      ac        c     a        ag

                               ↓

--     --       g      gu        c     c        au
  auauu  ugcaugu uguaua  uuugugug augug augugugu  a
  |||||  ||||||| ||||||  |||||||| ||||| ||||||||
  uauag  acguacg acauau  aaacacac uacau uacauaua  u
ca     ac       a      ac        c     a        ag
```

Appendix table 2 Detailed information about 69 novel candidates

```
seq-1   No. long read =1      mmu17-mir-3572
UUGGGGAACAGGGCAAGGUGGACAGCAUCUGACAGCCUGUUUACACUUGUCCUUCUUUCCCCAGU
((((((((.(((((((((((((((((........)).))))))).))))))))...)))))))). No. short read
((((((((.((((((((((((                                             4
((((((((.(((((((((((((                                            7
((((((((.((((((((((((((                                           1
 (((((((.((((((((((((                                             3
 (((((((.(((((((((((((                                            2
 (((((((.((((((((((((((                                           2
                                        )))).))))))))...))))))    1
                                        )))).))))))))...)))))))   2
                                        )))).))))))))...))))))))  3
                                        )))).))))))))...)))))))). 2
                                         ))).))))))))...)))))     1
                                         ))).))))))))...))))))))  1

seq-2   No. long read =1      mmu17-mir-5123
UGUAGAUCCAUAUGCCAUGGUGUGUAGACCCAUAUGCCAUGGUGUGUAGAUCCAUAUG
(((.((((.((((((((((((((((......)))))))))))))))).)))).))).. No. short read
(((.((((.(((((((((((((                                     9
(((.((((.((((((((((((((                                    7
(((.((((.((((((((((((((((                                  1
  (.((((.((((((((((((((((                                  1
     (((.(((((((((((((                                     2
     (((.(((((((((((((((                                   1
                                  ))))))))))))).)))).      1
                                  ))))))))))))).)))).)     2
```

```
seq-3   No. long read =1     mmu17-mir-5104
UGAGGCAUCUCUCUAGCUCCAGAGAGCACGGUUUUAUAACCAAUUGUUCUGUGCUAGUGAGGUGGCUCAG
((((.((((((.(((((..(((((.....((((....)))).....))))).))))).)))))).)))).    No. short read
((((.((((((.(((((..(((                                                    1
((((.((((((.(((((..((((                                                   1
                                               )))).))))).)))))).         1
                                               )))).))))).)))))).)        2
                                               )))).))))).)))))).))       2
                                               )))).))))).)))))).)))      4
                                               )))).))))).)))))).))))     3
                                                 )).))))).)))))).)))).    1

seq-4   No. long read =3    mmu17-mir-5107
UUGGGCAGAGGAGGCAGGGACAACAAACUGGUGGCCCAGCUGUAUCAACCUGUGCUUCUCUUCCCAGU
(((((.((((((((((((....(((..((((....)))).))).....))))).))))))).)))))).      No. short read
(((((.((((((((((((...                                                     2
(((((.((((((((((((....                                                    2
 ((((.((((((((((((...                                                     1
 ((((.((((((((((((....                                                    1
                                             ...))))).))))))).)))))       2
                                             ...))))).))))))).))))).      8
                                               .))))).))))))).))))).      1

seq-5   No. long read =1
UCCUUCACUAGCUGAGACCUGAGCUUGGCCAGUGUGCUGCAACAACUCUGCUAGUGGAGAGA
..((((((((((.(((..(((.((...)))))(((......))).))).))))))))))...    No. short read
..((((((((((.(((..(((.                                            5
..((((((((((.(((..(((.(                                           1
..((((((((((.(((..(((.((                                          2
                                       ..))).))).))))))))))..     1
                                        .))).))).)))))))))).      1
                                        .))).))).))))))))))..     2
                                        .))).))).))))))))))...    8
                                         ))).))).))))))))))...    3

seq-6   No. long read =1
GUGAAUAUUGAAUUCGGGAGCGGCCAGCCUGGCUAUUUAGCUGGCCCUGCAACUUCCCGCAGCUCUCCUUCAGUUCGAUGUCCACCCC
(((.((((((((((.(((((.(((((((...........)))))))((((........))))..)))))..))))))))))).)))...No.
short read
(((.((((((((((.(((((.((((                                                                1
 ((.((((((((((.(((((.((                                                                  3
 ((.((((((((((.(((((.(((                                                                 1
                                    ...)))))))((((........)                              1
seq-7   No. long read =3
CUACCCAGGGUUGUGGGCAGUGUGAGUGUCAGCUCCUGUGGCCACUGCUUCCUACCCAGGGUUGUG
..((((.((((...((((((((.((((....)))).......))))))))...)))).))))....No. short read
..((((.((((...((((((                                              1
..((((.((((...(((((((                                             1
..((((.((((...((((((((                                            3
..((((.((((...((((((((.                                           2
 .((((.((((...(((((((                                             1
                                            ))))))...)))).))))... 3
                                            ))))))...)))).))))....1

seq-8   No. long read =1
AAGGAAGGAGAGUCAGCAAGCACCUGGCUGGCCCAGGCUUCAGCUGUCCUCCUUUCUGUAG
.(((((((((...((((.....(((((.....))))).....))))..)))))))))....     No. short read
.(((((((((...((((.....                                            6
.(((((((((...((((.....(                                           11
.(((((((((...((((.....((                                          9
                                     .....))))..)))))))))         1
                                     .....))))..)))))))))..       1
                                     .....))))..)))))))))....     1

seq-9   No. long read =1
UUGGCCCUGAAUCAAGGCCGCAGUUUACUGAAGCUGUUGGUUUCAAGCAGGAGCCUAAAG
..(((((((.....((((((((((((....)))))).))))))....)))).))).....      No. short read
..(((((((.....(((                                                 1
..(((((((.....((((                                                3
                      ((((....)))))).)))                          1
                                       ))))....)))).)))....       1
seq-10  No. long read =1
CUCAUCAGAUGGCUUCUUGGGUUUCCUAGUAAGAUUCUCCCAAGGAGACAUCUUUGAGG
((((..(((((.(((((((((..((.......))....))))))))).))))).)))).       No. short read
 (((..(((((.(((((((                                               4
 (((..(((((.((((((((                                              16
   (..(((((.(((((((((                                             3

seq-11  No. long read =1
UGUGGGAGGGGACUGUAGAGAGGAGGGUGCCUAACCUCUGUUCUGCUCACCCUUCUCACAGU
(((((((((((...((((((.(((((........))))).))))))...)))))))))))..    No. short read
(((((((((((...(((                                                 1
(((((((((((...((((                                                1
(((((((((((...(((((                                               3
(((((((((((...((((((                                              5
(((((((((((...((((((.                                             28
(((((((((((...((((((.(                                            3
(((((((((((...((((((.((                                           11
(((((((((((...((((((.(((                                          1
(((((((((((...((((((.(((((                                        1
  (((((((((...((((((.((                                           1
                                    ))).))))))...))))))           1
                                    ))).))))))...)))))))          6
                                    ))).))))))...))))))))         4
                                    ))).))))))...)))))))))        6
                                    ))).))))))...))))))))))       9
                                    ))).))))))...)))))))))))      3
                                    ))).))))))...))))))))))).     4
                                    ))).))))))...)))))))))))..    5
                                      ).))))))...))))))))))       1
                                      ).))))))...))))))))))).     1
                                      ).))))))...)))))))))))..    2

seq-12  No. long read =7
CAGAGGCCUCAGCUCACCGCCCGCUGCCCCGGUGUGGGAGGGUGAAACCCAGGCCCCUACAUUC
.((.(((((....(((((.(((((.((....)))))))..))))).....))))).))......  No. short read
.((.(((((....(((((.(                                              6
.((.(((((....(((((.((                                             4
.((.(((((....(((((.(((                                            3
             (((((.(((((.((....)))))                              1
                       (.((....)))))))..)                         1
                                        )))))....))))).))......   1
                                         )))).....))))).)).....   1
                                          ))).....))))).))......  3
                                              ....))))).))......  1

seq-13  No. long read =3
AUGGAGAGACUUUGACAGCUCAGGUCAGCACAGUGCCUGCAGCUGCCACUCAUCUUUCCUAUAUAU
..(((((((...((.(((((((((.((......)))))).))))).))....)))))))).......No. short read
..(((((((...((.((                                                 4
..(((((((...((.(((                                                12
..(((((((...((.((((                                               157
..(((((((...((.(((((                                              1675
..(((((((...((.((((((                                             1099
..(((((((...((.(((((((                                            844
..(((((((...((.((((((((                                           3
..(((((((...((.(((((((((.                                         1
 .(((((((...((.(((((                                              1
 .(((((((...((.(((((((                                            2
 .(((((((...((.(((((((((                                          1
  (((((((...((.(((((                                              1
  (((((((...((.(((((((                                            1
        (...((.(((((((((.((..                                     1
          ..((.(((((((((.((..                                     1
            ((.(((((((((.((......)))))).))                        1
                      ((.((......))))))                           1
                                        ))))).))....)))))))...    1
                                        ))))).))....)))))))....   3
                                          ))).))....)))))))).......4
                                                 ...))))))).......1

seq-14  No. long read =1
UGGUUUUGGGGGGCAUGACUUGGGGUCCGGGGCUCGCGGAGCCAACCAUGUCUUCUUUCCCAG
(((....((((((((((.....((.((((.......)))).))...))))))))))...))).   No. short read
                                         ))...))))))))))...))).   1
                                          )...))))))))))...))).   6

seq-15  No. long read =1
AGGGCUGGAGAGAUGGCUCAGUGGUUAAAAGCACUGACUGCUCUUCCAAAGGUCCUGA
((((((((((((.(((.((((((........)))))))))))))))....)))))).. No. short read
((((((((((((.(((.                                          14
((((((((((((.(((.(                                         20
((((((((((((.(((.((                                        7
((((((((((((.(((.(((                                       8
((((((((((((.(((.((((                                      10
((((((((((((.(((.(((((                                     4
((((((((((((.(((.((((((.                                   4
((((((((((((.(((.((((((..                                  4
((((((((((((.(((.((((((...                                 2
((((((((((((.(((.((((((....                                1
((((((((((((.(((.((((((.....                               2
 (((((((((((.(((.(                                         308
 (((((((((((.(((.((                                        364
 (((((((((((.(((.(((                                       133
 (((((((((((.(((.((((                                      40
 (((((((((((.(((.(((((                                     4
 (((((((((((.(((.((((((..                                  2
 (((((((((((.(((.((((((...                                 2
  ((((((((((.(((.((                                        10
  ((((((((((.(((.(((                                       2
   (((((((((.(((.(((                                       6
    ((((((((.(((.((((                                      2
    ((((((((.(((.((((((.                                   1
     (((((((.(((.((((((                                    4
       (((((.(((.((((((.                                   1
                     ((........))))))))))))))              1
                               )))))))))))))))....))       1
                                ))))))))))))))...          1
                                 )))))))))))))....))))     1
                                 )))))))))))))....)))))    1
                                 )))))))))))))....)))))).. 1
                                  ))))))))))))....))       1
                                  ))))))))))))....)))      4
                                  ))))))))))))....))))     2
                                  ))))))))))))....)))))).. 1
                                   )))))))))))....)))))).. 1
                                    ))))))))))....)))))    1
                                    ))))))))))....)))))).. 6

seq-16 No. long read =1
UGGGACAGGGUGACAGGGUGAGACCCAUAGAUCAGAGCUGGGCUUCACCCAUUUCUCCCUGCUUCCAGU
((((((((((.((..((((((..((((...........))))..))))))...)).))))).)))))..    No. short read
((((((((((.((..(((((                                                     1
((((((((((.((..((((((                                                    3
((((((((((.((..((((((.                                                   1
                                         )..))))))...)).))))).))))       1

seq-17 No. long read =1
CUCUCUCCAACCUUCAUAUUUGUAUUUGUUAGACAAGAAGUACAAAAGUGAGGCUGGGGAGAGG
(((((((((.(((.(((.(((((((((.((....)).))))))))).)))))).)))))))))).  No. short read
                                         ))))).)))))).))))))      1
                                         ))))).)))))).)))))))     1
                                         ))))).)))))).))))))))    1
                                          )))).)))))).)))))       1
                                          )))).)))))).)))))))     1
                                          )))).)))))).))))))))    1
                                          )))).)))))).)))))))))   1

seq-18 No. long read =1
CUGCAUCCACUGAUAGACCUUGAACAGUUUGUGGUUGUUCUUCUGGUUUGCACUAGGAUGCAAAAGG
.(((((((..((.((((((..((((((.......))))))....))))))))...))))))).....      No. short read
.(((((((..((.((((                                                        2
.(((((((..((.((((((                                                      4
 (((((((..((.((((((                                                      1

seq-19 No. long read =1
GAUCGGAGCAGCUCAGAGCAGAUGGCGGCUUCAGCUGCUGCUCUGGCUCCUAAA
....(((((....((((((((.((((.......)))))))))))))))))....    No. short read
 ...(((((....((((((((.                                    1
 ...(((((....((((((((.(                                   2
 ...(((((....((((((((.((                                  1
 ...(((((....((((((((.(((                                 1
 ...(((((....((((((((.((((                                2
  ..(((((....((((((((                                     7
  ..(((((....((((((((.                                    6
  ..(((((....((((((((.(                                   25
  ..(((((....((((((((.((                                  24
  ..(((((....((((((((.(((                                 72
  ..(((((....((((((((.((((                                 27
   .(((((....((((((((.(((                                  7
   .(((((....((((((((.((((                                 3
    (((((....((((((((.(                                    1

seq-20 No. long read =9
UUAGGGGAUGUGGAGCCGGGAUUGGAGAGAUUGUCACAACCUGAUCCCGUUUCCUUCCUCCUGUCCCCUAG
.(((((((((.((((.(((((((((...............)))))))))........)))))))))))))).   No. short read
.(((((((((.((((.(((((                                                     1
.(((((((((.((((.((((((                                                    2
 (((((((((.((((.(((((                                                     1
 (((((((((.((((.((((((                                                    2
 (((((((((.((((.(((((((                                                   2
 (((((((((.((((.(((((((((                                                 1
                                          ))))))).......)))))))          1
                                          ))))))).......))))))))))))     1

seq-21 No. long read =5
ACAGCGCCAGCUGCCUAAUUGAUUGUCUCUGUUAAUUAGGCUCUGGAUCUGUGA
((((..((((..(((((((((((.......))))))))))).))))..))))..     No. short read
                                )))))))))).))))..)))).     3
                                ))))))))).))))..))))..     79
                                 )))))))).))))..)))).      2
                                 )))))))).))))..))))..     72
                                   )))))).))))..))))..     2

seq-22 No. long read =1
UUGUAUUUGUGUGAUUAAAGUAUUAGAAGUAUUGAAAUUCUGAGUUUUGCUUUUUUCUUCACAAAUACAG
.((((((((((.((..((((((((((((.........))))))....))))))..))..)))))))))).    No. short read
.((((((((((.((..((((                                                      3
.((((((((((.((..(((((                                                     3
.((((((((((.((..((((((                                                    2
 ((((((((((.((..(((((                                                     2
 ((((((((((.((..((((((                                                    1

seq-23 No. long read =4
UAGCAGAGGUACCCAUUCCAUUCCCAGUUUGCUCGGUAGCUGGUGAUUGGAAGACACUCUGCAACAUUA
..((((((((.....(((((.((((((((........)))))).)).))))).)).)))))).......     No. short read
..((((((((.....(((((.(                                                    4
..((((((((.....(((((.((                                                   3
..((((((((.....(((((.(((                                                  1
..((((((((.....(((((.((((                                                 2
                               ......)))))).)).)))                        1
                               ......)))))).)).))))).)).                  1
                                .....)))))).)).))))).)).))))              1
                                 ....)))))).)).))))).)).)))               1
                                 ....)))))).)).))))).)).)))))             1
                                  ...)))))).)).))))).)).)))))             1
                                   ..)))))).)).))))).))                   1
                                   ..)))))).)).))))).)).))))))...         1
                                       )))).)).))))).)).)))))).......     1
                                        ))).)).))))).)).)))))             2
                                        ))).)).))))).)).))))))            9
                                        ))).)).))))).)).)))))).           2
                                        ))).)).))))).)).))))))...         1
                                        ))).)).))))).)).)))))).......     1
                                         )).)).))))).)).)))))).           1
                                         )).)).))))).)).)))))).....       1
                                         )).)).))))).)).)))))).......     7
                                          ).)).))))).)).)))))             3
                                          ).)).))))).)).))))))            98
                                          ).)).))))).)).)))))).           338
                                          ).)).))))).)).))))))..          280
                                          ).)).))))).)).))))))...         584
                                          ).)).))))).)).))))))....        204
                                          ).)).))))).)).)))))).....       6
                                          ).)).))))).)).)))))).......     9
                                           .)).))))).)).))))              1
                                           .)).))))).)).)))))             12
                                           .)).))))).)).))))))            63
                                           .)).))))).)).)))))).           488
                                           .)).))))).)).))))))..          993
                                           .)).))))).)).))))))...         2706
                                           .)).))))).)).))))))....        2721
                                           .)).))))).)).)))))).....       81
                                           .)).))))).)).))))))......      2
                                           .)).))))).)).).)))))).......     8
                                            )).))))).)).)))))).           2
                                            )).))))).)).))))))..          6
                                            )).))))).)).)))))))...        7
                                            )).))))).)).))))))....        11
                                            )).))))).)).))))))......      3
                                            )).))))).)).)))))).......     47
                                             ).))))).)).))))))...         2
                                             ).))))).)).))))))....        8
                                             ).))))).)).)))))).....       1
                                             ).))))).)).))))))......      3
                                             ).))))).)).)))))).......     38
                                              .))))).)).))))))....        3
                                              .))))).)).)))))).....       9
                                              .))))).)).)))))).......     8
                                               ))))).)).)))))).......     6
                                                )))).)).)))))).....       1
                                                )))).)).)))))).......     12
                                                 ))).)).))))))......      1
                                                 ))).)).)))))).......     16
                                                  )).)).)))))).......     8
                                                   ).)).))))))......      4
                                                   ).)).)))))).......     126
                                                    .)).)))))).......     107
seq-24  No. long read =1
UCUGUGUGUCAGCAGCAUGUUCCUGCAUUGGGCCCAUAGCAGUGACUGCCUGCUCUCCCACAGUA
.(((((....(((((((.((..((((..((....))..))))..)))).)))))....)))))..  No. short read
  ((((....(((((((.((..(((                                          1
                                         )..)))).)))))....))))).   3
                                          ..)))).)))))....)))))    9
                                          ..)))).)))))....))))).   33
                                          ..)))).)))))....))))).. 1

seq-25  No. long read =1
UUGCAAGCAACACUCUGUGGCAGAUGGACAAAACCGUCUGACACAAUUUGAGCUUGCUAUAGCAAGG
..((((((..((...((((.(((((((......))))))).))))...)).)))))).........       No. short read
..((((((..((...((                                                         60
..((((((..((...(((                                                        105
..((((((..((...((((                                                       32
..((((((..((...((((.                                                      33
..((((((..((...((((.(                                                     125
..((((((..((...((((.((                                                    63
..((((((..((...((((.(((                                                   20
..((((((..((...((((.((((                                                  39
..((((((..((...((((.(((((                                                 9
..((((((..((...((((.((((((                                                15
 .((((((..((...(((                                                        4
 .((((((..((...((((                                                       13
 .((((((..((...((((.(                                                     5
 .((((((..((...((((.((                                                    3
 .((((((..((...((((.(((                                                   1
 .((((((..((...((((.((((                                                  4
    ((((..((...((((.(                                                     2
    ((((..((...((((.((((((                                                1
       (..((...((((.((((                                                  1
                           ......))))))).))))...)).)))))                  1
                                   ))))).))))...)).))                     1
                                   ))))).))))...)).)))))).....            1
                                    )))).))))...)).))))))........         1
                                      )).))))...)).))))))...              1
                                       ).))))...)).))))))....             1

seq-26  No. long read =2
UAGGUGAGCUCUUGGUACCUUGGCGACAAGAAGGUGAGCCCAGGAACAAGGUCUCAGUCCGAGG
..((((((..((((...(((.(((.((......))..))).)))..))))..))))..))....  No. short read
 .((((((..((((...(((.(((                                          1
  ((((((..((((...(((.(                                            2
  ((((((..((((...(((.((                                           28
  ((((((..((((...(((.(((                                          103
   (((((..((((...(((.((                                           1
   (((((..((((...(((.(((                                          3
                                       ).)))..))))..))))..))..    2
seq-27  No. long read =1
UGCAGGGAGAGCGCAGGUCGUUGACAUAUAAUUAGUGCACUCACCUGGCCUUCUCCUUGCCCCAUU
.(((((((((((.(((((...((.(((.......)))))...)))))))..)))))))))......No. short read
.(((((((((((.(((((..                                              3
.(((((((((((.(((((...                                             8
.(((((((((((.(((((...(                                            66
.(((((((((((.(((((...((.                                          2

seq-28  No. long read =1
GAGAGGACGACCGCCGAGGAGCGCGCGGGUCCGGGAACGUGUCCUCGGUAGUCUUAGUCCCGCU
....(((((((..(((((((.((((((....))....)))))))))))..)))...))))....  No. short read
                                       )))))))))..)))...))        1
                                       )))))))))..)))...)))       2
                                       )))))))))..)))...))))..    1
                                       )))))))))..)))...))))...   5
                                       )))))))))..)))...))))....  1

seq-29  No. long read =3
UCUGUGACUCCUGAGCUCUGUUCCCCAUGCUAGGUUCAGGGAUAAAUGGAGUCACAGAC
(((((((((((....(((((..((........))..)))))......))))))))))).       No. short read
(((((((((((....(((                                                1
(((((((((((....((((                                               1
(((((((((((....(((((                                              2
(((((((((((....(((((.                                             24
(((((((((((....(((((..                                            23
(((((((((((....(((((..(                                           2
(((((((((((....(((((..((                                          3
(((((((((((....(((((..((.                                         2
(((((((((((....(((((..((..                                        2
  (((((((((....(((((..                                            1
  (((((((((....(((((..(                                           1
  (((((((((....(((((..((                                          2
  (((((((((....(((((..((.                                         1
  (((((((((....(((((..((.......                                   1
           ....(((((..((.......                                   1
                                  ..)))))......))))))))))         2
                                  ..)))))......)))))))))))        16
                                   .)))))......))))))))           1
                                   .)))))......))))))))))         1
                                   .)))))......)))))))))))        3
                                     ))))......)))))))))))        2
                                     ))))......))))))))))).       7
                                         ......)))))))))))        1

seq-30  No. long read =1
UCAUUUCUGUAGCUUCCACGGGGCUGUUGGUCUUUCAAAGAAUUUAGCCUCGGGCUGGUGAGAUGGCU
(((((((..(((((....((((((((....((((...))))...)))))))))))))..)))))))..      No. short read
                                                 ))))))))..)))))))        1
                                                 ))))))))..)))))))..      2
                                                  )))))))..))))))).       2
                                                  )))))))..)))))))..      3
                                                   ))))))..)))))))..      1

seq-31  No. long read =1
CAGGGAUCGCCGGGAGCUAUGGUGGGGGCAGACCCGACCAGACCCCGGGGACCGCUGAGCUAUUGGG
(((((.((.(((((.....((((.(((.....))).))))...))))).)))).)))..........       No. short read
 ((((.((.(((((.....                                                       4
 ((((.((.(((((.....((                                                     1
 ((((.((.(((((.....(((                                                    1
  (((.((.(((((.....(((                                                    1
                       .(((.....))).)))).                                 1
                                ))).))))...))))).))                       1

seq-32  No. long read =8
AGGGGUUAAGGGGCUGGUUUAGGGCCCUGAUACUAAGUGGGACAACCAGACUUAAACCCAGG
.(((.(((((...((((((.....(((...........)))..)))))).))))).)))...    No. short read
                                        )..)))))).))))).)))..     4
                                        )..)))))).))))).)))...    9

seq-33  No. long read =1
AUAGACCUGUAUAGCUAUCUAUAUGUAUACUGAUCUAUAGAUCUAUAUAGGUCUGUAUA
((((((((((((((..(((((((.((......)).)))))))))))))))))))))...       No. short read
((((((((((((((..((((                                              3
((((((((((((((..((((((                                            2
                                     )))))))))))))))))))..        1

seq-34  No. long read =1
UCUGGGGGUACUGGUUGGUUCAUAUUGUUGUUUCUCUUAUGGGCCUGCAAACCCCUUCAG
...((((((....((.((((((((.............)))))))).))..))))))....      No. short read
...((((((....((.((((((                                            3
...((((((....((.(((((((                                           2
...((((((....((.((((((((                                          2

seq-35  No. long read =3
ACUGUUACACAAUUUAAUGCCUCUUUCUUAGCCACACAGGAGGAGGAUUAUGUGUGACAGACACAAGG
.((((((((((...((((.(((((((((.........)))))))))))))))))))))))........     No. short read
.((((((((((...((((.                                                      2
.((((((((((...((((.(                                                     7
.((((((((((...((((.((                                                    3
.((((((((((...((((.(((                                                   2
.((((((((((...((((.((((                                                  2
.((((((((((...((((.(((((                                                 5
.((((((((((...((((.((((((                                                2
.((((((((((...((((.(((((((                                               1
                                    .)))))))))))))))))))))))......       1
                                     )))))))))))))))))))))))).....       1
                                     )))))))))))))))))))))))......       2
                                      ))))))))))))))))))))))..           1
                                      ))))))))))))))))))))))......       9
                                       ))))))))))))))))))))).            15
                                       )))))))))))))))))))))..           1
                                       )))))))))))))))))))))...          1
                                        )))))))))))))))))))              2
                                        ))))))))))))))))))))             3
                                        )))))))))))))))))))).            20
                                        ))))))))))))))))))))..           5
                                        ))))))))))))))))))))...          10
                                        ))))))))))))))))))))......       1
                                         )))))))))))))))))               1
                                         ))))))))))))))))))              10
                                         )))))))))))))))))))             82
                                         ))))))))))))))))))).            1990
                                         )))))))))))))))))))..           3518
                                         )))))))))))))))))))...          4462
                                         )))))))))))))))))))....         25
                                         ))))))))))))))))))).....        3
                                         )))))))))))))))))))......       2
                                          )))))))))))))))))).            10
                                          ))))))))))))))))))..           33
                                          ))))))))))))))))))...          99
                                          ))))))))))))))))))....         12
                                          )))))))))))))))))).....        4
                                          ))))))))))))))))))......       3
                                           ))))))))))))))))).            2
                                           )))))))))))))))))..           5
                                           )))))))))))))))))...          6
                                           )))))))))))))))))....         8
                                           ))))))))))))))))).....        2
                                           )))))))))))))))))......       6
                                            )))))))))))))))).            2
                                            ))))))))))))))))..           12
                                            ))))))))))))))))...          15
                                            ))))))))))))))))....         2
                                              ))))))))))))))...          1
                                              ))))))))))))))......       1
                                               )))))))))))))......       1
                                                  ))))))))))........     1

seq-36  No. long read =8
UGGGAAACCUGUGUCGGGCUGUGAGUGUUUUCUGGGACCUGAUACUCACUGCACCGACCUGCUCUCCCAGU
(((((.......(((((((.(((((((((...........))))))))).)).)))))......)))))..  No. short read
(((((.......(((((((.                                                     1
(((((.......(((((((.(                                                    1
(((((.......(((((((.((                                                   8
(((((.......(((((((.(((                                                  8
(((((.......(((((((.((((                                                 2
           .(((((((.(((((((((......                                      1
                                                ).)).)))))......)))))..  3

seq-37  No. long read =1
UGUGUGGAAGCCUCUAGCCUGCUGUCUCCCAUGCAGGUGGCAGCAGCUGGCGCCUUCGCACAGA
((((((((.((..(((((.(((((((.((......)).)))))))))))))).)).))))))))..  No. short read
((((((((.((..(((((.                                               1
((((((((.((..(((((.(((                                            5
((((((((.((..(((((.((((                                           3
seq-38  No. long read =4
UUUCGGGCUCCAGACAUCUGUCCACCUUCCUGGCCAGAGGAUAGGUGUCUGCCCAGCCCUAGAG
....(((((.(((((((((((((.((.....)).....)))))))))))))...))))).....  No. short read
....(((((.(((((((((((                                             2
....(((((.(((((((((((((                                           2
  ..(((((.(((((((((((((                                           1
                                         ))))))))))...)))))....   1

seq-39  No. long read =1
CCCACCGCUGCCACCAAUGGCCAGGGCAUUGACCUGUCAUGCCUCGCCUGGAGGCAGCAGCAUUGGGGA
((((..((((((.(((..(((..((((((.((....)))))))).)))))).)))))).....))))..     No. short read
((((..((((((.(((..                                                        5
 (((..((((((.(((..                                                        4

seq-40  No. long read =4
GUGAGCAGACAGGGAGUGGUGGGGGAAGUCUUCUCCUGGGCUCCCUGAGCCCACACGUAACCCUCACCCUGCUGCCGGCUUGCAGU
((((((.....((.(((((((((((.((....))..(((((((...))))))).......))))))))..))).)).))))))...  No.
short read
((((((.....((.(((((((((                                                                 1
                                                               )))))..))).)).)))))      1
                                                               )))))..))).)).))))))     3
                                                               )))))..))).)).)))))).    2
                                                               )))))..))).)).))))))..   7
                                                               )))))..))).)).))))))...  1

seq-41  No. long read =1
AACCCUUGGGGCCCCUCACCAAUCAGACAGUGGAAAUGCAGGGGCAACAGGGGAAGUCA
..((((((..(((((((((((.........)))...)).))))))..))))))......       No. short read
            (((((((((.........))).                                1
                                 ...)).))))))..))))))....         1
                                  ..)).))))))..))))))...          4
                                  ..)).))))))..))))))....         1

seq-42  No. long read =3
UCUGCUAGCGCAUAACUGGGGCCGCCUGCCCUUCGCGGGCGGCCCUUUUAACCGCUAGCUACAGGC
...(((((((..(((..((((((((((((.....)))))))))))).)))..))))))).......No. short read
...(((((((..(((..((                                               7
...(((((((..(((..(((                                              13
...(((((((..(((..((((                                             28
...(((((((..(((..(((((                                            80
...(((((((..(((..((((((                                           18
...(((((((..(((..(((((((                                          31
...(((((((..(((..((((((((                                         4
 ..(((((((..(((..((((                                             3
 ..(((((((..(((..(((((((                                          2
  .(((((((..(((..(((                                              1
  .(((((((..(((..((((((((                                         1
                  (((((((((((.....)))))                           1
                                    )))))))))).)))..))))))).......4
                                     ))))))))).)))..)))))))...... 1
                                       ))))))).)))..))))))).......1
                                         ))))).)))..)))))))...... 1
                                         ))))).)))..))))))).......5
                                          )))).)))..)))))))...... 1
                                          )))).)))..))))))).......51
                                           ))).)))..))))))).....  1
                                           ))).)))..))))))).......16
                                            )).)))..))))))).......6
                                             ).)))..))))))).......2
                                              .)))..)))))))...... 1
                                              .)))..))))))).......5
                                               )))..))))))).......5
                                                ))..)))))))...... 2
                                                ))..))))))).......2
                                                 )..))))))).......31

seq-43  No. long read =1
UGGGGAGUCGGGCUGCCGCGGGGCUGUUCAGACUGAUGGCAAAUGCAACGCUGUAAUCCCUCUCCAGU
(((((((..((..(((.(((..(((((.((......)))))...))..))).)))..)))))))))..      No. short read
                                              ..))).)))..))))))))).       4
                                              ..))).)))..)))))))))..      17

seq-44  No. long read =2
CUGGCCUGGGCAGGACAGGGCAAGCUUUGGCCUAAUUGUCCUUCCCUGUCUCCUCCCUCUUGGCCACU
.(((((.(((..((((((((.(((....(((......)))))))))))))))...)))....)))))..      No. short read
 (((((.(((..((((((((                                                      1
 (((((.(((..((((((((.                                                    1
 (((((.(((..((((((((.(                                                   3
                                          ))))))))...)))....)))          1

seq-45  No. long read =5
CAGCAGAGGAAAUCCAGACGGGUCGUUUCCAUCUGCCCUGGGGCCUGUCUCUACAACUCUGCCACA
..((((((.......(((((((((....(((.......))))))))))))......))))))....No. short read
..((((((.......(((((                                              1
..((((((.......((((((                                             1
..((((((.......(((((((                                            3
..((((((.......((((((((                                           13
 .((((((.......((((((((                                           3
      ((.......((((((((                                           2
               (((((((((....(((.......)))))))                     1
                                     .))))))))))))......))))))....1
                                       )))))))))))......))))))    1
                                       )))))))))))......))))))....1
                                          ))))))))......))))))....1
                                            ))))))......)))))).   1
                                            ))))))......))))))....1
                                             )))))......))))))....3
                                              ))))......))))))....1

seq-46  No. long read =1
GUGAGGCUCAGUAUGGGGUGGGGGUGUCGUCGACUGCCCGACUGACCACCCACUCACCCUGGACUGACUCUCAGA
.((((..(((((.(((((((.(((((((((((......)))).)).)))))...))))))).)))))..))))..      No.      short
read
.((((..(((((.((((                                                                1
.((((..(((((.((((((                                                              1
.((((..(((((.(((((((.                                                            2
.((((..(((((.(((((((.(                                                           2
.((((..(((((.(((((((.((                                                          1
                                  ....)))).)).)))))...)                          1
                                                   ...))))))).)))))..)))         1
                                                   ...))))))).)))))..))))..      1
                                                     .))))))).)))))..))))        10
                                                     .))))))).)))))..)))).       7
                                                     .))))))).)))))..))))..      67
                                                      ))))))).)))))..))))..      2

seq-47  No. long read =117
GUGGGAGAAGGUCUGGGAGACCUGCAUCGUGGGCACAGGCUCAUGGGGACCUGCUGACCGCUGCCUGAUCUUACUCCCAGA
.((((((((((((.((....)).(((..((((.((((((((.....)).)))).)).)))))))..)))))).))))))..No.      short
read
.((((((((((((.((..                                                               2
.((((((((((((.((....                                                             4
.((((((((((((.((....)                                                            2
.((((((((((((.((....))                                                           4
.((((((((((((.((....)).                                                          1
                      .(((..((((.((((((                                          1
                                        (.....)).)))).)).))))))).                1
                                                           )))))..)))))).))))))..1

seq-48  No. long read =1
UGCAUAGACUUGACCAUUUCUAUCGUAAGUUAAAUGGUGAAAUGGAAAAGUCUUG
.....((((((..(((((((.(((((.......))))))))))))..))))))..    No. short read
.....((((((..(((((                                         1
.....((((((..((((((                                        1
.....((((((..(((((((                                       2
       ((((..(((((((.(((((....                             1
                               ..))))))))))))..))))))..    2
                                .))))))))))))..))))))      1
                                .))))))))))))..)))))).     1
                                .))))))))))))..))))))..    2
seq-49  No. long read =1
CUGUAGGCCAACGGGGGAAGGAAGGUAACCUGUUGCCCUUUCUCCUGCUGCCCUGCA
..(((((.((.((((((((((..(((((....))))))))))))))).)).))))).  No. short read
 .(((((.((.((((((((((.                                     2
 .(((((.((.((((((((((..                                    2
 .(((((.((.((((((((((..((                                  2
 .(((((.((.((((((((((..(((                                 1
                                   ))))))))))))).)).)))))   1

seq-50  No. long read =89
UCCGAGGCUAGAGUCACGCUCAGGUAUUGCUUGUUGCCUUAGUGUGCUUAAGUCCUCGAAGA
..(((((((.(((.((((((.(((((........))))).))))))))).)).)))))....    No. short read
```

```
..(((((((.(((.((((                                                 3
..(((((((.(((.(((((                                                11
..(((((((.(((.((((((                                               21
..(((((((.(((.((((((.                                              28
..(((((((.(((.((((((.(                                             68
..(((((((.(((.((((((.((                                            3
..(((((((.(((.((((((.(((                                           13
..(((((((.(((.((((((.(((((                                         1
..(((((((.(((.((((((.(((((..                                       1
 .(((((((.(((.((((((.(                                             1
 .(((((((.(((.((((((.(((                                           1
  (((((((.(((.((((((.                                              1
  (((((((.(((.((((((.(                                             3
    (((((.(((.((((((.(((((..                                       1
      (((.(((.((((((.(((                                           1
       ((.(((.((((((.(((((                                         1
                                      ).))))))))).)).)))))...      2
                                       .))))))))).)).)))))..       1
                                       .))))))))).)).)))))...      23
                                       .))))))))).)).)))))....     15
                                        ))))))))).)).)))))...      2
                                        ))))))))).)).)))))....     2

seq-51  No. long read =1
UUAGCUGGGCAUGAUCUGAUGAGCUCACAGCCAGGCUCUGGCUUAUGCCUAGCUUUC
..((((((((((((.((...(((((........))))).)).))))))))))))...  No. short read
..((((((((((((.((.                                         1
..((((((((((((.((..                                        3
..((((((((((((.((...                                       6
..((((((((((((.((...((                                     45
..((((((((((((.((...(((                                    7
..((((((((((((.((...((((                                   134
..((((((((((((.((...(((((                                  5
 .((((((((((((.((..                                        3
 .((((((((((((.((...((                                     14
 .((((((((((((.((...(((                                    3
 .((((((((((((.((...((((                                   184
 .((((((((((((.((...(((((                                  19
  ((((((((((((.((...((((                                   1
          ((((.((...(((((.....                             1
                                    )).)).))))))))))))...  1
seq-52  No. long read =1
UCCUGUGUUAGAGCUCAGGGUUGAGAUCAUGUGAUCUAUCCGGGUUUCUAACACACUAGA
...(((((((((((((.((((..(((((....)))))))))))).))))))))))).....       No. short read
...(((((((((((((.((((                                              4
...(((((((((((((.((((..(                                           1
  .(((((((((((((.((((..(                                           1
                           (....)))))))))))).                      1
                                    )))))))).))))))))))).          4
                                    )))))))).))))))))))..          32
                                    )))))))).))))))))))...         4
                                    )))))))).))))))))))....        19
                                     ))))))).))))))))))..          1
                                     ))))))).))))))))))).....      1
                                      )))))).))))))))))....        8
                                      )))))).))))))))))).....      1
                                        )))).))))))))))..          1

seq-53  No. long read =1
CACAGGCACCACAGGUUUGAGCAUUUUGAUUGAAUUGCCAAACCAGGCGUGCCUGUGGA
(((((((((....((((((.(((.(((....))).)))))))))....)))))))))..        No. short read
 ((((((((....((((((.((                                             2
 ((((((((....((((((.(((                                            3
                                     )))))))....))))))))).         8
                                     )))))))....)))))))))..        1
                                      ))))))....)))))))))..        1

seq-54  No. long read =1
ACAGGACUGUUGGGACUCCUGGACAGGACAACCCAGGAGUCUCCCUGCACCCUCUGU
(((((..(((.(((((((((((..........)))))))))))...)))...)))))  No. short read
(((((..(((.((((((                                          1
(((((..(((.((((((((((                                      2
(((((..(((.(((((((((((                                     4
(((((..(((.(((((((((((.                                    4
(((((..(((.(((((((((((........                             1
 ((((..(((.(((((((((((..                                   1
seq-55 No. long read =2
AACCCUUGAUUACAUCCUUGCCCUGAUACUGUACCAGUGGCAGCUGUUACUCAAGGGAACAGUG
..(((((((..(((...((((((((.........))).))))).)))...))))))).......  No. short read
..(((((((..(((...(((                                              1
..(((((((..(((...((((                                             4
..(((((((..(((...(((((                                            10
..(((((((..(((...((((((                                           1
..(((((((..(((...(((((((                                          1
                      (((.........))).))                          1
                                     .))))).)))...)))))))...      2
                                      ))))).)))...).......  1
                                       )))).)))...)))))))...      8
                                       )))).)))...)))))))....     5
                                        ))).)))...))))))).        13
                                        ))).)))...)))))))..       11
                                        ))).)))...)))))))...      1013
                                        ))).)))...)))))))....     267
                                        ))).)))...))))))).....    5
                                        ))).)))...)))))))......   2
                                        ))).)))...))))))).......  2
                                         )).)))...)))))))...      2
                                         )).)))...)))))))....     5
                                         )).)))...))))))).......  12
                                          ).)))...))))))).......  7
                                            )))...))))))).......  2
                                             ))...))))))).......  3
                                              )...))))))).......  1
                                               ...))))))).......  5
```

```
seq-56 No. long read =2
CCAUGGAGUAACAGGUGCUUGGUGGGGUUCUGUGAGUAGGAAGUCCCACUUGGGCCUGUCUCCACAGA
...(((((..((((((.((.((((((.((((.......)))).))))))).)))))))))))))....     No. short read
 ..(((((..((((((.((.((                                                   2
 ..(((((..((((((.((.(((                                                  3
 ..(((((..((((((.((.(((((                                                1
  .(((((..((((((.((.(((                                                  6
  .(((((..((((((.((.(((((                                                3
   (((((..((((((.((.(((                                                  1
   (((((..((((((.((.(((((                                                2
   (((((..((((((.((.((((((                                               1
                                               ))).)))))))))))))....     1
```

```
seq-57 No. long read =1
GUGAGGACCCAGGUGUGAUGGGAGGGCGGAGUCCAGAUUAGGACCUCAGCUCAAAAAUCCACCUCAUCUGGUUCUCUGUCCUCAGUUU
.((((((((((((((...((((.((((.((((((......))).))).)))).....))))...)))))))......)))))))....No.
short read
.((((((((((((((..                                                                       2
.((((((((((((((...                                                                      1
.((((((((((((((...(                                                                     1
.((((((((((((((...((                                                                    1
.((((((((((((((...(((                                                                   7
.((((((((((((((...((((                                                                  2
                                                               .)))))))......)))))))..  1
                                                                )))))))......)))))))    1
                                                                )))))))......))))))).   34
                                                                )))))))......)))))))..  15
                                                                )))))))......)))))))... 6
                                                                )))))))......)))))))....1
```

```
seq-58 No. long read =1
GUGAGCCAUGGCGUGUGCAGAGGCAGGGGCUGGGGUGGAGGAAGGCCCAUACUUCUGACUGCCAUUCCUUACAGAUA
(((((..((((((.((.((((((...(((((............)))))...))))))))))))))..)))))).....    No.      short
read
(((((..((((((.((.((((((                                                          1
                                                     ))))))))))))..)))))...      5
```

```
seq-59 No. long read =1
CUUGCAUGUGGGCCUGUGUGCUAUAAUGUCACUUGCUAACAGACGGGCUCUCAUGCUGACAG
...(((((.((((((((.((.(((((......))).)).)).)))))))).)))))......    No. short read
...(((((.((((((((.((.                                             3
...(((((.((((((((.((.(                                            3
 ..(((((.((((((((.((.                                             1
  .(((((.((((((((.((.(                                            1
  .(((((.((((((((.((.((                                           1
                                     ).)).)))))))).)))))...       1
```

```
                                      .)).)))))))).)))))...       2
                                      .)).)))))))).)))))....      2
                                      .)).)))))))).)))))).....     2
                                       )).)))))))).)))))...       2
                                       )).)))))))).)))))....      1
                                       )).)))))))).))))).....     3
                                       )).)))))))).)))))......    1
                                        ).)))))))).)))))).....     2
                                        ).)))))))).))))))......   2

seq-60  No. long read =2
UUGGGGACAGAGGCACAGGAUGCUGGCUGCUGACCACCUGUCCACUGUCUCUGGUCCCAGU
(((((..((((((((..((((..(((.......)))...))))..))))))))..)))))).     No. short read
(((((..((((((((..((((.                                            1
 ((((..((((((((..((((..                                           1
                                      .))))..))))))))..))         3
                                      .))))..))))))))..))))       1
                                      .))))..))))))))..)))))      6
                                      .))))..))))))))..))))).     20

seq-61  No. long read =1
UUGAUCAGCUGCAGGGUUUGCGGAACUCUGUCCGAAUGGAACUGCAGGUCCUGCCGCGGUCGCCA
..((((.((.(((((..((((((...(((((....))))).))))))..))))).))))))).... No. short read
            (((..((((((...(((((....)                              1
                                              )..))))).))))))..   1
                                              )..))))).)))))).... 16
                                               ..))))).))))))...  3
                                               ..))))).)))))).... 90
                                                .))))).)))))).... 58

seq-62  No. long read =2
UGGAUGUGUGAUGAGAUGAGGCUGUGCUUUCAAGGGCUCACCACUUCACACAUUAAAG
..(((((((((.....((((.((.((....)).)).)))).....))))))))).... No. short read
..(((((((((.....(((                                        1
..(((((((((.....((((                                       3
..(((((((((.....((((.                                      88
..(((((((((.....((((.(                                     55
..(((((((((.....((((.((                                    1
 .(((((((((.....((((                                       1
 .(((((((((.....((((.(                                     2

seq-63  No. long read =2
CCCUCCCCCUCUCCUGGCUGCUCUGGGCAGCAGCAACCUCGGCAGCCCGGGAGAUGGGAAGAGA
..((((((.(((((((((((((..((((....))..))..))))).)))))))).)))..))).  No. short read
..((((((.(((((((((((((                                            1
 .((((((.(((((((((((((..((                                        1
              ((((((((..((((....))                                1
                      ..((((....))..))..                          1
                                      ..))))).)))))))).))         1
                                      ..))))).)))))))).)))        1
                                      ..))))).)))))))).))).       7
                                      ..))))).)))))))).)))..      2
                                      ..))))).)))))))).)))..)     1
                                          ))).)))))))).)))..))).  2

seq-64  No. long read =1
GGGCAGGCGCGGAGGCGGGUCCGCAGAUCCCGAGCGGAGCUUGCCUGGCGUCUCUGUUU
.((.((((((..((((((((((((.........))))).))))))).))))))))....       No. short read
                                  )))).))))))).))))))))           2
                                  )))).))))))).)))))))).          12
                                  )))).))))))).))))))))..         19
                                   ))).))))))).)))))))).          1

seq-65  No. long read =1
CUCGCGGCCUCUAGCGUGACGUCUCCACGCCUCGGCGGAGAGGCCGCGUCGGGCCGCAGC
...(((((((...(((((..(((((..(((....)))..)))))))))).)))))))...      No. short read
                                  )))..)))))))))).))))            1
                                      .)))))))))).)))))))...      9

seq-66  No. long read =1
CUGCCUGUAGUCCCAGCUACUUGAGAGGCUGAGGUGGGAAGAUUGCUUCAGCUUUGGAGUUUGAAGCUGCAGUGAGC
((..(((((((..(((..((((..(((((((((((((.....))))))))))))).)))))))..)))))))..)).    No.      short
read
 (..(((((((..(((..                                                               2
 (..(((((((..(((..(                                                              1
 (..(((((((..(((..((                                                             2
 (..(((((((..(((..(((                                                           1
  ..(((((((..(((..(                                                             5
  ..(((((((..(((..((                                                            9
  ..(((((((..(((..(((                                                           3
   .(((((((..(((..((                                                            8
   .(((((((..(((..(((                                                           3
    (((((((..(((..(((                                                           1
             ((..((((..((((((((((                                               2
                .((((..((((((((((                                               1
```

```
seq-67  No. long read =1
UCUCCAGGAGUCUGAGGGGCAGGGUCGUCUCCAGCAUCUUGGGCUUGGCCUAACCUGCCUCCAUUCUCUUGGGCAG
..((((((((..((.((((((((((((..(((((....)))))..))))....))))))))))..))))))))...     No.      short
read
..((((((((..((.((((                                                              1
..((((((((..((.(((((((                                                           6
..((((((((..((.((((((((                                                          2
  ((((((((..((.((((                                                              2
                                                       ))))))))..))))))))...     1
```

```
seq-68  No. long read =1
UCUGAAGGCAGUCAGGUAGGCGGCACCUCUGACCUGCAGCCCCUGUCUGCCUCAUU
..(((.(((((.((((..(((.(((.........))).))))))).))))))))..   No. short read
..(((.(((((.((((..(                                        1
..(((.(((((.((((..((                                       8
..(((.(((((.((((..(((                                      8
..(((.(((((.((((..(((.                                     21
..(((.(((((.((((..(((.(                                    56
..(((.(((((.((((..(((.((                                   7
..(((.(((((.((((..(((.(((                                  7
..(((.(((((.((((..(((.(((.                                 2
 .(((.(((((.((((..(((.(                                    4
 .(((.(((((.((((..(((.(((                                  3
  (((.(((((.((((..(((.(                                    1
   ((.(((((.((((..(((.(((                                  1
```

```
seq-69  No. long read =1
GUGGGAAGGGGGAGGCCUUUCCAGGAUUUCAAACCUUGGCUCUGCCCCCUCCUUCCCAGU
.(((((((((((.(((....(((((.........)))))....))).)))))))))))..      No. short read
                                      )....))).))))))))))         2
                                      )....))).))))))))))).       11
                                      )....))).)))))))))))..      70
                                       ....))).)))))))))))..      2
```

PUBLICATIONS

1. **Na Li**, Xintian You, Sebastian Mackowiak, Marc Friedlaender, Andreas Gogol Doering, Yuhui Hu, Wei Chen. Global profiling of miRNA and the hairpin precursor: insights into miRNA processing and novel miRNA discovery. Submitted.
2. Friedländer, Marc; Mackowiak, Sebastian; **Li, Na**; Chen, Wei; Rajewsky, Nikolaus. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* NAR-00955-N-2011.R1.
3. SchwanhäusseB, Busse D, **Li N**, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature.* 2011 May 19;473(7347):337-42.
4. Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, He L, Hu Y, Hu H, **Li N**, Chen W, Khaitovich P. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* 2011 May 1;39(15):6596-6607.
5. Anna-Barbara Stittrich, Claudia Haftmann, Evridiki Sgouroudis, Anja Andrea Kühl, Ahmed Nabil Hegazy, Isabel Panse, Rene Riedel, Michael Flossdorf, Jun Dong, Franziska Fuhrmann, Gitta Anne Heinz, Zhuo Fang, **Na Li**, Ute Bissels, Farahnaz Hatam, Angelina Jahn, Ben Hammoud, Mareen Matz, Felix-Michael Schulze, Ria Baumgrass, Andreas Bosio, Hans-Joachim Mollenkopf, Joachim Grün, Andreas Thiel, Wei Chen, Thomas Höfer, Christoph Loddenkemper, Max Löhning, Hyun-Dong Chang, Nikolaus Rajewsky, Andreas Radbruch & Mir-Farzin Mashreghi. The microRNA miR-182 is induced by IL-2 and promotes clonal expansion of activated helper T lymphocytes. *Nature Immunology* 2010 Nov;11(11):1057-62.
6. Xu AG, He L, Li Z, Xu Y, Li M, Fu X, Yan Z, Yuan Y, Menzel C, **Li N**, Somel M, Hu H, Chen W, Pääbo S, Khaitovich P. Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-Seq. *PLoS Comput Biol.* 2010 Jul 1;6:e1000843.
7. Ning-Yi Shao, Hai Yang Hu, Zheng Yan, Ying Xu, Hao Hu, Corinna Menzel, **Na Li**, Wei Chen and Philipp Khaitovich. Comprehensive survey of human brain microRNA by deep sequencing. *BMC Genomics.* 2010, 11:409.
8. Weimin Qiu, Yuhui Hu, Tom E. Andersen, Abbas Jafari, **Na Li**, Wei Chen,

Moustapha Kassem. Tumor necrosis factor receptor superfamily member 19 (TNFRSF19) regulates differentiation fate of human mesenchymal (stromal) stem cells through canonical WNT signaling and C/EBP. *J. Biol. Chem.* 2010, 285(19):14438-49.

9. Huaqin Sun, Dan Li, Shu Chen, Yanyan Liu, Xiaolin Liao, Wenqian Deng, **Na Li**, Mei Zeng, Dachang Tao, Yongxin Ma. Zili Inhibits Transforming Growth Factor-β signaling by Interacting with Smad4. *J. Biol. Chem.* 2010, 285(6):4243-50.
10. Stoeckius M, Maaskola J, Colombo T, Rahn HP, Friedländer MR, **Li N**, Chen W, Piano F, Rajewsky N. Large-scale sorting of C. elegans embryos reveals the dynamics of small RNA expression. *Nat Methods*. 2009, 6(10):745-51.
11. Chen W, Kalscheu V, Tzschach A, Menzel C, Ullmann R, Schulz M, Erdogan F, **Li N**, Kijas Z, Arkesteijn G, Pajares IL, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser BG, Vingron M, Ropers HH. Mapping translocation breakpoints by next-generation sequencing. *Genome Res*. 2008, 18(7):1143-9.

For reasons of data protection,

the curriculum vitae is not included in the online version

For reasons of data protection,

the curriculum vitae is not included in the online version

For reasons of data protection,
the curriculum vitae is not included in the online version

SELBSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Ich versichere, dass diese Arbeit in dieser oder anderer Form noch keiner anderen Prüfungsbehörde vorgelegt wurde.

Hiermit erkläre ich, dass ich die Arbeit selbst verfasst habe sowie keine anderen als die angegebenen Quellen und Hilfsmittel in Anspruch genommen habe. Ich versichere, dass diese Arbeit in dieser oder anderer Form keiner anderen Prüfungsbehörde vorgelegt wurde.

Berlin, den

____

(Na Li)