Muller Fabbri *Editor*

# Non-coding RNAs and Cancer

Springer

Non-coding RNAs and Cancer

Muller Fabbri
Editor

# Non-coding RNAs and Cancer

Springer

*Editor*
Muller Fabbri
Pediatrics and Molecular Microbiology
Children's Hospital Los Angeles
University of Southern California
Los Angeles, CA, USA

ISBN 978-1-4614-8443-1 ISBN 978-1-4614-8444-8 (eBook)
DOI 10.1007/978-1-4614-8444-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013951653



Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Contents

# Chapter 1
# Introduction

**Muller Fabbri**

**Abstract** The discovery of microRNAs (miRNAs) as regulators of gene expression and of their involvement in human carcinogenesis represents one of the most important scientific discoveries of the last decade. This statement is supported by the plethora of works published in this field which reflect the broad interest of the scientific community for this topic. Recently, it has been shown that miRNAs not only modulate the expression of targeted genes but they are also involved in modulating intercellular signaling within the tumor microenvironment. Finally, new groups of noncoding RNAs are emerging as key players in carcinogenesis and a better understanding of their function and biology identifies new anticancer molecular targets.

**Keywords** MicroRNAs • Noncoding RNAs • Oncogenes • Tumor-suppressor genes • Ultraconserved regions • Inflammation • Tumor microenvironment • Toll-like receptors

In 1993 Victor Ambros et al. described for the first time a small noncoding RNA (ncRNA), called lin-4, able to regulate the expression of a protein-coding gene (PCG), named lin-14 in *Caenorhabditis elegans*, and affecting the development of the nematode [1]. This was the first description of a role of ncRNAs in affecting a phenotype. After a long period essentially devoid of publications in this field, in 2001 Thomas Tuschl et al. published a seminal manuscript identifying several of these small ncRNAs (now named microRNAs or miRNAs) not only in invertebrates but also in vertebrates [2]. The first connection between miRNA dys-regulation and

M. Fabbri, MD, Ph.D. (✉)
Departments of Pediatrics and Molecular Microbiology and Immunology, Norris Comprehensive Cancer Center and Keck School of Medicine of the University of Southern California, the Saban Research Institute and Children's Center for Cancer and Blood Diseases, Children's Hospital Los Angeles, 4650 Sunset Blvd, Mailstop 57, Los Angeles, CA 90027, USA
e-mail: mfabbri@chla.usc.edu

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_1,

cancer was provided by the group of Carlo Croce, who showed that a specific cluster of miRNAs (namely, the miR-15a/16-1 cluster) is located at 13q14, the most frequently deleted region in human chronic lymphocytic leukemia, and that miRNAs are located in cancer-associated genomic regions (CAGRs), including areas of genomic amplification/deletion, recombination, LOH, and fragile sites [3]. After these initial discoveries a plethora of publications have clearly demonstrated an involvement of miRNAs in human carcinogenesis and have dissected the molecular targets of miRNAs responsible for a specific cancer phenotype. While the most relevant of these studies will be presented in the next chapters of this book, I think it is important to identify the different "ages" of the scientific research in the field of ncRNAs. In my opinion this is relevant not only from a historical point of view but also to identify (and try to anticipate to my best) the next directions that scientists will take in this rapidly growing research field. Supported by the advent of high-throughput techniques for the profiling of miRNAs (such as microarrays, bead-based detection methods), we have assisted at an initial "profiling era," in which essentially the miRNome (defined as the full spectrum of miRNAs expressed in a given genome) was profiled in all human types of cancers (both solid and hematological malignancies) with respect to the normal tissue counterpart. This approach has allowed the identification of specific "signatures" of dys-regulated miRNAs, harboring important diagnostic implications. Subsequently, in an attempt to better understand the significance of this observed miRNA dys-regulation from a molecular point of view, scientists in the field have begun a new "target era." Essentially, each group focused on one target messenger RNA (mRNA) of interest, as the main responsible of the phenotype observed by modulating the expression of a given miRNA. This approach has soon revealed inadequate, since it has been shown that one single cluster of miRNAs (namely, the miR-15a/16-1 cluster) is able to modulate (directly and indirectly) about 14 % of the whole human genome in leukemic cells [4]. Therefore, the initial concept that one miRNA targets one mRNA has been rapidly replaced by the concepts that each miRNA targets actually several different mRNAs and that one mRNA can be targeted by several different miRNAs. It is now estimated that about 30 % of all human genes can be regulated by miRNAs [5]. The fact that several mRNAs can be target of a given miRNA has also led to abandoning the academic classification of miRNAs as oncogenes (OG) or tumor-suppressor genes (TSGs), but it has revealed that miRNAs have a dual nature and can act both as OGs or TSGs, based on the type of cancer and in some cases also in a species-specific fashion [6]. In an attempt to determine which factors were responsible for the dys-regulation of miRNAs observed in cancer, researchers started investigating which factors regulate miRNA expression. Since miRNAs are genes, it was no surprise to discover that they essentially undergo the same regulatory mechanisms of any other PCG. In particular, miRNAs are regulated by transcription factors and epigenetic phenomena (e.g., promoter hypermethylation, histone, and chromatin modifications) described for any other gene, and these aspects will be examined in detail in the book. More recently, we are witnessing the beginning of a new age in miRNA research: the "pathway era." This concept stems from the observation that miRNAs do not completely silence the expression of a targeted gene, but

they reduce the protein expression by 40–50 % at best. However, scientists observed that the same miRNA (or a limited number of miRNAs) affects a specific phenotype by targeting multiple genes of a specific pathway or belonging to interacting pathways. This observation is not trivial, since it represents the rationale for the use of miRNA-based therapies versus siRNA-based therapies. Indeed, while siRNAs designed specifically against a given mRNA are way more effective than miRNAs in silencing the expression of that gene, they also work specifically to silence only the gene for which they have been designed. Conversely, miRNAs regulating different effectors of a common (or related) pathway exert a broader and bigger impact on the processes regulated by that pathway and are more likely to interfere with the plasticity of cancer cells and their renowned ability to find escape routes to the silencing of a specific gene.

More recently, it has been shown that miRNAs belong to a bigger family of ncRNAs, whose members are also of great significance in human carcinogenesis. One of the first groups of "other" ncRNAs is composed of the so-called transcribed ultraconserved regions (T-UCRs) of the genome, highly conserved transcripts also frequently dys-regulated in cancer, undergoing transcription factor and epigenetic regulation as well, but whose mechanism of action is essentially still unknown [7]. Intriguingly, it has been shown that in addition to targeting mRNAs, miRNAs are also able to directly target T-UCRs [7]. This discovery reveals a network of cross-regulation among different members of the ncRNA family and adds a layer of complexity to gene expression regulation. In the last years we have assisted at the identification of a plethora of other ncRNAs, relevant to cancer and in some cases with well-defined mechanisms of action. For this reason, a specific chapter will be dedicated to these non-miRNA ncRNAs and their implications in human carcinogenesis. The last few years have been particularly exciting in the field of ncRNAs, since new mechanisms of action for miRNAs have been identified and the results of the first miRNA-based clinical trial conducted in patients affected by HCV have been published [8]. An increasing interest is currently geared towards exosomes, microvesicles released by all cells (including cancer cells), their content, and the significance of exosome cargo for intercellular communication. MiRNAs have been identified in cancer-released exosomes [9], and it has been shown that they can be secreted by cancer cells and engulfed by surrounding immune cells expressing Toll-like receptor 8 (TLR8). Interestingly, exosomic miRNAs (specifically miR-21 and miR-29a) are able to bind to and activate TLR8 in the cancer-surrounding immune cells, in a paracrine, ligand-like fashion, and induce the secretion of pro-inflammatory and pro-tumoral cytokines [10]. This discovery showed that miRNAs can actually also function as ligands for receptors and modulate the intercellular signaling within the tumor microenvironment. Also, circulating miRNAs can be successfully used to identify cancer patients from healthy individuals, and they can predict prognosis and response to treatment in certain cases, revealing a role as diagnostic, prognostic, and predictive-of-response-to-treatment cancer biomarkers. Exciting times are ahead of the research in the ncRNA field. Lots of efforts are put in the creation of nanoparticles specifically designed to target and release their content in cancer cells. The creation of nanoparticles delivering specific cocktails of miRNAs (and/or other

ncRNAs) able to interfere with cancer growth and metastatization represents an appealing yet-not-too-far-to-come perspective in cancer therapy. In parallel, several groups are investigating the effects of currently used drugs in modulating the miR-Nome. These studies are providing the rationale for new drug combinations and/or miRNA–drug combinations in the treatment of tumors. Finally, several lines of evidence reveal that ncRNAs are involved in the development of resistance to therapy, which is essentially the main cause of insuccess in the treatment of cancer. The ability to revert these therapy resistance mechanisms will provide new (and hopefully also less toxic) treatments against cancer and will provide cancer patients with a new realistic hope to finally be cured.

**Acknowledgments** Dr Fabbri is supported by the Jean Perkins Foundation, by Nautica Malibu Triathlon Funds by the St. Baldrick's Foundation, by a Southern California CTSI grant, by the Pablove Foundation, by Funds from the Saban Research Institute, by award number P30CA014089 from the National Cancer Institute, by the Hugh and Audy Lou Colvin Foundation, and by the T.J. Martell Foundation.

## References

1. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843–54.
2. Lagos-Quintana M et al. Identification of novel genes coding for small expressed RNAs. Science. 2001;294(5543):853–8.
3. Calin GA et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101(9):2999–3004.
4. Calin GA et al. MiR-15a and miR-16-1 cluster functions in human leukemia. Proc Natl Acad Sci U S A. 2008;105(13):5166–71.
5. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15–20.
6. Fabbri M et al. Regulatory mechanisms of microRNAs involvement in cancer. Expert Opin Biol Ther. 2007;7(7):1009–19.
7. Calin GA et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell. 2007;12(3):215–29.
8. Janssen HL et al. Treatment of HCV infection by targeting MicroRNA. N Engl J Med. 2013;368(18):1685–94.
9. Mitchell PS et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105(30):10513–8.
10. Fabbri M et al. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. Proc Natl Acad Sci U S A. 2012;109(31):E2110–6.

# Chapter 2
# Biogenesis and Physiology of MicroRNAs

**Carlos A. Melo and Sonia A. Melo**

**Abstract** MicroRNAs (miRNAs) are small noncoding RNAs 17–25 nucleotides long that control gene expression by promoting degradation or repressing translation of target mRNAs. Since each miRNA regulates the expression of hundreds of target mRNAs, miRNAs can be seen as master-coordinators, efficiently regulating fundamental cellular processes such as proliferation, apoptosis, and development. MiRNAs are synthesized in the cell through a multistep coordinated process that starts in the nucleus and proceeds to the cytoplasm culminating with the production of the biological active form, the miRNA. This maturation process consists of a series of biochemical steps that convert the primary miRNA transcript into an intermediate precursor miRNA hairpin and culminates with the formation of the mature miRNA. The spatiotemporal control of miRNA abundance is made possible, in part, by the regulation of its biosynthesis pathway, where alterations can lead to global miRNA deregulation. Since miRNAs are involved in a broad range of developmental and physiological processes their deregulation appears to play a fundamental role in the onset, progression, and dissemination of many cancers as well as in many other human diseases.

**Keywords** miRNAs • Biogenesis • Maturation • Drosha • Dicer • Cropping • Dicing • Slicing

C.A. Melo
Division of Gene Regulation, The Netherlands Cancer Institute,
Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands

Doctoral Program in Biomedicine and Experimental Biology, Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra 3030, Portugal

S.A. Melo (✉)
Division of Matrix Biology, Beth Israel Deaconess Medical Center,
Harvard Medical School, Boston, MA, USA
e-mail: smelo@mdanderson.org

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_2,

## 2.1 MicroRNAs

### *2.1.1 Definition and Genomic Organization*

MicroRNAs (miRNAs) are a class of small (17–25 nucleotide; nt), single-stranded noncoding RNAs that control gene expression in animals, plants, and unicellular eukaryotes that have emerged as key posttranscriptional regulators of gene expression [1–3]. The first studied miRNA gene was lin-4, which was identified in *C. elegans* in 1993 by Victor Ambros and colleagues during genetic studies of defects in larval development [4]. Seven years later, Reinhart et al. discovered a second 22-nucleotide small RNA, let-7, a gene that was also involved in *C. elegans* developmental timing [5]. These reports found that the 22-nt lin-4 and 21-nt let-7 are both translational repressors of mRNAs [4, 5]. These small molecules gained the attention of the scientific community for two main reasons: homologs of the let-7 gene were identified in other animals including humans that suggested an important and fundamental biological role for this small RNA; likewise RNA interference (RNAi) was discovered at that time, and it became clear that both pathways were linked and shared common components. Within the following year, more than 100 additional small RNAs were identified in worms, Drosophila, and in humans [6]. Currently, more than 18,226 miRNAs from 50 organisms are registered in the miRNA database (http://www.mirbase.org/). These small RNAs have emerged as key posttranscriptional regulators of gene expression controlling about ~60 % of all protein-coding genes in mammals. Thus, miRNAs participate in the regulation of almost every cellular event such as development, differentiation, proliferation, and apoptosis.

The genomic distribution of miRNA genes is characterized by the presence of families of several identical or closely related mature miRNAs, encoded within the same genomic cluster. Therefore it is expected that a certain degree of redundancy will exist among miRNAs [7, 8]. In humans approximately 50 % of known human miRNAs are found in clusters and they are transcribed as polycistronic primary transcripts [6, 9]. There are usually two or three genes per cluster and the largest cluster, at 13q31, is composed of seven [10, 11]. This genomic organization confers simultaneous expression of similar miRNAs, possibly leading to a synergistic effect in their ultimate function. Moreover, a significant portion of miRNAs are located in the intronic region of protein-coding and/or -noncoding transcription units, whereas a minor subset of miRNAs are mapped to repetitive sequences such as long interspersed nuclear elements [12, 13]. There are four groups of miRNA genes according to their genomic location: intronic miRNA in noncoding transcription units; exonic miRNA in noncoding transcription units; intronic miRNA in protein-coding transcript units or exonic miRNAs in protein-coding transcripts (Fig. 2.1). About one-third of miRNA genes are located in the introns of protein-coding genes; these are found on the sense strand, implying some linkage of miRNA and host mRNA transcription [12, 14]. Human miRNAs are located in all chromosomes except Y chromosome and they are nonrandomly distributed in the human genome.

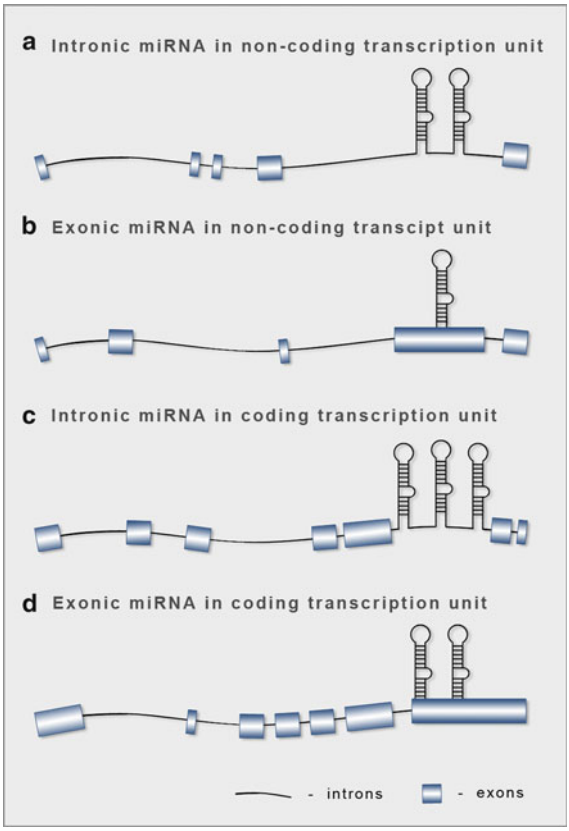


**Fig. 2.1** MicroRNA genomic location and gene structure. MicroRNAs can be categorized into four groups according to their genomic location: (**a**) MiRNAs located in introns of noncoding transcripts, (**b**) miRNAs located in exons of noncoding transcripts, (**c**) miRNAs in introns of protein-coding transcripts, and (**d**) miRNAs in exons of protein-coding transcripts

## 2.2 Biogenesis and Maturation of MicroRNAs

The biosynthesis of miRNAs is a tightly regulated multistep process that starts in the nucleus of the cell, following transcription, and continues through the cytoplasm where finally the mature miRNA molecule exerts its main function (Fig. 2.2). Each one of the multiple steps that compose miRNA biosynthesis seems to be remarkably well coordinated. Drosha initiates the processing by specific cropping of the stem–loop precursor in the nucleus [15]. The resulting structure, precursor miRNA (pre-miRNA), seems to be a signature motif for all dsRNAs that are involved in small-RNA pathways. Exportin-5 recognizes this signature motif and exports pre-miRNAs to the cytoplasm through nuclear pores on a GTP–GDP gradient [16, 17]. Following export, pre-miRNA is handed over to another RNase III enzyme, Dicer, that dices the pre-miRNA into a miRNA duplex that is further unwinded giving rise to the mature functional miRNA molecule.

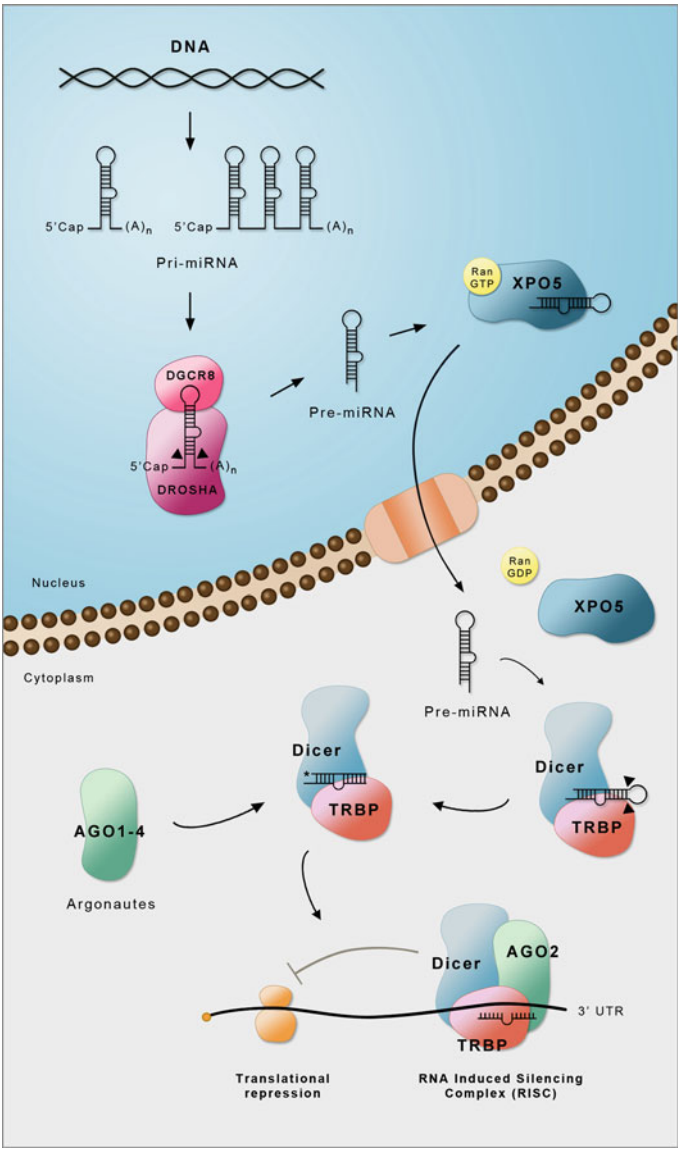


**Fig. 2.2** MicroRNA biogenesis pathway. Long primary transcripts (pri-miRNAs) containing one or several miRNAs are transcribed by RNA polymerase II and cleaved by the microprocessor complex, containing at least Drosha (RNAase III endonuclease) and DGCR8 in humans (a double-stranded RNA-binding protein). This complex recognizes the double-stranded RNA structure of the pri-miRNA and specifically cleaves at the base of the stem–loop, hence releasing a 60- to 70-nucleotide precursor (pre)-miRNA. This pre-miRNA is then exported through the exportin-5 pathway into the cytoplasm where it is further processed into a mature miR/miR* duplex by DICER1, a second RNase III endonuclease together with its catalytic partner TAR-binding protein (TRBP). The miR/miR* duplex is then loaded into a multicomponent complex, the RNA-induced silencing complex (RISC), constituted of at least TRBP, DICER1, and one Argonaute (Ago2 in human). The miR serves as a guide for target recognition while the miR* passenger strand is cleaved by Ago2. Most of animal miRNAs harbor an imperfect homology with their targets and, therefore, inhibit translation by an RISC-dependent mechanism

### 2.2.1 *Canonical Versus Noncanonical Processing of MicroRNAs*

#### 2.2.1.1 MiRNA Transcription and Drosha "Cropping"

The biogenesis of a miRNA begins with the synthesis of a long transcript known as primary miRNA (pri-miRNA). The promoter region of autonomously expressed miRNA genes is highly similar to that of protein-coding genes [18, 19]. pri-miRNAs are typically (although not exclusively) transcribed by RNA polymerase II, which generates the pri-miRNA that consists of one or more hairpin structures, each composed of a stem and a terminal loop. pri-miRNAs are structurally analogous to mRNAs; they are 5′-capped and spliced and bear a 3′ poly-A tail, and they often can produce more than one functional miRNA [20–23]. However, other pathways generate a minor set of miRNAs, especially from genomic repeats (i.e., Alu repeats) whose transcription is carried out by RNA polymerase III [24].

In the canonical miRNA biogenesis pathway the first step of miRNA maturation starts in the nucleus where the pri-miRNA is "cropped" into a ~70 nt hairpin-structured pre-miRNA. This phenomenon is catalyzed by a multiprotein complex called the microprocessor. The core components of this complex are Drosha, an RNase III enzyme, together with its interacting partner DGCR8 (DiGeorge syndrome critical region gene 8), a double-stranded RNA-binding domain (dsRBD) protein [15, 25, 26]. The DGCR8 protein recognizes the stem and the flanking single-stranded RNA (ssRNA) and serves as a ruler for Drosha to cut the stem approximately 11 nt away from the stem–ssRNA junction releasing the pre-miRNA [27]. The pre-miRNA has a 5′ phosphate and a 2 nt 3′ overhang, characteristic for RNase III endonuclease products [15, 28]. RNase III enzymes, like Drosha, are a family of double-stranded RNA (dsRNA) ribonucleases that are expressed in all living cells [29]. The microprocessor appears to represent a heterotetramer consisting of two Drosha and two DGCR8 molecules. Importantly, the ~120 kDa DGCR8 protein contains two consensus dsRNA-binding domains; thus DGCR8 may play a critical role in mediating pri-miRNA recognition and/or binding by the microprocessor complex. Moreover DGCR8 gene is located in a region of chromosome 22 that is monoallelically deleted in patients suffering from a relatively common and very complex genetic disease termed DiGeorge syndrome that raises the possibility of the existence of a link between DiGeorge syndrome and global impairment of miRNA processing [30]. Other cofactors play an important role on the action of the microprocessor complex, including the DEAD box RNA helicases p68 (DDX5) and p72 (DDX17), as well as heterogenous ribonucleoproteins (hnRNPs) [26]. All the components of this complex are necessary for pri-miRNA processing in vivo, as reduction on the level of either Drosha or DGCR8 led to the reduction of both pre-miRNAs and mature miRNAs [25, 26]. Drosha cleavage generates a product with a 2 nt 3′ overhang that is recognized by exportin-5 (XPO5), which transports the pre-miRNA into the cytoplasm through a Ran-GTP-dependent mechanism [21, 31].

While most miRNAs are generated through the concerted action of the referred enzymes, some animal miRNAs adopt alternative pathways for their maturation. A particular subset of miRNAs known as mirtrons are pre-miRNA-like hairpins that bypass the Drosha requirement step by splicing and debranching of short-hairpin introns [32–34]. Mirtron products appear as pre-miRNA mimics from splicing reaction and enter the canonical biogenesis pathway as XPO5 substrates. Also, some small nucleolar RNAs (snoRNAs), transfer RNAs (tRNAs), and endogenous short-hairpin RNAs (shRNAs) are also processed into miRNA-like molecules independently of the microprocessor complex [35–38]. In addition terminal hairpins of endogenous siRNA long-stem–loop precursors are another source of miRNAs that do not go through the Drosha processing step [39]. Therefore, despite Drosha processing being the main route by which biosynthesis of miRNAs takes place in the nucleus, some miRNAs have been described to bypass the microprocessor step and enter the biogenesis pathway directly to XPO5.

#### 2.2.1.2 Nuclear Export and Cytoplasmic Processing of MicroRNAs

Following nuclear processing by Drosha the pre-miRNA is recognized by exportin-5 and transits to the cytoplasm to be further processed [16, 17]. Owing to compartmentalization of the processing events, nuclear export of pre-miRNAs is a crucial step in miRNA biogenesis [40, 41]. Although XPO5 was originally known as a minor export factor for tRNAs, the major cargos of XPO5 turned out to be pre-miRNAs. Exportin-5 is a member of the karyopherin family of nucleocytoplasmic transport factors that depend on their cofactor Ran for their function. As with other nuclear transport receptors, XPO5 binds cooperatively to its cargo and the GTP-bound form of the cofactor Ran in the nucleus, and releases the cargo following the hydrolysis of GTP in the cytoplasm. Ran is a GTPase that binds in the GTP-bound form to karyopherins (XPO5 in this case) and forms a heterotrimer with pre-miRNAs [16, 17]. The XPO5/Ran-GTP heterodimer binds small RNAs bearing a terminal $\geq$16 nt and a short 3′ overhang, i.e., precisely the structure of the pre-miRNA [42]. Knockdown experiments of XPO5 showed little effect on the level of expression of the pri-miRNAs, but did significantly reduce the level of expression of the mature miRNA [17]. Loss and mutation of XPO5 also lead to the nuclear retention of pre-miRNAs [43]. These results identify the major function of XPO5, as a pre-miRNA nuclear exporter, and define it as a cofactor for miRNA biogenesis and function [16, 17].

Once liberated in the cytoplasm, pre-miRNAs are subsequently processed into ~22 nt miRNA duplexes by the cytoplasmic RNase III Dicer together with its catalytic partner *Trans*-activator RNA (tar)-binding protein (TRBP) [44–48]. Dicer has a single processing center with intramolecular dimerization of the two RNase III domains. Each RNase domain cuts independently one RNA strand of the duplex and generates products with 2-nt 3′ overhangs [49]. Mammalian cells have a single Dicer protein, with a molecular weight of ~200 kDa. Dicer contains an ATPase/RNA helicase domain, a DUF domain (a divergent dsRBD), a PAZ domain (a nucleic

acid-binding domain, with a strong preference for single-stranded nucleic acids or RNA duplexes with single-stranded 3′ overhangs), two RNase III domains, and a dsRBD [44]. The PAZ domain of Dicer, also found in Argonaute proteins, is thought to bind to the 2 nt 3′ overhang present at the base of the pre-miRNA hairpin, and the dsRBD of Dicer binds the stem and defines the distance of the cleavage from the base [50]. The two strands of the duplex are separated by Dicer and the strand whose 5′ end forms the more unstable duplex with its partner seems to preferentially survive as the mature miRNA [27, 51–53]. Dicer knockout ($Dcr^{-/-}$) mice and cells are not viable, indicating a key role for this protein during developmental and normal cell function [44].

Dicer associates with several other proteins including TRBP and Argonaute family. TRBP protein is necessary for RNAi function; nonetheless TRBP is not required for the cleavage reaction itself; instead, it has various roles in maintaining Dicer stability and "dicing" action on the cytoplasm [54–57]. The TRBP C-terminal–Dicer interaction and its function as part of the RNA-induced silencing complex (RISC) have been identified as an important component of the RNAi pathway [58]. TRBP binds Dicer in the amino-terminal DExD/H-box helicase domain and activates Dicer through a conformational rearrangement [59]. Following Dicer/TRBP cleavage, the resulting ~22-nt RNA duplex is loaded onto Argonaute proteins so as to generate the effector complex, RISC. RISC is the cytoplasmic effector machine of the miRNA pathway and contains a single-stranded miRNA guiding it to its target mRNAs.

Cytoplasmic miRNA processing and RISC assembly are mediated by the RISC loading complex (RLC). RLC is a multi-protein complex composed of Dicer, TRBP, and the core component Argonaute-2 (AGO2), which also mediates RISC effects on mRNA targets [60, 61]. The slicer activity of AGO2 cleaves the 3′ arm of the hairpin in the middle to generate a nicked hairpin, producing the AGO2-cleaved pre-miRNA [62].

Recently a Dicer-independent AGO2-dependent mechanism has been described for miR-451 [63, 64]. The precursor miR-451 (pre-miR-451) has an unusual secondary structure with a predicted stem of only 17 nt in length that seems to be too short to serve as substrate for Dicer, which requires >19 nt stem in addition to a 2 nt 3′ overhang for efficient hairpin cleavage [65, 66]. Thus it has given rise to a conserved miRNA biogenesis pathway that requires AGO2 catalysis [63, 64]. The conserved miR-451 matures by direct cleavage of its corresponding pre-miRNA via the "slicer" activity of AGO2 [63, 64].

#### 2.2.1.3 MicroRNAs and Posttranscriptional Regulation of Gene Expression

The small size of miRNAs provides a limited amount of sequence information for specificity. Partial base-pairing between a miRNA and its target mRNA is sufficient for repression and/or degradation of the target; thus a wide net of mRNAs can be regulated by the same miRNA. MiRNAs select mRNA targets for downregulation

through the association with a large, multi-protein complex, the RISC. This selection requires the presence of sequences within the target mRNA which are imperfectly complementary to the miRNA sequence. MiRNA-binding sites commonly occur within the 3′-untranslated region (3′-UTR) of the mRNA, but functional miRNA-binding sites can also occur with the 5′UTR or coding region [67, 68].

The 3′-UTRs of messenger RNAs serve as docking platforms for miRNAs and RNA-binding proteins, which control mRNA stability, localization, and translation [69, 70]. The most common feature is perfect base-pairing between nucleotides 2 and 7 at the 5′ end of the miRNA, which is called the "seed" sequence, and the target site [69]. There is still a great deal of uncertainty regarding the exact composition of the miRNA complex as well as the mechanisms used to control target gene expression. What is known is that miRISC inhibits the expression of mRNAs basically in one of the two ways depending on the degree of complementarity between miRNA and the target. There are mismatches and bulges in most miRNA target sites; therefore the degree of complementarity between the miRNA and the target is thought to be a major determinant in distinguishing the two mechanisms of post-transcriptional silencing: translation inhibition or mRNA degradation. Several other mechanisms have been however documented, including translational inhibition at the level of initiation and elongation, rapid degradation of the nascent peptide, and mRNA degradation (Fig. 2.3). The core component of the mRNP complex is the Argonaute protein and in mammals one of the four Argonaute proteins (AGO1-4) is recruited into the complex. AGO2 is the only Argonaute with "slicer" activity and is responsible for the cleavage of the mRNA target midway into the complementary region. Other proteins are also essential for this silencing complex to work, such as the RNA-binding protein fragile-X-mental-retardation protein (FMRP), the p-body marker GW182, and the decapping activator RCK/p54, which likely dictate how the silencing of the mRNA target will occur [71–76].

## 2.3 Regulation of MicroRNA Biogenesis

Most miRNAs are under the control of developmental and tissue-specific signaling [77]. Stringent control of miRNA levels is crucial to maintain normal cellular functions, and deregulation of miRNA expression is often associated with human diseases, such as cancer [78]. Therefore it is highly important to have the maturation pathway of miRNAs subject to intense regulation at various levels, from stability, processing, sequence identity, and binding to target mRNAs. Likewise the miRNA biogenesis pathway is often subjected to feedback regulation. These autoregulatory feedback loops are a common mechanism particularly important in cell fate determination and development. MiRNAs are perfectly suited to participate in these networks owing to their potential to directly repress mRNAs that encode factors involved in the biogenesis or function of the same miRNAs. The control of the miRNA biosynthesis pathway is emerging as an important mechanism in defining the spatiotemporal pattern of miRNA expression in cancer cells. Interestingly transcriptional regulation is probably the main control mechanism.

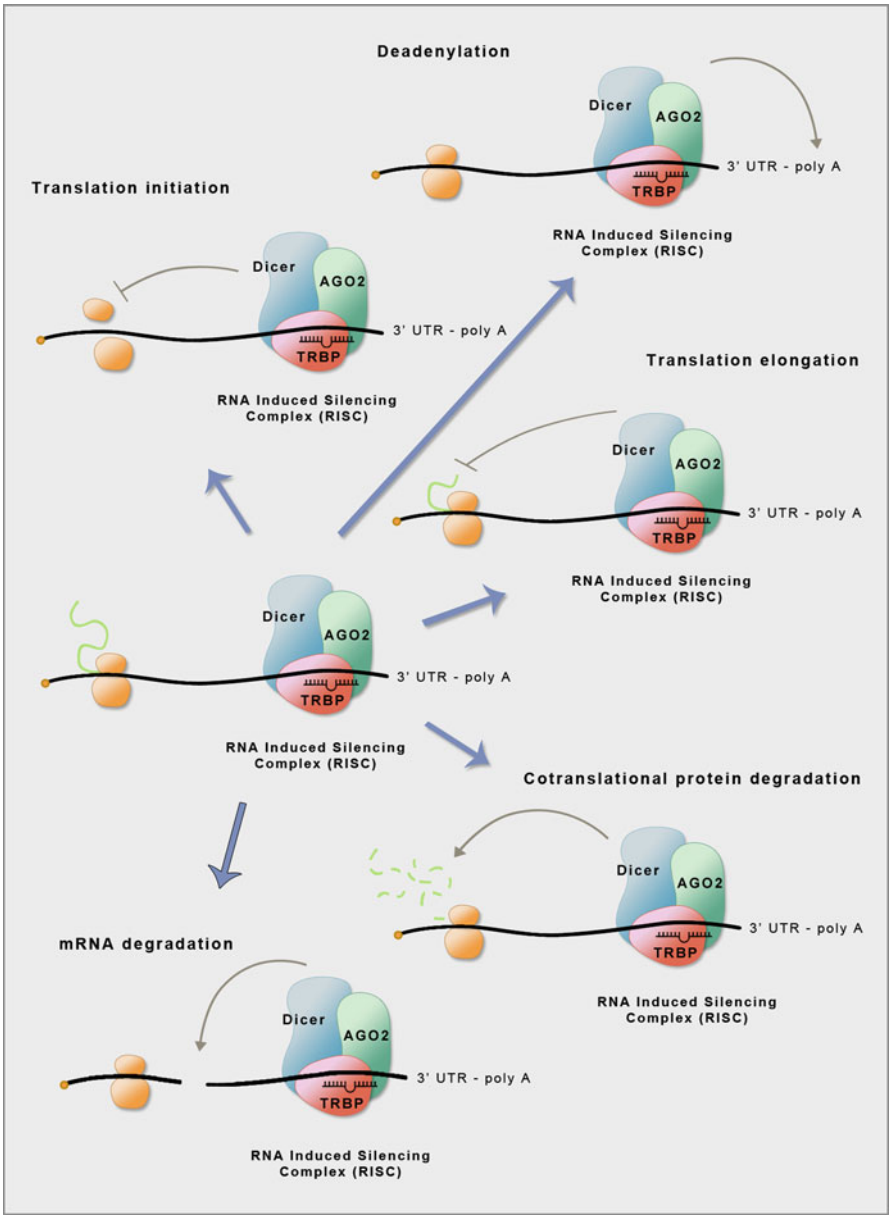


**Fig. 2.3** Mechanisms of action of miRNAs. The silencing mechanisms modulated by miRNAs remain under debate. The ones clarified until now involve translational repression, degradation, and/or deadenylation of target mRNAs, as well as cotranslational protein degradation

### 2.3.1 Regulation of MicroRNA Transcription

Transcription is a major point of regulation in miRNA biogenesis. Many characteristics of miRNA gene promoters, such as the existence of CpG islands, TATA box, TFIIB recognition, initiator elements, and histone modifications, are similar to the promoters of protein-coding genes; therefore the same mechanisms of gene expression control are applicable to miRNAs [18, 19]. MiRNA genes can be transcribed by RNA polymerase II or III; as both RNA polymerases are regulated differently and recognize specific promoter and terminator elements, they facilitate a wide range of regulatory options. In addition, many miRNAs are encoded in the genome as clusters and can be transcribed as long polycistronic primary transcripts; however it has been shown that each miRNA located in the same genomic cluster can be transcribed and regulated independently [79, 80].

MiRNA genes also contain DNA-binding factors that can be regulated by the action of transcription factors such as p53 and c-myc, as well as MEF2, PU.1, and REST [81–84]. The tumor-suppressor p53 activates the miR-34 family of miRNAs, whereas the oncogenic protein c-myc transactivates or represses a number of miRNAs that are involved in the cell cycle and apoptosis [11, 85]. It is well demonstrated that c-myc binds to E-boxes and activates transcription of the miR-17-92 cluster [84]. Consistent with the c-myc activation in tumors, miR-17-92 is often highly expressed in tumors [86]. Moreover, c-myc is shown to increase transcription of Lin-28B that mediates the posttranscriptional repression of let-7 miRNA family members [87]. Furthermore, transcription activation can occur in response to growth factor stimulation, including platelet-derived growth factor (PDGF), transforming growth factor-β (TGF-β), and bone-derived neurotrophic factor [19, 84, 88–90].

Epigenetic control also contributes to miRNA gene regulation through the methylation of CpG islands at the promoter of miRNA genes that silences their transcription [91–93]. MiRNA promoters are also regulated by histone modifications during development and disease [94].

### 2.3.2 Regulation of the Cropping, Dicing, and Slicing Events

Drosha processing (cropping) constitutes an important point of regulation. The total levels of Drosha and DGCR8 in the cell are tightly controlled and play a role in the regulation of pri-miRNA processing. DGCR8 protein has a stabilizing effect on Drosha through the interaction with its middle domain, whereas Drosha controls DGCR8 levels by cleaving hairpins present in the DGCR8 mRNA, thereby inducing its degradation [95, 96]. Therefore a tight equilibrium needs to exist between the amounts of Drosha and DGCR8 to maintain the processing of pri-miRNAs flowing. In addition SMAD proteins, activated by BMP/TGFβ, interact with Drosha and DDX5/p68 stimulating Drosha processing [97]. Likewise it was recently described that the tumor-suppressor breast cancer 1 (BRCA1) accelerates the processing of pri-miRNAs by direct interaction with Drosha and DDX5 [98]. In addition the

authors described that BRCA1 recognizes the pri-miRNA secondary structure and directly binds the pri-miRNA before Drosha interaction [98].

In the cytoplasm accumulation of Dicer is dependent on its partner TRBP, and a decrease in TRBP leads to Dicer destabilization and pre-miRNA processing impairment [54, 55, 99]. Truncating mutations of TRBP that cause lower TRBP protein expression are associated with both a defect in the processing of miRNAs and destabilization of the Dicer protein [54]. TRBP itself is stabilized by phosphorylation mediated by mitogen-activated protein kinase (MAPK)/extracellular regulated kinase (ERK) [99]. Notably, let-7 and miR-107 can target Dicer mRNA forming a feedback loop with the potential to influence miRNA biogenesis at the Dicer dicing step [68, 100].

Moreover, there are a significant number of RNA-binding proteins that influence miRNA processing at the Drosha and Dicer levels through specific interaction with a subset of pri-miRNAs or pre-miRNAs. The RNA-binding protein LIN28 is the best-studied negative regulator of miRNA biogenesis, which can act at different levels. LIN-28 binding to the terminal loop of pri-let-7 interferes with cleavage by Drosha; likewise binding of LIN-28 to the pre-let-7 can also block its processing by Dicer through polyuridylation of pre-let-7 performed by the enzyme TUT4 [101–105]. In addition, Kedde and colleagues demonstrated that the expression of dead end 1 protein (Dnd1), an evolutionary conserved RNA-binding protein, counteracts the function of several miRNAs in human cells and in primordial germ cells of zebrafish by binding mRNAs and prohibiting miRNAs from associating with their target sites. These effects of Dnd1 are mediated through uridine-rich regions present in the miRNA-targeted mRNAs [106]. Likewise, the RNA-binding protein RBM38 has been described to control miRNA access to target mRNAs [107]. Target selectivity is determined by the interaction of RBM38 with uridine-rich regions near miRNA target sequences [107].

RNA editing is another possible way of regulating miRNA biogenesis through the action of adenosine deaminases acting on RNA (ADARs), enzymes that modify adenosine (A) into inosine (I). Adenosine-to-inosine editing of pri-miRNAs can lead to changes in miRNA stability or alter their target selection [108]. Editing can also change the target specificity of the miRNA if it occurs in miRNA sequences [108].

## 2.4 Biological Roles of MicroRNAs

The importance of faithful miRNA expression has been involved in numerous cellular events during animal development where it is pivotal for the timing and regulation of many key processes such as cell fate determination, proliferation, and cell death [109]. In addition to these vital processes, miRNAs are implicated in other multiple biological functions, such as immune response, insulin secretion, neurotransmitter synthesis, circadian rhythm, and viral replication [110–113]. This list will undoubtedly expand as experimental data accumulates.

Representative examples of the wide scope of miRNA regulation are found in the miRNA let-7, a miRNA critical for developmental timing [5]; a developmentally regulated miRNA, bantam, that controls cell proliferation via regulation of apoptosis [114]; miRNAs that control ES cell differentiation by stabilizing the self-renewing versus differentiated cell fates [115]; and stem cell division [116]. Specifically, let-7 miRNA is essential for normal development and has a temporally regulated expression pattern. The misexpression of let-7 has been linked to several types of human cancer. Other examples are miR-196, which is involved in hind limb development [117], and the brain-specific miR-134, which is necessary for synaptic development and plasticity [118]. Skin differentiation is promoted by miR-203, which represses p63 in stratified epithelial tissues, while precise levels of miR-1 are critical in cardiogenesis [119]. Normal immune function is dependent on miR-155 and B-cell differentiation is controlled by miR-150-mediated repression of the transcription factor c-Myb [120, 121]. In addition, the pancreatic islet cell-specific miR-375 regulates insulin secretion by inhibiting myotrophin, a component of the exocytosis pathway [110].

Recent studies implicated miRNAs with several different diseases besides cancer. MiRNAs are not only required for the development of early embryonic stem cell survival and differentiation, but also plays an important role in maintaining the survival of mature neurons and their function. In neurological diseases, the loss of miR-20a/b-1 cluster has been implicated in Alzheimer's disease and the loss of miR-133b may contribute to the decrease in dopaminergic neurons seen in Parkinson's disease [122, 123]. In heart disease the expression of miR-21 in cardiac fibroblasts contributes to interstitial fibrosis and cardiac hypertrophy, while miR-1 and miR-133 in cardiomyocytes protect against hypertrophy [124]. It was also shown that autosomal dominant progressive hearing loss is a consequence of point mutations in the seed region of miR-96, a miRNA expressed in hair cells of the inner ear [125]. Therefore miRNAs have been found deregulated in a wide range of human diseases but specifically have been described to play an important role in human cancer. However, it remains uncertain whether altered miRNA expression is a cause or a consequence of pathological processes.

## 2.5 MicroRNA Biogenesis and Cancer

### *2.5.1 Oncogenic or Tumor-Suppressor MicroRNAs*

In line with their broad effects, miRNAs have been proposed to function as oncogenes or tumor-suppressor genes given their inhibition of a variety of tumor-suppressive and oncogenic mRNAs, respectively [126, 127]. Overexpressed miRNAs in cancer, such as miR-17-92, may function as oncogenes and promote cancer development by negatively regulating tumor-suppressor genes and/or genes that control cell differentiation or apoptosis. Non-expressed or low-expressed miRNAs in cancer, such as let-7, function as tumor-suppressor genes and may inhibit

cancers by regulating oncogenes. In particular, three distinct mechanisms have been posited. First, oncogenic miRNAs can undergo a gain of function in tumors. This has been more clearly demonstrated for the miR-17~92 cluster, whose amplification in B-cell lymphomas promotes their development, potentially through its control of B-cell differentiation [11, 128]. Furthermore, tumor-suppressive miRNAs could undergo loss of function in tumors. This has been shown for several miRNAs, including the let-7 family, whose expression can limit tumorigenesis through inhibition of oncogenes like the *RAS* family and *HMGA2* [129, 130]. In particular, let-7 family members are in sites frequently deleted in human tumors, and their processing is inhibited by the oncogenic Lin-28 proteins [103, 131, 132]. Finally, oncogenes can acquire mutations to remove miRNA-binding sites in tumors. This has been described for *HMGA2*, whose translocation promotes lipoma development by releasing the transcript from let-7-mediated tumor suppression [133]. Furthermore, Voorhoeve and colleagues have identified miR-372 and miR-373 as oncogenes in human testicular germ cell tumors by numbing the p53 pathway [134]. These miRNAs neutralize p53-mediated CDK inhibition, possibly through direct inhibition of the expression of the tumor-suppressor LATS2 [134].

### *2.5.2 Impairment of MicroRNA Biogenesis in Cancer*

Widespread downregulation of miRNAs is a common feature of human cancers when compared with normal tissues [135]. Whether this tendency is a reflection of a pattern associated with specific cells of origin, is a consequence of the malignant state, or actively contributes to cancer development is still unclear. Two main mechanisms have been proposed as the cause of the global downregulation of miRNAs in cancer. One is related with transcriptional repression by oncogenic transcription factors like MYC oncoprotein [85]. The other mechanism involves impaired miRNA biogenesis and is based on the observation that cancer cells often display reduced levels of proteins involved in the miRNA biogenesis pathway [136]. In addition, global repression of miRNA biogenesis by suppression of the key components of miRNA processing machinery, such as Drosha, DGCR8, Dicer, TRBP, and XPO5, promotes cellular transformation and tumorigenesis [43, 54, 85, 137]. While the mechanism(s) remains to be fully elucidated, it suggests that miRNAs might have an intrinsic function in tumor suppression and its downregulation eventually accelerates oncogenesis.

After transcription, global levels of miRNAs can be reduced by impaired miRNA biogenesis. Inactivating mutations and reduced expression have been described for almost all the members of the miRNA processing machinery. Furthermore the existence of frameshift mutations in the TARBP2 and XPO5 genes in cancer cell lines and primary tumors has been recently reported [43, 54, 138]. These genetic alterations in miRNA biogenesis genes are associated with nonfunctional truncated forms of both proteins. The first mutation to be described was in the TARBP2 gene that codes for the TRBP protein, the catalytic partner of Dicer. Frameshift mutations in

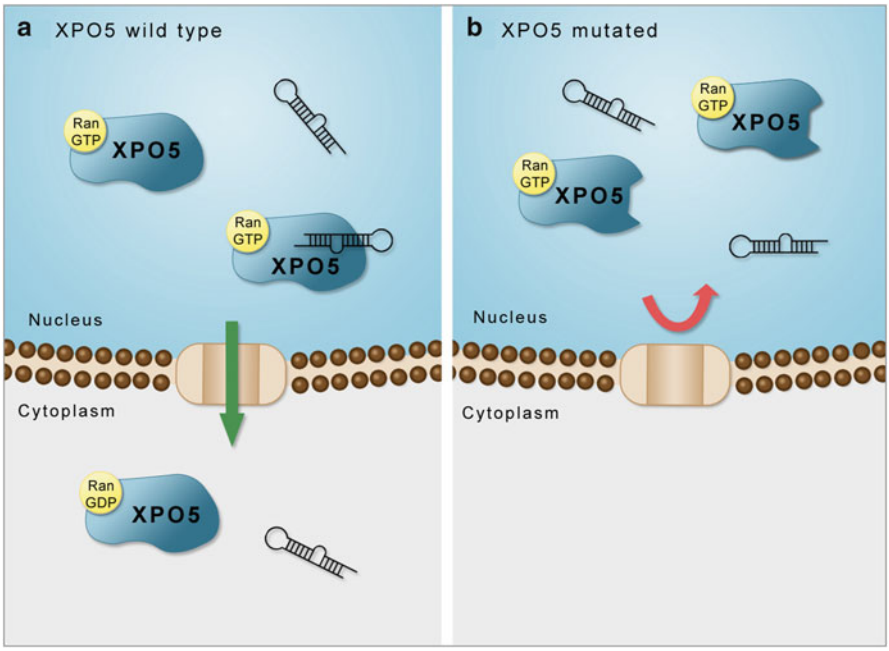


**Fig. 2.4** Impairment of miRNA biogenesis in cancer at the exportin-5 step. (**a**) In a normal setting the formation of a heterotetrameric complex between XPO5/Ran/GTP/pre-miRNA results in export of the pre-miRNA from the nucelus to the cytoplasm to be further processed. (**b**) In cancer, heterozygous frameshift mutations in the XPO5 gene lead to a truncated form of the XPO5 protein that does not recognize/bind the pre-miRNA molecule, therefore preventing the formation of a functional XPO5/Ran/GTP/pre-miRNA complex and its translocation from the nucleus to the cytoplasm (*right panel*)

TARBP2 result in monoallelic expression of wild-type TRBP. Thus, due to TRBP downregulation Dicer protein is destabilized slowing down the dicing process of pre-miRNAs culminating in a global downregulation of miRNAs due to impaired processing [54]. Likewise, the mutations found in exon 32 of XPO5 alter and truncate the protein sequence and prevent XPO5 from associating with its pre-miRNA cargo and exiting the nucleus (Fig. 2.4). In XPO5 heterozygous mutant cells, less pre-miRNA was accessible to processing by the cytoplasmic machinery, resulting in decreased mature miRNA levels and enhanced tumorigenicity. Many pre-miRNAs are also targeted by ADARs at various stages of their processing, and the modification can also prevent export of pre-miRNAs. In addition germline truncating mutations in Dicer, that is essential for processing miRNAs, have been observed in families with the pleuropulmonary blastoma family tumor and dysplasia syndrome [139]. Recent work has suggested that other components of the miRNA biogenesis pathway, Dicer and TARBP2, are haploinsufficient tumor suppressors [54, 140]. Moreover, biallelic deletion was found to impair cell viability, hence preventing the phenomenon of loss-of-heterozygosity (LOH) [140]. This is also the case for XPO5,

where mimicking LOH by RNA interference against the XPO5 wild-type transcript rendered cells unviable [43]. In addition, the miRISC components AGO2, TNRC6A, and TNRC6C can also be mutated in cancer, although the functional consequences remain to be evaluated [138].

Two other groups have also reported the accumulation of let7 miRNA precursors at various stages during fruit fly and sea urchin development [46, 141]. Although this could represent a defect in RNA processing, it is also possible that the nuclear export of the let7 precursor, and hence access to cytoplasmic Dicer and TRBP, is developmentally regulated.

Hence, it is possible that global dys-regulation of miRNAs in cancer has its roots in impaired miRNA processing through genetic mutations of their components contributing to cancer progression.

**Acknowledgment** We would like to show our special thanks to Dr. Nuno Barros for the design of the figures in this chapter.

## References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281–97.
2. Molnar A et al. miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature. 2007;447(7148):1126–9.
3. Zhao T et al. A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. Genes Dev. 2007;21(10):1190–203.
4. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843–54.
5. Reinhart BJ et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901–6.
6. Lagos-Quintana M et al. Identification of novel genes coding for small expressed RNAs. Science. 2001;294(5543):853–8.
7. Abbott AL et al. The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. Dev Cell. 2005;9(3):403–14.
8. Miska EA et al. Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS Genet. 2007;3(12):e215.
9. Lau NC et al. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science. 2001;294(5543):858–62.
10. Calin GA et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101(9):2999–3004.
11. He L et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043):828–33.
12. Rodriguez A et al. Identification of mammalian microRNA host genes and transcription units. Genome Res. 2004;14(10A):1902–10.
13. Smalheiser NR, Torvik VI. Mammalian microRNAs derived from genomic repeats. Trends Genet. 2005;21(6):322–6.
14. Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA. 2005;11(3):241–7.

15. Lee Y et al. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003;425(6956):415–9.
16. Lund E et al. Nuclear export of microRNA precursors. Science. 2004;303(5654):95–8.
17. Yi R et al. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003;17(24):3011–6.
18. Ozsolak F et al. Chromatin structure analyses identify miRNA promoters. Genes Dev. 2008;22(22):3172–83.
19. Corcoran DL et al. Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. PLoS One. 2009;4(4):e5279.
20. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009;136(4):642–55.
21. Kim W et al. Histone acetyltransferase GCN5 interferes with the miRNA pathway in Arabidopsis. Cell Res. 2009;19(7):899–909.
22. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004;10(12):1957–66.
23. Lee Y et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051–60.
24. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006;13(12):1097–101.
25. Denli AM et al. Processing of primary microRNAs by the Microprocessor complex. Nature. 2004;432(7014):231–5.
26. Gregory RI et al. The Microprocessor complex mediates the genesis of microRNAs. Nature. 2004;432(7014):235–40.
27. Han J et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006;125(5):887–901.
28. Basyuk E et al. Human let-7 stem-loop precursors harbor features of RNase III cleavage products. Nucleic Acids Res. 2003;31(22):6593–7.
29. Conrad C, Rauhut R. Ribonuclease III: new sense from nuisance. Int J Biochem Cell Biol. 2002;34(2):116–29.
30. Shiohama A et al. Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. Biochem Biophys Res Commun. 2003;304(1):184–90.
31. Okada C et al. A high-resolution structure of the pre-microRNA nuclear export machinery. Science. 2009;326(5957):1275–9.
32. Berezikov E et al. Mammalian mirtron genes. Mol Cell. 2007;28(2):328–36.
33. Okamura K et al. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell. 2007;130(1):89–100.
34. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature. 2007;448(7149):83–6.
35. Babiarz JE et al. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. 2008;22(20): 2773–85.
36. Ender C et al. A human snoRNA with microRNA-like functions. Mol Cell. 2008; 32(4):519–28.
37. Saraiya AA, Wang CC. snoRNA, a novel precursor of microRNA in Giardia lamblia. PLoS Pathog. 2008;4(11):e1000224.
38. Cole C et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA. 2009;15(12):2147–60.
39. Miyoshi K et al. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. RNA. 2010;16(3):506–15.
40. Kim VN. MicroRNA precursors in motion: exportin-5 mediates their nuclear export. Trends Cell Biol. 2004;14(4):156–9.

41. Murchison EP, Hannon GJ. miRNAs on the move: miRNA biogenesis and the RNAi machinery. Curr Opin Cell Biol. 2004;16(3):223–9.
42. Gwizdek C et al. Terminal minihelix, a novel RNA motif that directs polymerase III transcripts to the cell cytoplasm. Terminal minihelix and RNA export. J Biol Chem. 2001;276(28): 25910–8.
43. Melo SA et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. Cancer Cell. 2010;18(4):303–15.
44. Bernstein E et al. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001;409(6818):363–6.
45. Grishok A et al. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell. 2001; 106(1):23–34.
46. Hutvagner G et al. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001;293(5531):834–8.
47. Ketting RF et al. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. 2001;15(20):2654–9.
48. Knight SW, Bass BL. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science. 2001;293(5538):2269–71.
49. Filipowicz W. RNAi: the nuts and bolts of the RISC machine. Cell. 2005;122(1):17–20.
50. Zhang H et al. Single processing center models for human Dicer and bacterial RNase III. Cell. 2004;118(1):57–68.
51. Salzman DW, Shubert-Coleman J, Furneaux H. P68 RNA helicase unwinds the human let-7 microRNA precursor duplex and is required for let-7-directed silencing of gene expression. J Biol Chem. 2007;282(45):32773–9.
52. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003;115(2):209–16.
53. Schwarz DS et al. Asymmetry in the assembly of the RNAi enzyme complex. Cell. 2003;115(2):199–208.
54. Melo SA et al. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. Nat Genet. 2009;41(3):365–70.
55. Chendrimada TP et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature. 2005;436(7051):740–4.
56. Haase AD et al. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. EMBO Rep. 2005;6(10):961–7.
57. Rossi JJ. RNAi and the P-body connection. Nat Cell Biol. 2005;7(7):643–4.
58. Daniels SM et al. Characterization of the TRBP domain required for dicer interaction and function in RNA interference. BMC Mol Biol. 2009;10:38.
59. Ma E et al. Autoinhibition of human dicer by its internal helicase domain. J Mol Biol. 2008;380(1):237–43.
60. Gregory RI et al. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell. 2005;123(4):631–40.
61. MacRae IJ et al. In vitro reconstitution of the human RISC-loading complex. Proc Natl Acad Sci U S A. 2008;105(2):512–7.
62. Matranga C et al. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell. 2005;123(4):607–20.
63. Cheloufi S et al. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 2010;465(7298):584–9.
64. Cifuentes D et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science. 2010;328(5986):1694–8.
65. Siolas D et al. Synthetic shRNAs as potent RNAi triggers. Nat Biotechnol. 2005;23(2): 227–31.
66. Griffiths-Jones S et al. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008;36(Database issue):D154–8.

67. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. Proc Natl Acad Sci U S A. 2007; 104(23):9667–72.
68. Forman JJ, Legesse-Miller A, Coller HA. 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. 2008;105(39):14879–84.
69. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136(2):215–33.
70. Rigoutsos I. New tricks for animal microRNAS: targeting of amino acid coding regions at conserved and nonconserved sites. Cancer Res. 2009;69(8):3245–8.
71. Jin P, Alisch RS, Warren ST. RNA and microRNAs in fragile X mental retardation. Nat Cell Biol. 2004;6(11):1048–53.
72. Wulczyn FG et al. Post-transcriptional regulation of the let-7 microRNA during neural cell specification. FASEB J. 2007;21(2):415–26.
73. Jakymiw A et al. Disruption of GW bodies impairs mammalian RNA interference. Nat Cell Biol. 2005;7(12):1267–74.
74. Liu J et al. A role for the P-body component GW182 in microRNA function. Nat Cell Biol. 2005;7(12):1261–6.
75. Chu CY, Rana TM. Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol. 2006;4(7):e210.
76. Eulalio A et al. Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes Dev. 2007;21(20):2558–70.
77. Landgraf P et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007;129(7):1401–14.
78. Jiang Q et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009;37(Database issue):D98–104.
79. Diederichs S, Haber DA. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. Cancer Res. 2006;66(12):6097–104.
80. Yanaihara N et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9(3):189–98.
81. Sampson VB et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res. 2007;67(20):9762–70.
82. Raver-Shapira N et al. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. Mol Cell. 2007;26(5):731–43.
83. He L et al. A microRNA component of the p53 tumour suppressor network. Nature. 2007;447(7148):1130–4.
84. Davis BN et al. Induction of microRNA-221 by platelet-derived growth factor signaling is critical for modulation of vascular smooth muscle phenotype. J Biol Chem. 2009;284(6):3728–38.
85. Chang TC et al. Widespread microRNA repression by Myc contributes to tumorigenesis. Nat Genet. 2008;40(1):43–50.
86. Garzon R et al. MicroRNA fingerprints during human megakaryocytopoiesis. Proc Natl Acad Sci U S A. 2006;103(13):5078–83.
87. Chang TC et al. Lin-28B transactivation is necessary for Myc-mediated let-7 repression and proliferation. Proc Natl Acad Sci U S A. 2009;106(9):3384–9.
88. Chan MC et al. Molecular basis for antagonism between PDGF and the TGFbeta family of signalling pathways by control of miR-24 expression. EMBO J. 2010;29(3):559–73.
89. Kato M et al. MicroRNA-192 in diabetic kidney glomeruli and its function in TGF-beta-induced collagen expression via inhibition of E-box repressors. Proc Natl Acad Sci U S A. 2007;104(9):3432–7.
90. Vo N et al. A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. Proc Natl Acad Sci U S A. 2005;102(45):16426–31.
91. Lujambio A et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res. 2007;67(4):1424–9.

92. Lujambio A et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A. 2008;105(36):13556–61.
93. Davalos V et al. Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. Oncogene. 2012;31(16): 2062–74.
94. Scott GK et al. Rapid alteration of microRNA levels by histone deacetylase inhibition. Cancer Res. 2006;66(3):1277–81.
95. Han J et al. Posttranscriptional crossregulation between Drosha and DGCR8. Cell. 2009; 136(1):75–84.
96. Triboulet R et al. Post-transcriptional control of DGCR8 expression by the Microprocessor. RNA. 2009;15(6):1005–11.
97. Davis BN et al. SMAD proteins control DROSHA-mediated microRNA maturation. Nature. 2008;454(7200):56–61.
98. Kawai S, Amano A. BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. J Cell Biol. 2012;197(2):201–8.
99. Paroo Z et al. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. Cell. 2009;139(1):112–22.
100. Graziano F et al. Genetic modulation of the Let-7 microRNA binding to KRAS 3′-untranslated region and survival of metastatic colorectal cancer patients treated with salvage cetuximab-irinotecan. Pharmacogenomics J. 2010;10(5):458–64.
101. Lehrbach NJ et al. LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. Nat Struct Mol Biol. 2009;16(10):1016–20.
102. Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol. 2009;16(10):1021–5.
103. Heo I et al. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell. 2008;32(2):276–84.
104. Heo I et al. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell. 2009;138(4):696–708.
105. Jones MR et al. Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nat Cell Biol. 2009;11(9):1157–63.
106. Kedde M et al. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell. 2007;131(7):1273–86.
107. Leveille N et al. Selective inhibition of microRNA accessibility by RBM38 is required for p53 activity. Nat Commun. 2011;2:513.
108. Wulff BE, Nishikura K. Modulation of microRNA expression and function by ADARs. Curr Top Microbiol Immunol. 2012;353:91–109.
109. Stefani G, Slack FJ. Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol. 2008;9(3):219–30.
110. Poy MN et al. A pancreatic islet-specific microRNA regulates insulin secretion. Nature. 2004;432(7014):226–30.
111. Jopling CL et al. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science. 2005;309(5740):1577–81.
112. Greco SJ, Rameshwar P. MicroRNAs regulate synthesis of the neurotransmitter substance P in human mesenchymal stem cell-derived neuronal cells. Proc Natl Acad Sci U S A. 2007; 104(39):15484–9.
113. Gantier MP, Sadler AJ, Williams BR. Fine-tuning of the innate immune response by microRNAs. Immunol Cell Biol. 2007;85(6):458–62.
114. Brennecke J et al. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell. 2003;113(1): 25–36.
115. Tay Y et al. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature. 2008;455(7216):1124–8.
116. Hatfield SD et al. Stem cell division is regulated by the microRNA pathway. Nature. 2005;435(7044):974–8.

117. Hornstein E et al. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. Nature. 2005;438(7068):671–4.
118. Schratt GM et al. A brain-specific microRNA regulates dendritic spine development. Nature. 2006;439(7074):283–9.
119. Kwon C et al. MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. Proc Natl Acad Sci U S A. 2005;102(52):18986–91.
120. Rodriguez A et al. Requirement of bic/microRNA-155 for normal immune function. Science. 2007;316(5824):608–11.
121. Xiao C et al. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell. 2007;131(1):146–59.
122. Hebert SS et al. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. Proc Natl Acad Sci U S A. 2008;105(17):6415–20.
123. Kim J et al. A MicroRNA feedback circuit in midbrain dopamine neurons. Science. 2007;317(5842):1220–4.
124. Care A et al. MicroRNA-133 controls cardiac hypertrophy. Nat Med. 2007;13(5):613–8.
125. Mencia A et al. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet. 2009;41(5):609–13.
126. Plasterk RH. Micro RNAs in animal development. Cell. 2006;124(5):877–81.
127. Ventura A, Jacks T. MicroRNAs and cancer: short RNAs go a long way. Cell. 2009;136(4):586–91.
128. Koralov SB et al. Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. Cell. 2008;132(5):860–74.
129. Esquela-Kerscher A et al. The let-7 microRNA reduces tumor growth in mouse models of lung cancer. Cell Cycle. 2008;7(6):759–64.
130. Kumar MS et al. Suppression of non-small cell lung tumor development by the let-7 microRNA family. Proc Natl Acad Sci U S A. 2008;105(10):3903–8.
131. Newman MA, Thomson JM, Hammond SM. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. RNA. 2008;14(8):1539–49.
132. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. Science. 2008;320(5872):97–100.
133. Mayr C, Hemann MT, Bartel DP. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science. 2007;315(5818):1576–9.
134. Voorhoeve PM et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell. 2006;124(6):1169–81.
135. Lu J et al. MicroRNA expression profiles classify human cancers. Nature. 2005; 435(7043):834–8.
136. Thomson JM et al. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. Genes Dev. 2006;20(16):2202–7.
137. Kumar MS et al. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet. 2007;39(5):673–7.
138. Kim MS et al. Somatic mutations and losses of expression of microRNA regulation-related genes AGO2 and TNRC6A in gastric and colorectal cancers. J Pathol. 2010;221(2):139–46.
139. Heravi-Moussavi A et al. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. N Engl J Med. 2012;366(3):234–42.
140. Kumar MS et al. Dicer1 functions as a haploinsufficient tumor suppressor. Genes Dev. 2009;23(23):2700–4.
141. Pasquinelli AE et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature. 2000;408(6808):86–9.

# Chapter 3
# Detecting Noncoding RNA Expression: From Arrays to Next-Generation Sequencing

**Catalina Perdomo, Joshua Campbell, and Frank Schembri**

**Abstract** Detection and quantification of noncoding(nc) RNA species can present specific challenges as compared to mRNA. Among them, ncRNA sequences are generally less well annotated, can include extraordinarily small species such as miRNA or piRNA, can have repetitive sequences or have high GC content, or even have antisense expression. Despite this, many different traditional technologies have been adapted to measure ncRNAs and include those based on a priori knowledge of sequence such as probe detection (qRT-PCR) and hybridization (arrays), and those where such knowledge is not required (next-generation sequencing). A summary of these experimental techniques is reviewed in this chapter, including the available chemistries, throughput, starting material, and species interrogated. Subsequent focus is on the computational analysis for next-generation sequencing since this technology can not only detect and measure ncRNAs but more excitingly also lead to the discovery of new or isomeric forms.

**Keywords** microRNA • LncRNA • Real-time PCR • Microarrays • Next-generation sequencing

C. Perdomo (✉) • J. Campbell
Division of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, 72 East Concord Street, Evans 637, Boston, MA 02118, USA
e-mail: cperdomo@bu.edu

F. Schembri
Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_3,

## 3.1 Introduction

There are many different technologies currently available for ncRNA detection which are essentially reflective of, and adapted from, those used to measure coding RNAs. Traditional older methods of RNA detection and quantification such as in situ hybridization [1] and northern blot [2] are indeed applicable to noncoding species, both large and small; however, this chapter focuses on more modern and higher throughput methods. ncRNA detection methods can be considered in two broad groups: those based on prior knowledge and annotation of sequences (i.e., probe based detection like qRT-PCR [3], microarrays [4]) and those where a priori knowledge of sequence is not required (i.e., next-generation sequencing [5]). Since regulatory ncRNAs are generally classified by size (greater or less than 200 nt), in our discussion we further subdivide those technologies which have been optimized for each of these sizes.

Designing assays for ncRNA, both large and small species, pose some unique challenges as compared to messenger RNA. First and foremost, a comprehensive annotation database must be in place in order to facilitate primer or probe design. Over the last 10 years, the public repository of miRNA sequences, miRBase [6], has been expanded and refined, defining both their primary structure (i.e., sequence) and secondary structures (i.e., hairpin). These have led to an explosion in the available commercial products for microRNA detection [7]. Notably, several characteristics of miRNA create significant challenges for PCR or probe based assays such as their small size (~20–25 nt), lack of polyadenylation, high conservation between family members (differing by only 1 or 2 nucleotides), and tendency to bind targets with imperfect homology [8, 9]. Additionally, the presence of "isomirs" has been reported, which are essentially variants of a miRNA canonical sequence which may be present in relatively high proportions and not differentiated by a probe or primer designed for the primary sequence [10].

Primer and probe design for larger species of ncRNA (lncRNA) is limited by annotation, which while quickly improving, is still at an earlier stage than microRNA, with differing numbers of annotations spread across smaller databases such lncRNAdb [11], Human Body Map lincRNAs [12], Ensembl [13], USCS [14], RNAdb [15], RefSeq [16] but without uniform agreement. Primer and probe design for long noncoding RNAs is considered difficult, since many lncRNAs are found in long intergenic regions with high GC content that have repetitive sequences or even circular structures [17, 18]. Additionally, a large percentage of lncRNAs are antisense transcripts of known protein coding genes, but present in far less abundance. Therefore, strategies such as poly-A tailing and tagging are sometimes employed to ensure strand specific PCR amplification. Because of these difficulties in computational lncRNA primer design, the most robust of primers available are clearly those which have been experimentally validated. Next-generation sequencing is particularly exciting in the area of ncRNA detection since it avoids primer and probe design issues and allows for the discovery of novel ncRNA species and isomirs.

## 3.2 Probe Based Detection, Assays Based on qRT-PCR

One of the most popular methods of detecting and quantifying both short and long ncRNAs is qRT-PCR. This method has first the advantage that usually no additional specialized hardware is necessary for the vast numbers of laboratories already performing qRT-PCR for mRNA. There are two general strategies for quantifying ncRNAs, based on TaqMan technology or SYBR green chemistries (Fig. 3.1) [19, 20]. A number of different manufactures including Life Technologies, Exiqon, and WaferGen offer pre-designed assays for both long and short ncRNAs which are discussed below. Many have compiled different primers and probes onto a single well or plate in order to create PCR-based arrays that allow higher throughput. Countless numbers of studies have shown than miRNA can be accurately quantified by qRT-PCR and are consistent with other technologies (like arrays) [21–23].

### *3.2.1 Small ncRNA qRT-PCR*

Real-time PCR assays for small species of ncRNA (such as piRNA and miRNA) are based on novel strategies compared to mRNA. As mentioned, the small target sequences of miRNA, along with close homology within families (e.g., Let 7 family) pose difficulties for primer and probes design [9]. Several manufacturers offer SYBR green detection methods for small RNA species. In general, these methods such as the miScript system from Qiagen and WaferGen system rely first

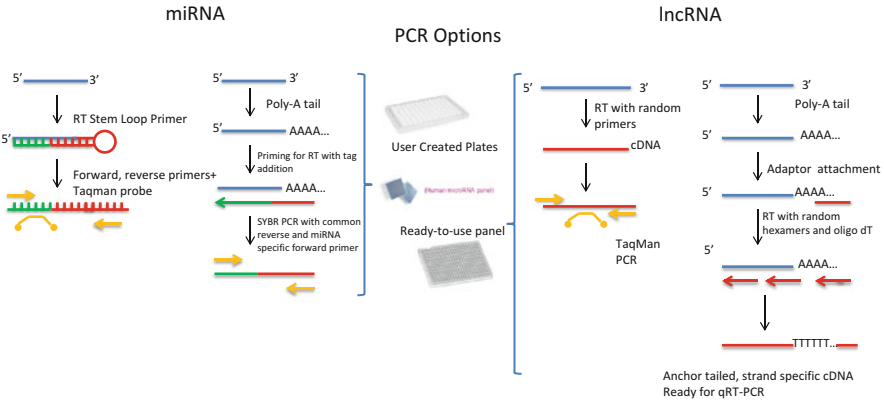


**Fig. 3.1** Illustration of qRT-PCR methods for detecting noncoding RNA. qRT-PCR methods designed for microRNA (*left*) include both TaqMan qRT-PCR and SYBR green assays. TaqMan microRNA assays rely on a miRNA-specific RT with a stem loop primer followed by the usual PCR. SYBR usually include addition of poly-A tail or tag to lengthen the sequence to allow for primer binding. Long noncoding assays (*right*) can also involve both TaqMan or SYBR chemistry, utilizing strand-specific PCR or primers and probes carefully selected to detect only ncRNA

on polyadenylation of small RNA species, followed by a reverse transcription using an oligo-dT primer with a common tag sequence. This tag sequence is then used as a universal reverse primer site and combined with a miRNA specific forward primer. Amplification proceeds with SYBR-green detection [24]. Qiagen maintains specificity for small RNA species using a proprietary Hi-Spec buffer, which inhibits the reverse transcription of longer coding and noncoding RNAs.

Exiqon miRNA PCR is also SYBR green based but utilizes locked nucleic acid technology to increase specificity [25, 26]. As before, the reaction starts with polyadenylation of mature miRNA, followed by reverse transcription using poly-dT primers. Specificity for the microRNA of interest is obtained by using LNA-enhanced forward and reverse primers for the amplification. Locked nucleic acids, because of their ribose modifications, increase the avidity of Watson–Crick binding and specificity of primers allowing for similar primer Tms with shorter sequences facilitating the detection of small ncRNA species. A recent study suggests that for lowly expressed miRNA, LNA technology has improved sensitivity and linearity as compared to TaqMan [27].

Another PCR-based method for miRNA detection is based on TaqMan chemistry [19, 28]. Life Technologies offers predesigned assays which includes stem loop, forward and reverse primers. The stem-loop reverse transcription primer is specific for the small RNA of interest, and serves to elongate the sequence allowing for subsequent PCR priming and amplification. During the amplification, the TaqMan probe is hydrolyzed, releasing the fluorescent dye from inhibition by the quencher and allowing for detection.

As an extension of the individual PCRs, many of above companies also offer array based PCR assays to increase throughput. Qiagen for example, offers several options based on their miScript assays which include prefabricated 96-well, 384-well, and 100-well rotor disc. These plates already contain miRNA specific primers in each well and allow for amplification by loading a common RT product. Currently, Qiagen offers nearly 20 different panels, which include miRNAs which are most abundantly expressed, as well as those that have been implicated in specific pathways or disease processes such as inflammation, cancer, and apoptosis. Similarly, Exiqon offers SYBR-green PCR based arrays which start with a common RT reaction followed by use of multiwell plates preloaded with desiccated LNA enhanced primers for each microRNA. Using two panels, 742 different human microRNAs can be measured in one experiment. Lastly, WaferGen manufactures a Smart Chip microRNA panel which is user-customizable and preloaded with primers for up to 1,200 different miRNAs across 5,000 different SYBR-green PCR reactions in one experiment. Primers design has been optimized for the nanoliter and microliter scale of these PCR reactions and an automated dispensing station is mandatory. Likewise, using TaqMan chemistry, Life Technologies offers a single array with many different human qRT-PCR assays. As mentioned above, TaqMan reverse transcription utilizes a specific stem loop primer for each miRNA, but this system simplifies the reverse transcriptase reaction by using a multiplex of up to 377 different stem loop RT primers. Afterwards, cDNA is loaded onto prefabricated cards

already containing different amplification primers and TaqMan probes for specific miRNAs. The card is run on a standard qRT-PCR machine making it is possible to interrogate expression of over 300 miRNAs in one run.

### *3.2.2 Long ncRNA qRT-PCR*

As eluded to earlier, primer design for qRT-PCR assays for lnc-RNAs also present challenges when compared to mRNA. Many lncRNAs are found in long intergenic regions with high GC content and have repetitive sequences or even circular structures. Additionally, a large percentage of lncRNAs are antisense of known protein coding genes, but present in far less abundance [17, 18, 29]. Life Technologies has overcome many of these limitations by using stringent methodology for lnc-RNA primer design. In sum, long ncRNA sequences are culled from NCBI and other databases and mapped back to the genome. Suitable location for primer designs are identified, which involves avoiding SNPs, repeats, and discrepant areas and focusing on exon–exon junctions. Further refinement of primer design is based on the chemical and thermodynamic properties of primer binding including Tm, GC content, secondary structure, amplicon size, and primer-dimer affinity. In silico quality control is finally done to confirm the highest specificity and sensitivity of the proposed primers. Using this design process, Life technologies reports close to 25,000 available TaqMan primer assays for long ncRNAs. The company has extensively experimentally validated many of these assays and there are a number of publications demonstrating the accuracy and dynamic range of these assays along with consistency as compared to alternative platforms like genomic tiling arrays [30]. To date, this product is available as individual assays only.

There are two manufacturers who have combined multiple individual qRT-PCR reactions onto a single array, allowing for increased throughput. WaferGen for example, using their SmartChip technology, offers a qRT-PCR based array as a single panel of lnc-RNAs containing over 1,700 RNAs in triplicate. After a common RT, the cDNA is loaded robotically into nanoscale capacity wells containing predispensed PCR primers. The platform then performs thousands of quantitative PCR reactions in a single run, essentially combining the screening capability of microarrays with the sensitivity of qRT-PCR. Similarly, Systems Biosciences markets a qRT-PCR array. Sample preparation begins with RNA polyadenylation, and adaptor ligation to the poly-A tail. The reverse transcription reaction is done with both oligo-dT and random hexamer primers which significantly increase cDNA yield and allow for detection of both sense and antisense transcripts. Lastly, a common primer is added which is complementary to the adaptor sequence, and then cDNA is loaded onto a prefabricated plate which contains a forward primer specific for each lncRNA. While customizable to meet user needs, the current stock platform allows for measurement of 90 different lncRNAs simultaneously.

## 3.3 Hybridization Based Techniques

Microarrays have been widely used for many years to profile large numbers of messenger RNA, and recently, the technology has been adapted to the many types of noncoding RNAs [31, 32]. As with mRNA, microarray assays for ncRNA rely on synthesis of cDNA, labeling the product with a fluorophore, followed by dissociation and hybridization to complementary probes immobilized on a surface. Similar to PCR primer design for small or large ncRNA, there are challenges to rigorous microarray probe design. The small size of microRNA limits available probe binding options. Additionally, the ability of microRNA to bind without perfect complementarity and the fact there are similar sequences across family members can lead to nonspecific microarray binding [33]. Long ncRNAs probe design is complicated by high GC content, repetitive or circular structures, and antisense transcription. Despite these challenges, multiple platforms exist and a summary of the most common for both small and long ncRNA is illustrated in Fig. 3.2 and summarized in Table 3.1.

While many of these microarrays share common techniques, some noticeable differences include Exiqon's microRNA arrays which feature Tm-normalized locked nucleic acid-enhanced capture probes, which have improved specificity for microRNAs [34]. Also considered a hybridization based technology, NanoString offers the nCounter Analysis System which can measure both coding and ncRNA species. The system is based on solution-based hybridization of a specific probe followed by immobilization and imaging. Specifically, a target RNA species is combined with a specific reporter probe, containing a color barcode made up by a sequence of six fluorophores. Excess probe is then washed and the complex is immobilized on an nCounter Cartridge via a capture probe. An electrical charge application across the cartridge serves to align the immobolized complexes in a linear fashion so the barcode color probes can be imaged, counted and tabulated for each RNA species [35]. The human v2 miRNA expression assay can measure 800 miRNA simulatenously (based on miRBase 18), while the lnc-RNA product is a custom product allowing the investigator to select up to 800 lncRNAs to multiplex.

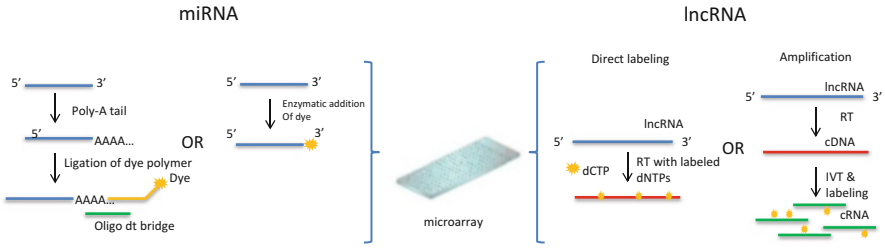


**Fig. 3.2** Illustration of hybridization-based detecting both small and large species of noncoding RNA. Hybridization based methods for assaying microRNA (*left*) are usually sensitive enough to allow for direct miRNA labeling without amplification followed by hybridization to complementary probes on array. Protocols for lnc-RNA can include direct labeling or an IVT-based amplification step

**Table 3.1** Summary of hybridization based methods for assaying both long and small noncoding RNA

| Type of ncRNA | Assay | Manufacturer | RNA profiled | Total RNA needed | Number of probes (human array) | Species available |
|---|---|---|---|---|---|---|
| Long ncRNA | Human Gene 2.0 ST | Affymetrix | lncRNA, mRNA | 50–500 ng | ~40,000 (11,000 lncRNA) | Human, mouse |
| | NCode noncoding RNA arrays | Life technologies | lncRNA, mRNA | 100 ng–5 μg | ~39,000 (22,000 lncRNA) | Human, mouse |
| | Human lncRNA microarray v2.0 | Arraystar (service only) | lncRNA, mRNA | 2 μg | ~63,000 (33,000 lncRNA) | Human, rat, mouse |
| | SurePrint G3 | Agilent | lncRNA, mRNA | 10 ng–1 μg | ~28,000 (7,400 lncRNA) | Human |
| Small ncRNA | NCode V3 | Life technologies | Mature miRNA, snoRNA | 0.25–5 μg | 1,230 | Human |
| | Genechip miRNA array | Affymetrix | Mature miRNA, PremiRNA, scaRNA, snoRNA | 0.13–1 μg | ~25,000 (~5,600 human) | Multispecies on single array |
| | Mercury LNA | Exiqon | Mature miRNA | 0.25–1 μg | ~3,100 (1,900 human) | Multispecies on single array |
| | GenoExplorer | GenoSensor | Mature miRNA, PremiRNA | 1 μg | | Human, rat, mouse, C. elegans, arabidopsis, drosophila |
| | microRNA microarray | Agilent | Mature miRNA, Viral miRNA | 100 ng | 1,309 (1,205 human) | Human, rat, mouse |
| | OneArray | Phalanx | Mature miRNA | 0.5–2.5 μg | 1,884 | Human, rat, mouse, plant |
| | μParaFlo biochip array | LC biosciences | Mature miRNA | 5 μg | 2,042 | All species on miRBase 19 |

## 3.4 Non-probe Based Dectection/Next Generation Sequencing

Classical Sanger sequencing [36], originally described in 1977 has contributed enormously to the understanding of the human genome. Since its discovery, Sanger sequencing has been continuously modified, refined, and automated in order to facilitate the timely completion of the Human Genome Sequencing Project [37, 38]. The technique has been limited largely by its low throughput and ability to only interrogate one predefined target gene at a time. No single modification of sequencing technology has been as critical as the development of "massive parallelization" which has now allowed many millions of genetic fragments to be sequenced at once without a priori sequence or gene information [39]. Massively parallel or next-generation sequencing has found applications in traditional DNA sequencing, gene expression analysis, methylation and of course discovery and quantification of non-coding RNAs. A full review of massively parallel sequencing technologies is beyond the scope of this chapter, but briefly illustrated in Fig. 3.3 are the general categories of library preparation and sequencing technology commercially available today. Most importantly, we discuss data processing techniques in order to use this sequencing data for detection and discovery of ncRNA.

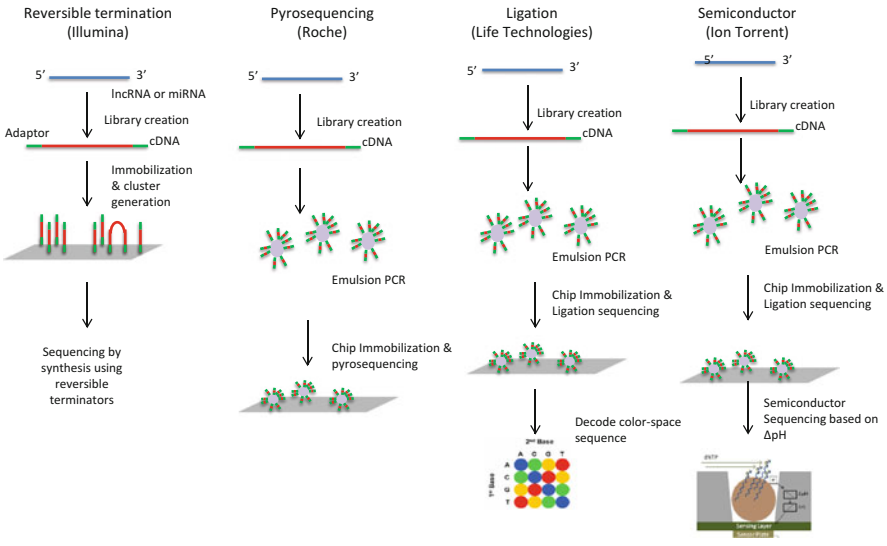


**Fig. 3.3** Illustration of the four most common deep sequencing technologies all starting with library creation, followed by some form of clonal amplification, and lastly a variety of approaches to sequencing

### 3.4.1 Sequencing by Reversible Termination (Illumina)

This method retains a feature of classical Sanger sequencing by utilizing a sequencing by synthesis approach with the sequential addition of fluorescent terminator nucleotides [40]. After fragmentation of the DNA or cDNA of interest, adaptor oligonucleotides are ligated to each fragment end. Fragments are then immobilized on a flow cell where an isothermal PCR based "bridge amplification" occurs. The product of this is hundreds of millions of dense clusters containing identical DNA sequences. Sequencing itself starts with a common primer, followed by the sequential addition of one of four fluorescently labeled, reversibly terminated nucleotides. After each round of synthesis, the clusters are excited with a laser, which emits a color that identifies the newly added base. The fluorescent label and terminator group are then removed, allowing the next cycle of base incorporation and imaging to start. These steps are repeated over several hundred cycles, generating a pattern of color images which is decoded into sequence information for each cluster.

### 3.4.2 Pyrosequencing (Roche)

Similar to the Illumina approach, pyrosequencing is also based on sequencing by synthesis with fluorescently labeled nucleotides [41]. The process starts with fragmentation of the subject DNA followed by adaptor ligation. Next, the fragments are hybridized to the surface of a tiny bead, which are present in far excess to allow each bead to bind only a maximum of one DNA fragment. An amplification occurs on the surface of this bead, in an emulsion created by a mixture of beads, PCR reagents and oil (termed emulsion PCR). Millions of beads, each with their surfaces full of clonally amplified DNA, are then loaded into microscopic wells on small plates only large enough for a single bead to settle inside. Smaller beads, or packing beads, which contain all the needed reagents for the sequencing reaction (except nucleotides) are then added into each well. The sequencing reaction at each well position proceeds by the sequential application of each fluorescently labeled nucleotide in turn to the entire plate. If a specific nucleotide is complementary to the template strand, and therefore incorporated into the growing DNA strand, a phosphate ion is released which leads to an enzymatic cascade and release of light. The intensity and location of the light emission across the plate is recorded and translated into a sequence read and the process is repeated several hundred times.

### 3.4.3 Sequencing by Ligation (Life Technologies)

The last of the "next-generation" technologies involves similar features as discussed above but is termed "sequencing by ligation" instead of "sequencing by synthesis" [42].

As before, fragmentation of the target DNA is necessary followed by a bead emulsion PCR. The millions of beads are then immobilized on a glass slide. The sequencing reaction proceeds by first adding a universal primer, followed by a series of fluorescent oligonucleotide probes, each 8 bases in length with the first two bases being a unique combination and the remaining 6 able to pair with any nucleotide. One of the total 16 possible probes binds specifically to the first 2 bases of the target sequence and is chemically linked via DNA ligase. The cell is washed, and imaged to determine which oligonucleotide was bound at each flow cell site. Subsequently, the terminal 3 bases and dye are cleaved from the bound oligo, and washed. Multiple cycles of ligation provide partial sequence information given there is a gap of three oligonucleotides between each 2 bases that are interrogated (those that were not cleaved from the prior oligo) and the fact that the same color dye is used on multiple different oligos. Therefore, it is necessary to repeat this series of ligation reactions five times, with each time using an initial primer 1 base shorter than the last. In this way, each base on the template strand is interrogated twice (by two separate oligos binding) and the arrangement of colors (known as color space) can be translated into actual sequence data.

### *3.4.4 Semiconductor Sequencing (Life Technologies/Ion Torrent)*

Similar to pyrosequencing, this third-generation sequencing technology uses the sequencing-by-synthesis approach, but instead of detecting a pyrophosphate cascade, it detects hydrogen ions ($H^+$) released as a byproduct of nucleotide incorporation [43]. The release of hydrogen ions produces a change in pH that is proportional to the number of nucleotides incorporated into the nascent DNA strand. This change in pH is detected by a sensor and is converted to voltage that is electronically digitized. A key innovation of this system is that it does not require imaging technology, electromagnetic intermediates (i.e., light or X-rays), or specialized reagents (e.g., chemically modified nucleotides), simplifying the sequencing instruments and reducing the costs. In addition, because the system detects natural polymerase-mediated nucleotide incorporation, the sequencing reaction is measured in real-time (each incorporation measurement takes only 4 s). Another advantage of semiconductor sequencing is its scalability. The semiconductor technology has exponentially improved over the past years leading to an increase in the number of reads per chip and this trend is expected to continue.

### *3.4.5 Single Molecule Sequencing*

Even though this nascent field is at its early stages, there are a number of technologies that already offer the possibility of sequencing single DNA molecules without

the need of reverse transcription or clonal amplification, resulting in reduced preparation time and higher accuracy. In addition, these technologies are expected to produce longer reads, which is ideal for accurate DNA mapping and will facilitate the identification of alternative splicing forms (isoforms) and haplotypes. Most of these new technologies are based on the sequencing-by-synthesis method. For example, Heliscope (Helicos Bioscience) utilizes a method similar to the one described for Solexa/Illumina [44]. After fragmentation and polyadenylation, each strand of DNA is labeled with a fluorescent adenosine and hybridized to a flow cell that contains billions of immobilized oligo dT universal primers. Each of these templates generates its own sequencing reaction. After hybridization, a laser illuminates the surface of the flow cell showing the location of each labeled template and a camera produces a map of the templates. The sequencing reaction proceeds by the addition of a DNA polymerase and fluorescently labeled nucleotides (one at a time) [45].

Another fluorescence-based single molecule sequencing technology is the SMRT™ chemistry from Pacific Biosciences. This technology offers two key innovations: First, the dyes are attached to the terminal phosphate of the nucleotide instead of to the base itself (which is the standard of other sequencing technologies). This allows the polymerase to cleave away the fluorescent label as part of the incorporation process leaving behind a natural strand of DNA. The second innovation consists on an optical waveguide called "the zero-mode waveguide," which enables the observation of individual molecules against a background of labeled nucleotides while maintaining a high signal to noise ratio [46]. This technology is therefore, single molecule and real-time.

## 3.5 Computational Methods for Analysis

After sequencing of the RNA transcriptome, the next critical challenge is to manage the terabytes of data produced which requires powerful computing and storage resources along with efficient algorithms. In the next two sections, we will outline the general techniques and resources available for analysis of next-generation sequencing data, with particular attention first to miRNA and then to long ncRNA.

### *3.5.1 Small (miRNA) Analysis*

Many different types of analyses can be performed to gain insight into the roles small RNAs play in processes such as development and cancer. The raw data obtained from a sequencer consists of millions of "reads" with each read corresponding to a single small RNA molecule that was present in the sample. Before any meaningful information can be gained, the reads need to be preprocessed and aligned to a reference genome. Preprocessing generally includes trimming the adapter off of the 3′ end of the read, filtering out reads that are too short (e.g., less

than 15 nt), and filtering out reads with low quality or missing bases. A critical step is the identification of the location from the genome where the read was originally transcribed and this process is referred to as a sequence alignment. Many algorithms and tools have previously been developed to align a set of sequences to a database of reference genomes such as the Basic Local Alignment Search Tool (BLAST) and the BLAST-Like Alignment Tool (BLAT) [47, 48]. Although these tools have a high degree of accuracy, they are not capable of efficiently aligning millions to billions of reads within a reasonable period of time. To this end, several algorithms and tools have been developed to process and align the reads generated from DNA, large RNA, or small RNA sequencing data. Aligners that can handle small RNA sequencing data include Bowtie, BWA, PerM, and SHRiMP (Table 3.2).

Once the reads are aligned to the genome, several types of analyses can be performed including quantification of expression of known small RNAs, detection of novel small RNAs, and characterization of small RNA isoforms. Over the past several years, hundreds of novel microRNAs have been discovered in many species using small RNA sequencing. Algorithms designed to predict novel microRNAs rely on the overall number of reads that align to a specific locus as well as several specific features of microRNA biogenesis. MicroRNAs during their biogenesis will form a stem loop structure that is thermodynamically stable (Fig. 3.4a) after the primary transcript has been cleaved by Drosha. Once a microRNA stem loop is exported from the nucleus, it will be cleaved again by Dicer into three pieces. The two "arms" of the stem loop, one from the 5′ side and the other from the 3′ side, have the potential to become mature microRNAs and inhibit mRNA expression after being incorporated into the Ago complex. Often, one arm will be preferentially incorporated into the Ago complex while the other arm will be degraded. In these cases, the expression levels of each arm will be different from each other within a cell. The loop that held the two arms together will also be degraded and only found in low levels within a cell. An example of an algorithm which uses these features is MiRDeep, which was one of the first algorithms to predict novel microRNAs from small RNA sequencing data. MiRDeep will score each potential novel microRNA locus based on the total number of reads that align to that region, the ability of a potential RNA sequence from that region to form a thermodynamically stable stem loop structure, and the distribution of reads within that region (Fig. 3.4b) (i.e., that many reads align to the arms of the stem while a few reads align to the loop). Algorithms to predict novel microRNAs are constantly being developed or improved.

Sequencing of the microRNA transcriptome has also revealed the presence of microRNA isoforms. Isoforms can be observed by examining any heterogeneity in reads that align to a particular mature microRNA locus. Each mature microRNA has a canonical sequence which can be found in the miRBase database [49]. Many reads will match the canonical microRNA sequence and therefore align perfectly to the canonical microRNA locus. However, some reads will have extra nucleotides at the beginning (5′ variability) and/or at the end (3′ variability) of the read. These reads will have a slightly different length which will result in different starting and/or stopping locations in the genome. Other reads may match the canonical microRNA in length, but have one or more mismatches to the reference sequence. This

**Table 3.2** Algorithms for detection and quantification of small noncoding RNAs

| Algorithm | Reference | Description | Input data | Novel detection | Expression quantification |
|---|---|---|---|---|---|
| miRDeep | Friedländer MR, Nat Biotechnol. 2008 [50] | Uses a probabilistic model of miRNA biogenesis to score compatibility of the position and frequency of sequenced RNA with the secondary structure of the miRNA precursor | Reference genome + Small RNA-seq Data | x | x |
| miRAnalyzer | Hackenberg, Nucl. Acids Res. (2009) [51] | A web server tool that identifies known small RNAs and uses a random forest machine learning algorithm to predicted novel microRNAs | Small RNA-seq Data | x | x |
| Deep-sequencing small RNA analysis pipeline (DSAP) | Huang, Nucleic Acids Res. 2010 [52] | multiple-task web service for identifying and quantifying known small RNAs | Small RNA-seq Data | | x |
| MIReNA | Mathelier A, Bioinformatics. 2010 [53] | Detects new miRNAs by homology from known miRNAs or from deep sequencing data by utilizing characteristics of pre-miRNAs | Sets of known microRNA and/or small RNA-seq data | x | |

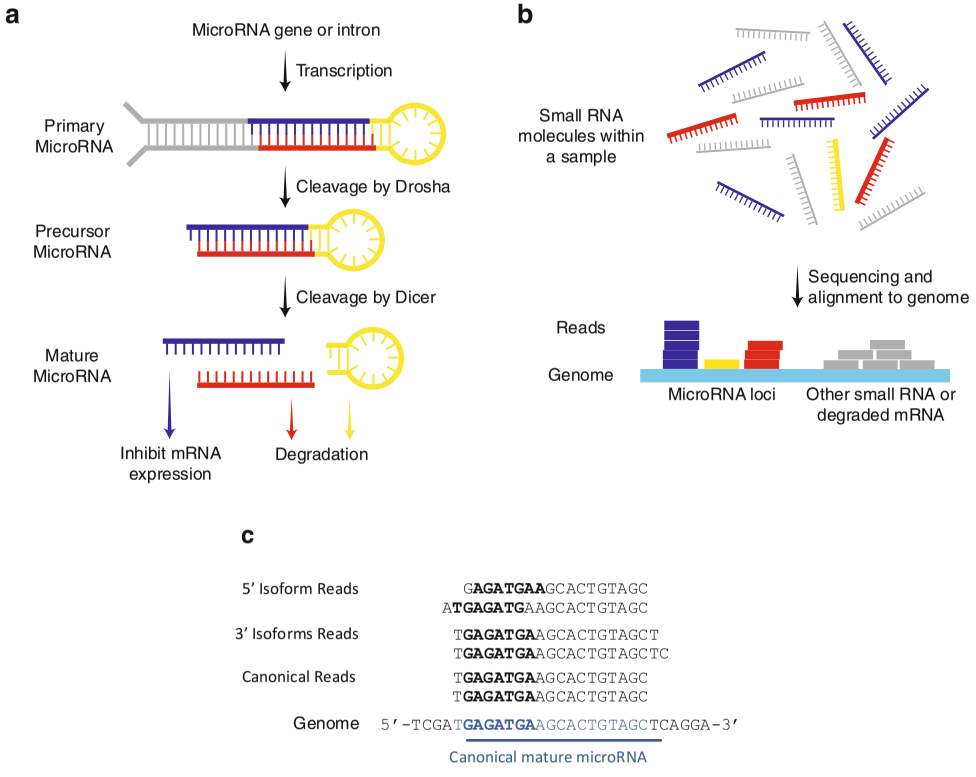


**Fig. 3.4** Measuring known and detecting novel miRNA with sequencing technology. Several features specific to microRNA biogenesis (**a**) such as formation of a stem loop structure and subsequent cleavage of two "arms" aid in sequencing analysis. One such algorithm, miRDeep (**b**) predicts novel miRNA by identifying patterns during alignment of these cleavage products including the two "arms," separated by a lowly expressed intervening region containing a sequence able to form a thermodynamically stable stem loop structure. (**c**) Sequencing of the microRNA transcriptome has also revealed the presence of microRNA isoforms known as "isomirs"

phenomenon is known as RNA editing and usually occurs at the last nucleotide of the microRNA. Detection of isomirs as described can be done by varying the number of mismatches tolerated during sequence alignment (Fig. 3.4c). While the role of noncanonical microRNA isoforms in development and disease is not well understood, some initial studies have suggested that these isoforms do have some differences in gene targets or in targeting efficiency compared to the canonical form of the microRNA. The existence of microRNA isoforms adds to the complexity of the microRNA transcriptome and provides a future avenue of enticing research for which small RNA sequencing is well suited.

These analyses highlight the benefits of using RNA sequencing technology compared to high-throughput probe-based technologies. While many microRNA microarrays have been developed and can measure the expression of known microRNAs within a sample, they do not have the ability to discover novel microRNAs or characterize microRNA isoforms since they rely on predetermined probes. Identifying novel microRNAs or microRNA isoforms through the use of small RNA sequencing allows for a more comprehensive characterization of the microRNA transcriptome.

### *3.5.2 Large ncRNA Analysis*

Similar to small RNA sequencing, the first step in analysis to detect ncRNAs starts with aligning each read to a reference genome in order to identify the genomic location from which the transcript was derived. Tools that work with large RNA sequencing data include MAQ, TopHat (which utilizes the Bowtie aligner), and BWA. These read aligners for large RNA sequencing data must handle complexity arising from mismatches in the sequence due to sequencing errors, single nucleotide polymorphisms (SNPs), mutations, and alternative splicing of coding and noncoding RNAs. In cases of alternative splicing, reads may overlap a splice junction resulting in each end of the read aligning to a different place in the reference genome. If paired-end sequencing was performed, then each read in the pair may align to different regions in the reference genome. Once the reads are properly aligned, other analyses may be performed including prediction of novel noncoding RNAs or identifying differentially expressed transcripts between various biological states.

Sequencing of the large RNA can reveal additional complexity in both the coding and noncoding transcriptomes including novel transcripts. Several long intergenic noncoding RNAs (lincRNAs) have been identified and shown to have important roles in development [30, 54]. However, the complete set of lincRNAs has not been fully characterized yet. Before the advent of large RNA sequencing technology, previous algorithms including QRNA, Moses, Dynalign, and RNAz predicted novel lincRNA loci using features such as sequence homology, substation rates, and predicted secondary structure or thermodynamic stability of potential RNA transcripts (Table 3.3). While these algorithms provided an initial picture of the lincRNA landscape, newer algorithms such as Cufflinks and Scripture are able to utilize RNA sequencing data to empirically identify novel transcripts within a given sample.

**Table 3.3** Algorithms for detection of novel and quantification of large noncoding RNA

| Algorithm | Reference | Description | Input data | Detection of novel | Expression quantification |
|---|---|---|---|---|---|
| Moses (*mo*dular *se*quence *s*uite) | Raasch BMC Bioinformatics 2010 [55] | Combines predictions from several algorithms including RNAz and Dynalign | Reference sequences | x | |
| QRNA | Rivas BMC Bioinformatics 2001 [56] | Tests the pattern of substitutions observed in a pairwise alignment of two homologous sequences | Pairwise sequence alignment between two genomes | x | |
| RNAz | Washietl, Curr Protoc Bioinformatics, 2007 [57] | Predicts such functional RNA structures on the basis of thermodynamic stability and evolutionary conservation of homologous sequences | Query genome and homologous genomes | x | |
| Dynalign | Uzilov, BMC Bioinformatics, 2006 [58] | Predicts secondary structures common to two input RNA sequences | Query sequences and known | x | |
| Scripture | Guttman, Nat. Biotechnol 2010 [59] | Reconstructs the transcriptome using RNA-sequencing reads and a reference genome | Reference genome + RNA sequencing data | x | x |
| Cufflinks | Roberts, Bioinformatics, 2011 [60] | Leverages RNA-sequencing reads and existing annotations to assemble the transcriptome | Reference genome + existing annotations + RNA sequencing data | x | x |

They rely on graph theory to reconstruct each transcript from the aligned reads, including the various isoforms produced from alternative splicing events. These reconstructed transcripts can then be compared to a database of known transcripts to identify which ones are known or novel. Novel transcripts of sufficient length that do not have coding potential can be considered putative lincRNAs [12]. As more novel lincRNAs are being discovered and characterized, their roles in regulation of the cell will become more apparent and we will gain greater insights into many biological processes such as development and cancer.

## 3.6 Conclusion

As described in this chapter, miRNA and larger ncRNA detection and discovery present unique challenges when compared to conventional mRNA. A number of different and novel platforms have been developed to address these limitations and include optimization of qRT-PCR, microarray, and next-generation sequencing. The latter requires extensive computational expertise but opens the opportunity for discovery of new RNA species.

## References

1. Jin L, Lloyd RV. In situ hybridization: methods and applications. J Clin Lab Anal. 1997;11:2–9.
2. Kevil CG, Walsh L, Laroux FS, Kalogeris T, Grisham MB, Alexander JS. An improved, rapid Northern protocol. Biochem Biophys Res Commun. 1997;238:277–9.
3. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques. 2008;44:619–26.
4. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270:467–70.
5. ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. J Mol Diagn. 2008;10:484–92.
6. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39:D152–7.
7. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. Nat Rev Genet. 2012;13:358–69.
8. Hammond SM. microRNA detection comes of age. Nat Methods. 2006;3:12–3.
9. Wark AW, Lee HJ, Corn RM. Multiplexed detection methods for profiling microRNA expression in biological samples. Angew Chem Int Ed Engl. 2008;47:644–52.
10. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008;18:610–21.
11. Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. 2011;39:D146–51.
12. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25:1915–27.

13. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, et al. Ensembl 2011. Nucleic Acids Res. 2011;39:D800–6.
14. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 2011;39:D876–82.
15. Pang KC, Stephen S, Dinger ME, Engstrom PG, Lenhard B, Mattick JS. RNAdb 2.0–an expanded database of mammalian non-coding RNAs. Nucleic Acids Res. 2007;35:D178–82.
16. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35:D61–5.
17. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nat Rev Genet. 2007;8:413–23.
18. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447:799–816.
19. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5′—3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A. 1991;88:7276–80.
20. Zipper H, Brunner H, Bernhagen J, Vitzthum F. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. Nucleic Acids Res. 2004;32:e103.
21. Sand M, Skrygan M, Sand D, Georgas D, Gambichler T, Hahn SA, Altmeyer P, Bechara FG. Comparative microarray analysis of microRNA expression profiles in primary cutaneous malignant melanoma, cutaneous malignant melanoma metastases, and benign melanocytic nevi. Cell Tissue Res. 2013;351:85–98.
22. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, Lu J, Liu G, Bowers J, Vaziri C, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A. 2009;106:2319–24.
23. Ho CY, Bar E, Giannini C, Marchionni L, Karajannis MA, Zagzag D, Gutmann DH, Eberhart CG, Rodriguez FJ. MicroRNA profiling in pediatric pilocytic astrocytoma reveals biologically relevant targets, including PBX3, NFIB, and METAP2. Neuro Oncol. 2013;15:69–82.
24. Shi R, Chiang VL. Facile means for quantifying microRNA expression by real-time PCR. Biotechniques. 2005;39:519–25.
25. Kaur H, Arora A, Wengel J, Maiti S. Thermodynamic, counterion and hydration effects for the incorporation of locked nucleic acid (LNA) nucleotides in duplex. Nucleic Acids Symp Ser (Oxf). 2008;52:425–426.
26. Owczarzy R, You Y, Groth CL, Tataurov AV. Stability and mismatch discrimination of locked nucleic acid-DNA duplexes. Biochemistry. 2011;50:9352–67.
27. Jensen SG, Lamy P, Rasmussen MH, Ostenfeld MS, Dyrskjot L, Orntoft TF, Andersen CL. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. BMC Genomics. 2011;12:435.
28. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 2005;33:e179.
29. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, Dinger ME, Mattick JS. Genome-wide analysis of long noncoding RNA stability. Genome Res. 2012;22:885–98.
30. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129:1311–23.
31. Yin JQ, Zhao RC, Morris KV. Profiling microRNA expression with microarrays. Trends Biotechnol. 2008;26:70–6.

32. Li W, Ruan K. MicroRNA detection by microarray. Anal Bioanal Chem. 2009;394:1117–24.
33. Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, Bertone P, Caldas C. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA. 2010;16:991–1006.
34. Castoldi M, Schmidt S, Benes V, Noerholm M, Kulozik AE, Hentze MW, Muckenthaler MU. A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA. 2006;12:913–20.
35. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008;26:317–25.
36. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74:5463–7.
37. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. Nature. 1986;321:674–9.
38. Luckey JA, Drossman H, Kostichka AJ, Mead DA, D'Cunha J, Norris TB, Smith LM. High speed DNA sequencing by capillary electrophoresis. Nucleic Acids Res. 1990;18:4417–21.
39. Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. Hugo J. 2011;5:1–12.
40. Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA, Erlich HA. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens. 2009;74:393–403.
41. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376–80.
42. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005;309:1728–32.
43. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348–52.
44. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, et al. Single-molecule DNA sequencing of a viral genome. Science. 2008;320:106–9.
45. Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, Jarosz M, Krzymanska-Olejnik E, Kung L, Lipson D, et al. Virtual terminator nucleotides for next-generation DNA sequencing. Nat Methods. 2009;6:593–5.
46. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.
47. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
48. Kent WJ. BLAT–the BLAST-like alignment tool. Genome Res. 2002;12:656–64.
49. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34:D140–4.
50. Friedlander MR, Chen W, Adamidi C, Masskola J, Einspanier R, Knespel S, et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol. 2008;4:407–15.
51. Hackenberg M, Sturm M, Langenberger D, Falcon-Perez J, Arasay A. miRanalzyer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucl Acids Res. 2009;37:68–76.
52. Huang P, Lui Y, Lee C, Lin W, Gan R, Lyu P, et al. DSAP: deep-sequencing small RNA analysis pipeline. Nucl Acids Res. 2010;38:385–91.
53. Mathelier A, Carbone A. MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. Bioinformatics. 2010;18:2226–34.

54. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322:750–6.
55. Raasch P, Schmitz U, Patenge N, Vera J, Kreikemeyer B, Wolkenhauer O. Non-coding RNA detection methods combined to improve usability, reproduction and precision. BMC Bioinforma. 2010;11:1–12.
56. Rivas E, Eddy S. Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinforma. 2001;2:1–19.
57. Washietl S, Hofacker I. Identifying structural noncoding RNA using RNAz. Curr Protoc Bioinformatics. 2007;10:1–12.
58. Uzilov A, Keegan J, Matthews D. Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. BMC Bioinforma. 2006;7:1–30.
59. Guttman M, Garber M, Levin J, Donaghey J, Robinson J, Adivonis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2008;28:503–10.
60. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics. 2011;27:2325–9.

# Chapter 4
# MicroRNAs in Solid Tumors

**Jin Wang and Subrata Sen**

**Abstract** MicroRNAs (miRNAs) are small noncoding RNAs (17–25 nt) that modulate the expression of genes by negatively regulating translation and stability of messenger RNAs (mRNAs). miRNAs have emerged as critical players in cancer initiation and progression processes due to their documented roles in either promoting (oncogenic) or suppressing (tumor-suppressors) oncogenesis. Dysregulation of miRNA expression levels and associated alterations in gene expressions have been demonstrated in various solid tumors. In this chapter, biogenesis of miRNAs and dysregulation of miRNAs in tumor formation, maintenance and metastasis as well as potential role of miRNAs as cancer biomarkers and therapeutic targets for anticancer treatment are discussed.

**Keywords** miRNAs • Solid tumors • Oncomirs • Biomarkers • Diagnosis and prognosis • Cancer therapy

## 4.1 Introduction

MicroRNAs (miRNAs) are a class of small noncoding RNAs of 17–25 nt in length, which are conserved across species. Since their discovery in *Caenorhabditis elegans* at the beginning of the 1990s [1], miRNAs have been found to be expressed in

J. Wang
Department of Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

S. Sen (✉)
Program in Human and Molecular Genetics, The University of Texas Graduate School of Biomedical Sciences, Houston, TX, USA

Departmental of Translational Molecular Pathology, Unit 951, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA
e-mail: ssen@mdanderson.org

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_4,

different tissues and cell types. It is predicted that miRNAs regulate expression of at least half of the human transcriptome by either repressing the translation or causing degradation of multiple target mRNAs [2, 3]. In the mammalian genome, miRNA species accounts for 1–3 % of genes and over 1,900 miRNAs have been reported, with these having critical regulatory functions in cell development, proliferation, differentiation, apoptosis, and stress response, and they are involved in the genetic networks regulating functional pathways dysregulated in cancer [4, 5]. miRNA genes are generated from the stem loop precursors by RNA polymerase II, forming large molecules containing a stem-loop structure and called primary precursor miRNAs (pri-miRNAs). pri-miRNAs initially contain a cap structure at the 5′ end and a 3′ poly-A tail. In the nucleus, the stem-loop is asymmetrically cleaved by a complex comprising the RNase III Drosha and its cofactor Dgcr8 to produce the precursor miRNA (pre-miRNA), approximately 70 nt in length. The pre-miRNA is transported to the cytoplasm, a process mediated by the nuclear transport receptor exportin-5 and the nuclear protein Ran-GTP (Exportin-5-Ran complex), and the exported pre-miRNA is cleaved by another RNase III endonuclease, Dicer, to produce a short duplex molecule (20 bp mi-RNA:mi-RNA duplexes). One of the strands of the duplex is selected in association with specific Argonaute proteins to form miRNA-induced silencing complex (*miRISC*) on the mRNA that negatively regulates its expression by either translational repression or by inducing mRNA cleavage. The silencing mechanism is possibly determined by the extent of base pairing between miRNA and target mRNA when mRNA binds to the complementary target sites located in the 3′ untranslated region (3′UTR) of target mRNA. Argonaute proteins have also been implicated in the transcriptional and posttranscriptional gene silencing of the targeted mRNA [5–7]. Aberrant expression of miRNAs affects the regulation of many cellular functions and gene networks [4] and is associated with a wide spectrum of diseases, including diabetes [8], cardiovascular disease [9], viral infection [10], kidney disease [11], neurodegenerative diseases [12], and cancer [13].

## 4.2 Discovery of miRNA Involvement in Cancer

The involvement of miRNA dysregulation in human cancer was first reported by Dr. Carlo Croce's group in 2002. They discovered downregulation of *miR-15a* and *miR-16* in about 70 % of chronic lymphocyte leukemia [13] suggesting these miRNAs have tumor suppressor function in normal cells. *miR-15* and *miR-16*, were later reported to target pro-survival proteins, *Bcl-2* and *Mcl-1* encoding transcripts, and thus regulate induction of apoptosis [14–16]. These two miRNAs were also shown to effectively inhibit growth of colon tumor xenografts [17]. Since the first discovery of miRNA dysregulation playing a role in chronic lymphocyte leukemia, aberrant expression profiles of many miRNAs have been implicated in various solid tumors, including those of the breast, colon, liver, lung, prostate, thyroid, and pancreas. Early large-scale miRNA microarray studies revealed that the samples from colon, liver, pancreas, and stomach all clustered together in unsupervised

hierarchical cluster analysis. The gut-associated miRNA cluster (*miR-192*, *miR-194* and *miR-215*) was found to have functional roles in specification of gut development and upregulated in gut-derived tumors [18]. *miR-17-5p*, *miR-20a*, *miR-21*, *miR-92*, *miR-106a*, and *miR-155*, on the other hand, represent miRNA signatures associated with breast, colon, lung, pancreas, prostate and stomach cancers suggesting their plausible involvement in these malignancies [19].

High throughput miRNA microarray, real-time PCR array, and the next-generation sequencing technologies have been performed to profile miRNAs of solid tumors in order to generate tumor tissue specific miRNA signatures. Microarray profiling of miRNAs have been used to identify the tissue of origin for unknown primary or poorly differentiated tumors and also to distinguish different tumor subtypes. Combination analysis using Illumina/Solexa deep sequencing and miRNA microarray have shown not only high sequence counts and concordant differential expression of miRNAs in prostate cancer [20] and breast cancer [21] but also miRNA expression signatures that may serve as accurate diagnostic and prognostic biomarkers for these malignancies. Importantly, large tissue specimens are not needed for accurate miRNA detection since their expression can be reliably measured in biopsy specimens, as shown by the miRNA profiling assays performed with formalin-fixed, paraffin-embedded (FFPE) tissues of hepatocellular carcinoma (HCC) [22], lung cancer [23], melanoma [24], renal tumor [25], and papillary thyroid carcinoma [26]. FFPE specimens also can be used reliably for miRNA deep-sequencing analysis, reported in renal cell carcinoma [27], and for locked nucleic acid (LNA)-based miRNA qPCR analysis successfully performed with multiple types of tumor tissues [28].

## 4.3 miRNA Functions in Solid Tumors

Aberrant expression of miRNAs is commonly detected in various solid tumors. Dysregulation of these cancer associated miRNAs has been implicated to play roles in tumor initiation and progression by promoting cell proliferation and invasion while enhancing survival by inactivating apoptosis response mechanisms. Upregulation and repression of several miRNAs have also been implicated in de-differentiation of cells facilitating the tumorigenic transformation process. Oncogenic and tumor-suppressing miRNAs, collectively termed “oncomirs,” act either as tumor suppressors [29, 30] or as oncogenes [31–34] depending on the cellular context and the genes targeted in each instance (Table 4.1). Furthermore, specific metastasis-regulating miRNAs collectively termed “metastamirs” have been found to be either positively or negatively associated with metastasis [35].

The miRNAs that are encoded by the *let-7* family were the first group of oncomirs shown to regulate the expression of the *Ras* gene [34]. The repression of *Let-7* targeted *Ras* [36] is associated with shortened postoperative survival in patients with lung tumors [37]. A variant allele of *KRAS* at *rs61764370* (referred to as the *KRAS* variant), which disrupts a *let-7*-binding site in *KRAS*, is a genetic marker for

**Table 4.1** MicroRNAs as oncomirs in solid tumors[a]

| miRNA | Tumor type | Role | Reference |
|---|---|---|---|
| Let-7 | Breast, cervical, colorectal, ovarian, urothelial | Tumor suppressor | [34, 36–40] |
| miR-9 | Brain, liver | Oncogene | [55, 60] |
| | Breast, colorectal, gastric, ovarian, renal | Tumor suppressor | [61–68] |
| | Breast | Metastasis-promoting | [57] |
| miR-10b | Breast | Metastasis-promoting | [31] |
| miR-15/16 | Colon, prostate | Tumor suppressor | [17, 34] |
| miR-17-92 cluster | Breast, colon, lung, pancreas, prostate | Oncogene | [19, 34] |
| | Liver | Tumor suppressor | [34] |
| miR-21 | Breast, colon, lung, prostate, stomach | Oncogene | [5, 19, 42, 43] |
| | Liver, pancreas | | [138] |
| | Breast, colon, lung | Metastasis-promoting | [44, 69] |
| miR-30a | Lung | Metastasis-suppressing | [54] |
| miR-30d | Lung, liver, ovarian | Oncogene | [49–51] |
| | Thyroid | Tumor suppressor | [52] |
| miR-31 | Breast, gastric, prostate, ovarian | Tumor suppressor | [76–81] |
| | Breast, urothelial | Metastasis-suppressing | [76, 82] |
| miR-143/miR-145 | Breast, colon, cervical, prostate | Tumor suppressor | [34] |
| miR-146 | Breast, prostate | Metastasis-suppressing | [35, 84, 86] |
| miR-155 | Breast, colon, lung, pancreas | Oncogene | [19, 34, 138] |
| miR-182 | Breast, endometrial, ovarian | Oncogene | [71, 74, 75] |
| | Breast, liver | Metastasis-promoting, EMT | [72, 98] |
| miR-200 family | Breast, ovarian | Metastasis-promoting, EMT | [91, 92] |
| miR-205 | Breast, nasopharyngeal, ovarian | Metastasis-promoting, EMT | [91, 94] |
| miR-221/miR-222 | Breast | EMT | [95–97] |
| miR-335 | Breast | Metastasis-suppressing | [87] |

[a]This is a partial list of oncomirs characterized in solid tumors and discussed in the chapter

increased risk of developing ovarian cancer [38], breast cancer [39], and colorectal cancer [40]. *miR-21* promotes invasion and migration as an oncomir, which drives tumorigenesis through inhibition of negative regulators of the *Ras/MEK/ERK* pathway and inhibition of apoptosis [41]. High *miR-21* levels have been found in a variety of solid tumors (lung, prostate, breast, colon, stomach, liver) and hematological malignancies [5, 42, 43]. Knockdown of *miR-21* significantly decreases experimental metastasis to lungs in breast cancer, colorectal carcinoma, and B16 melanoma cells [44, 45]. Elevated expression of a polycistronic *miR-17-92* cluster, located on the chromosome 13 open reading frame 25 (*C13orf25*) in the human genome has been validated to display both tumor suppressor and oncogenic potential in many solid tumors including glioma, non-small-cell lung cancer (NSCLC),

bladder cancer, squamous cell carcinoma of the head and neck, peripheral nerve sheath tumor, malignant fibrous histiocytoma, alveolar rhabdomyosarcoma, liposarcoma, osteosarcoma, and colon carcinoma [19, 46–48]. The ability of the *miR-17-92* cluster to function either as tumor suppressors or oncogenes depends on which cell types these miRNAs are expressed in and what tissue-specific target mRNAs are regulated [34].

Tumor-suppressor miRNAs are repressed in primary tumors [29, 30]. These miRNAs are expressed at relatively higher levels in differentiated tissues and at reduced levels in cancer. Such reduction in miRNA expression presents an interesting parallel between cancer and development. As in cancer, early developmental stages have reduced levels of many miRNAs that are abundant in differentiated tissue. Lower expression of these miRNAs possibly reflects loss of differentiation, which is a hallmark of cancer.

There are, however, examples of miRNAs, which can not be classified either as a tumor-suppressor miRNA or as an oncomir. For examples, *miR-30d* is upregulated in squamous lung carcinomas [49] and HCC and targets *GNAI2* [50]. The chromosomal locus harboring *miR-30d* is amplified in more than 30 % of human solid tumors, and higher levels of *miR-30d* expression has been associated with poor clinical outcomes in ovarian cancer patients [51]. Interestingly, on the other hand, repression of miR-30d is known to contribute to the process of thyroid cancer progression, leading to the development of anaplastic carcinomas with accompanying upregulation of the target protein *EZH2*, a component of the polycomb repressive complex [52]. Similarly, *miR-29* and *miR-30* upregulation was shown to induce *Rb* driven senescence through downregulation of *B-Myb* in cervical carcinoma cells [53]. Furthermore, *miR-30a*, repressed in NSCLC, was shown to inhibit invasion, EMT, and metastasis by targeting *Snai1* [54]. Another example of context dependent conflicting functional association of miRNA in different tumor types has been reported for *miR-9*. Overexpression of *miR-9* was detected in brain tumors [55], metastatic breast cancer [56–58] and in *c-myc*-induced mouse mammary tumors [59], while knockdown of *miR-9* reportedly decreased invasion of HCC cells [60] implying *miR-9* overexpression to be positively associated with malignancy of human cancers. However, *miR-9* downregulation was observed in gastric cancer [61, 62], ovarian cancer [63], colorectal cancer [64], renal cell carcinoma [65], and breast cancer [66–68]. Additionally, depletion of *miR-9* recapitulated the antiproliferative and pro-apoptotic effect of *miR-9* overexpression in MCF-7 breast cancer cells by targeting *MTHFD2* gene transcript [68], all of which indicate a tumor-suppressor role for this miRNA. Taken together, the opposing tumor-suppressing and oncogenic natures of the *miR-17-92* cluster as well as *miR-30* and *miR-9*, emphasize the complexities and intricacies of miRNA-mediated gene regulation in cancer progression.

Metastamirs have typically been discovered using in vitro screens for steps in the metastatic cascade including cell growth, EMT, adhesion, migration, invasion, apoptosis, and/or angiogenesis. The metastamirs have been shown to promote or inhibit metastasis in experimental models, acting as metastasis-promoting miRs and metastasis-suppressing miRs, affecting critical steps in the metastatic cascade [35].

Metastasis-promoting miRs include *miR-9*, *miR-10b*, *miR-21*, and *miR-182*. *miR-9* is considered a pro-metastatic miRNA since it negatively regulates the key metastasis suppressor E-cadherin, and *miR-9* levels were reported to be significantly elevated in breast tumor patients with metastatic disease compared with those with no metastasis [57]. *miR-10b*, highly expressed in metastatic breast cancer cells and metastatic breast tumors, positively regulates cell migration, tumor cell invasion, and metastasis by targeting homeobox D10 (*HOXD10*), leading to increase in the pro-metastatic gene *RHOC* [31]. As discussed above, while *miR-21* has earlier been shown to promote oncogenesis by potentiating oncogenic Ras signaling and attenuating apoptosis by repressing expression of multiple tumor suppressor genes [41], more recent data suggest that elevated *miR-21* also promotes growth and metastasis of lung cancer cells by targeting the expression of the tumor suppressor protein *PTEN* [69]. Overexpression of *miR-182* has been correlated with progression from primary to metastatic melanoma and inversely correlated with *FOXO3* and microphthalmia-associated transcription factor levels [70]. Oncogenic function of another metastamir, *miR-182* in ovarian cancer is reportedly mediated in part through its involvement in impairing DNA double strand break repair mechanism by negatively regulating *BRCA1* tumor suppressor and metastasis suppressor 1 as well as positive regulation of the oncogene *HMGA2* [71]. Published studies have suggested that the use of anti-miR-182 is a promising therapeutic strategy for metastatic melanoma and liver metastases [72]. *miR-182*, *miR-183*, and *miR-96*, highly conserved across species with similar sequences, are involved in zinc import and export and help regulate zinc homeostasis by targeting *hZIP1* [73]. Both *miR-96* and *miR-182* are reported to target *FOXO1* and *FOXO3* in breast and melanoma cancer cell lines and endometrial tumors [70, 74, 75].

The metastasis-suppressing miRs include *miR-31*, *miR-146*, *miR-206*, and *miR-335*. *miR-31* expression is repressed in human carcinomas of the breast [76, 77], prostate [78], ovary [79], and stomach [80, 81]. The homozygous loss of the miR-31-encoding genomic locus has been described in human urothelial carcinomas [82] and acute lymphoblastic leukemias [83]. The expression of *miR-31* correlated inversely with breast cancer progression in humans with inhibition of *miR-31* promoting metastasis in vivo [76], possibly as a consequence of targeting the expression of an actin remodeling protein, *WAVE3* [77] affecting the invasion-metastasis cascade. *miR-146a* and *miR-146b*, two distinct genes encoded on chromosomes *5q33* and *10q24* with similar targets were shown to inhibit both migration and invasion [35, 84]. Expression of *miR-146* attenuating metastatic potential of breast cancer cells and metastasis in breast cancer [35] suppresses *NF-κB* and signaling intermediates downstream of innate immune receptors [84, 85]. Silencing of another *miR-146* target *ROCK1* kinase contributing to metastatic and invasive activity of transformed cells when dysregulated along with its activating Rho kinase, has also been reported to diminish metastatic potential of prostate tumors [86]. *miR-335*, *miR-206* and *miR-126*, downregulated across a number of metastatic tumor cell lines, can suppress metastasis of breast cancer cells following restoration of their expression and *miR-335* was shown to suppress metastasis and migration of breast cancer cells by targeting the progenitor cell transcription factor *SOX4* and

extracellular matrix component tenascin C [87]. Intriguingly, however, another study reported that although the expression levels of *miR-335* was not found significantly different in normal adjacent colon tissue compared with primary tumors that did not metastasize, *miR-335* levels were significantly elevated in colorectal cancer primary tumors with metastatic capacity, retaining similar levels in their metastatic lesions [88].

Epithelial-mesenchymal transition (EMT) has key roles in metastatic disease and therapeutic resistance. Epithelial cancer cells that acquire migratory and invasive mesenchymal cell-like properties during EMT gain the ability to emigrate from the primary tumor mass and move to distant locations. Once metastasized, these cells undergo mesenchymal-epithelial transition (MET) to facilitate the subsequent settlement and proliferation of disseminated cancer cells at secondary locations [89, 90]. The *miR-200* family (*miR-200a*, *miR-200b*, *miR-200c*, *miR-141*, and *miR-429*) and *miR-205* are markedly downregulated in cells that undergo EMT in response to *TGF*-β or to ectopic expression of the protein tyrosine phosphatase *Pez* in MDCK cells, and repression of these miRNAs is an essential early step in tumor metastasis [91]. Ectopic overexpression of *miR-429* was shown to result in reversal of mesenchymal phenotype and significant reduction in migration and anchorage independent growth of ovarian cancer cells. The results revealed that *miR-429* was not only a useful biomarker of EMT in ovarian cancer, but also of potential therapeutic value in abating ovarian cancer metastasis [92]. miRNAs involved in regulating expression of *miR-200* family miRNAs, such as, *miR-103*/*107* can induce EMT by downregulating *miR-200* levels, and inhibition of *miR-103*/*107* abrogates migration and metastasis of malignant cells [93]. *miR-205*, implicated in metastasis is known to play a role in the development of radio-resistance of human nasopharyngeal carcinoma cells by directly targeting *PTEN* tumor suppressor protein and inhibition of *miR-205* induces apoptosis in these cells [94]. The *miR-221*/*222* miRNA cluster also can induce EMT in breast cancer cells and target *ESR1*, *Dicer*, and *TRPS1* protein expression [95–97]. Knockdown of the putative oncogenic *miR-182* expression has been shown to elevate the expression of chromobox homolog 7 (*CBX7*) and positively regulate the expression of E-cadherin, the protein regulating epithelial morphology of the cells [98]. Given these findings, it is evident that altered expression of the miRNAs involved in the determination of epithelial and mesenchymal lineages, discussed above, play roles in tumor metastasis.

Epigenetic modifications are heritable and reversible biochemical changes of the chromatin structure that regulate gene expression via chromatin remodeling including DNA methylation, histone modifications, and through miRNAs involved in these processes [99, 100]. DNA methylation is a type of covalent modification in which a methyl group is added to a cytosine in the genome via S-adenosylmethionine and restricted to the 5-position of the pyrimidine ring of cytosine residues that are located in *CpG* dinucleotides [101, 102]. *CpG* islands are frequently found within gene promoter regions, and DNA methylation of promoter-associated *CpG* islands is associated with gene repression, through either a direct or indirect influence on the chromatin structure that results in chromatin condensation [99, 103, 104]. Approximately 34–50 % of all annotated human miRNA genes are located in fragile

sites or in areas of the genome that are associated with cancer [13, 105–107]. On average, *CpG* islands are 32 % more frequent on fragile sites and also positively correlated with miRNA and protein-coding genes [107]. Epigenetic mechanisms regulate not only the expression of protein-encoding genes but also expression of miRNAs, such as *let-7a*, *miR-9*, *miR-34*, *miR-124*, *miR-137*, *miR-148*, and *miR-203* [108], and many miRNAs are found near *CpG* islands [109]. The *let-7a-3* locus was found to be hypomethylated in lung adenocarcinoma, and elevated expression of this locus resulted in enhanced oncogenic gene transcription [110]. A high correlation was found between miRNAs and *CpG* islands with several miRNA loci (including *miR-9*, *miR-193a*, *miR-137*, *miR-342*, *miR-203*, and *miR-34b/c*) found hypermethylated in multiple human cancers [107, 111]. *miR-127*, a tumor-suppressor miRNA targeting *BCL-6*, was found to be embedded in a *CpG* island and silenced by DNA methylation [112, 113]. *miR-34b/c CpG* island is a bidirectional promoter, and the *CpG* island methylation in colorectal cancer is associated with the silencing of both *miR-34b/c* and *BTG4* [114]. On the other hand, some miRNAs also can control epigenetic mechanisms and induce gene silencing by targeting a specific gene region for DNA methylation and histone modifications [108, 115]. The *miR-29* family directly targets DNA methyltransferases (*DNMTs*), and exogenous *miR-29* can reactivate methylation-silenced tumor-suppressor genes by restoring normal patterns of DNA methylation in NSCLC cells [116]. Interestingly, *miR-148* has also been shown to silence the human *DNMT3b* gene expression, though not by binding to the 3′UTR but by targeting the protein coding region of the transcript [117].

In view of the robust regulatory function of miRNAs as posttranscriptional repressors of gene expression, it is not unexpected that compromised miRNA activities have important biological implications in cancer [118]. Recent discovery of endogenous transcripts acting as natural miRNA decoys, called "competitive endogenous RNAs" or "*ceRNAs*" [119, 120] demonstrates that a very delicate balance is maintained between the miRNA seed families and their total pool of mRNAs in normal cells that is attenuated in tumor cells with major physiologic and phenotypic consequences due to altered regulatory functions of miRNAs. In view of the observations that tumor growth is accelerated in models of global miRNA depletion [121] and 3′ UTRs are frequently shortened in tumors due to alternative polyadenylation site choice [122], it is natural that such anomalies in miRNA regulatory processes cause profound changes in gene expression patterns contributing to accelerated evolution of tumor cells.

## 4.4 miRNAs as Novel Potential Biomarkers For Solid Tumors

Given the widespread tissue-specific dysregulation of miRNA expression detected in cancer, multiple studies have explored the potential usefulness of miRNA expression profiles as biomarkers of the disease initiation, progression and response to therapy. High throughput microarray analyses of miRNAs have been used to identify the

tissue of origin for tumors of unknown origin and in poorly differentiated tumors to distinguish different tumor subtypes. Such profiling of miRNAs has revealed that accumulation of *miR-21* and *miR-181a* in bone marrow appears to be associated with prognosis in breast cancer patients and is predictor for recurrence and metastasis [123]. The high expression level of *miR-21* was associated with poor therapeutic outcome and survival in colon adenocarcinoma as well as response to gemcitabine therapy among pancreatic ductal adenocarcinoma (PDAC) patients [124, 125]. For liver cancer patients, low *miR-26* expression correlated with shorter overall survival but better response to interferon therapy than those with high expression of this miRNA [126]. Reduced *let-7* expression in human lung cancers was associated with significantly shorter survival after resection [37]. Loss of *miR-126* and *miR-335* expression in breast tumor patients undergoing relapse associated with poor distal metastasis-free survival [87]. In 2008, two research groups identified tumor-specific alterations in circulating miRNAs of cancer patients and showed the potential of serum and plasma circulating miRNAs as novel noninvasive biomarkers for the early diagnosis of cancers and other diseases [127, 128], with the former publication validating the origin of these miRNAs from cancer tissues.

The high throughput technologies of miRNA microarray, real-time PCR array, and the next-generation Solexa sequencing generate circulating miRNA signatures from body fluids. The microarray platform enables simultaneous analysis of all human miRNAs by either fluorescent or electrochemical signals, and this technique can be redesigned to include newly identified miRNAs. The miRNA profiles in serum from human prostate cancer patients analyzed by high-density miRNA microarray revealed upregulation of 15 miRNAs (*miR-16*, *-92a*, *-103*, *-107*, *-197*, *-34b*, *-328*, *-485-3p*, *-486-5p*, *-92b*, *-574-3p*, *-636*, *-640*, *-766*, *-885-5p*) compared with normal donor serum [129]. Using microarray-based expression profiling followed by real-time quantitative PCR (RT-qPCR) validation, Zhao et al. found that 18–31 miRNAs were differentially expressed in subjects with early-stage breast cancer by comparing the levels of circulating miRNAs in plasma samples from 20 women with early-stage breast cancer and 20 matched healthy controls [130]. Following another low density microarray study, five miRNAs (*let-7f*, *miR-20b*, *miR-30e-3p*, *miR-223*, and *miR-301*) were validated by real-time PCR in plasma from 78 NSCLC patients and 48 controls and correlated with pathologic parameters and survival [131]. Next-generation Solexa sequencing with RT-qPCR assay validation in NSCLC demonstrated that a set of 11 serum miRNAs were differentially expressed between age, sex and stage matched patient cohorts stratified based on longer or shorter survival, and among this set, four (*miR-486*, *miR-30d*, *miR-1*, and *miR-499*) were associated with decreased overall survival [132]. A Solexa sequencing study also demonstrated that 19 serum miRNAs were markedly upregulated in gastric cancer patients and 25 circulating miRNAs were overexpressed in patients with esophageal squamous cell carcinoma (ESCC) compared with their matched controls [133, 134]. A logistic regression model with the best prediction defined on the basis of four genes (*miRNA-21*, *miR-126*, *miR-210*, and *miR-486-5p*), yielded 86.22 % sensitivity and 96.55 % specificity in distinguishing NSCLC patients from the healthy controls. Furthermore, the panel of miRNAs produced 73.33 %

sensitivity and 96.55 % specificity in identifying stage I NSCLC patients [135]. The expression levels of five serum miRNAs (*miR-1*, *miR-20a*, *miR-27a*, *miR-34*, and *miR-423-5p*) were found to correlate with gastric cancer tumor stage. The areas under the receiver operating characteristic (ROC) curve of this five-serum miRNA signature were 0.879 (95 % CI 0.822–0.936) and 0.831 (95 % CI 0.767–0.898) for the two sets of serum samples, which were markedly higher than the routinely used biomarkers *CEA* (0.503) and *CA19-9* (0.600) [136]. RT-qPCR analysis also identified a profile of seven serum miRNAs (*miR-10a*, *miR-22*, *miR-100*, *miR-148b*, *miR-223*, *miR-133a*, and *miR-127-3p*) as ESCC biomarkers. The area under the ROC curve for the selected miRNAs ranged from 0.817 to 0.949, significantly higher (0.549, $P<0.0005$) than that obtained for carcinoembryonic antigen (*CEA*). More importantly, this panel of seven miRNAs clearly distinguished stage I/II ESCC patients from controls [134]. Five serum miRNAs (*miR-21*, *miR-92*, *miR-93*, *miR-126*, and *miR-29a*) were reported to be significantly overexpressed in a set of 19 samples from epithelial ovarian cancer patients before therapy compared with healthy controls [137]. The combined expression analyses of *miR-21*, *miR-210*, *miR-155*, and *miR-196a* in plasma was found to distinguish pancreatic adenocarcinoma patients from controls [138]. Interestingly, members of this miRNA panel appear to be involved in multiple different human cancers. For example, *miR-21* has been found significantly increased in the plasma/serum of patients with breast cancer [139, 140], NSCLC [141], PDAC [138], gastric cancer [142], lung cancer [135, 136], HCC [143], prostate cancer [144], ESCC [145], and diffuse large B-cell lymphoma (DLBCL) [146]. The circulating *miR-21* has been suggested as an independent prognostic marker for cancer patients [139, 141]; associated with relapse-free survival in DLBCL patients [146] and involved in docetaxel-based drug resistance in patients with hormone-refractory prostate cancer [144]. Elevated expression level of *miR-155* also was detected in plasma/serum of patients with DLBCL [146], breast cancer [147, 148], and lung cancer [149]. Circulating *miR-210*, was found highly expressed in serum/plasma of patients with lung cancer [135], DLBCL [146], and as a marker of hypoxic signature associated with adverse outcome in pancreatic cancer [150]. However, some miRNAs have been reported to reveal discordant expression patterns in plasma/serum from different cancers: for example, the levels of *miR-92a* was found significantly higher in the plasma/serum from patients with advanced-stage colorectal cancer [151] and breast cancer [140], but detected at low levels in the plasma of patients with acute leukemia [152] and HCC [153] compared with their matched healthy counterparts.

Association of circulating miRNA levels with the presence of tumors is most convincingly established by observations that their levels in circulation get restored to the normal range after resection of the tumors. For example, serum levels of upregulated miRNAs such as *miR-21* and *miR-106b*, found significantly higher in preoperative plasma from patients with gastric cancer were reduced after tumor resection [142]. Similarly, elevated *miR-500* returned to normal after tumor resection in three HCC patients [154]. In breast cancer patients, increased systemic *miR-195* and *let-7a* levels in tumors, and circulation also were reported to go down after tumor resection [155]. Surgical removal of the primary tumor coincided with significant

reduction in plasma *miR-184* levels in patients with squamous cell carcinoma of the tongue [156]. Expression levels of *miR-96* and *miR-183* in urine was found to be low after surgery in urothelial carcinoma patients [157].

Although the majority of studies have assessed circulating miRNAs in serum and plasma, recent studies have indicated that tumor-specific miRNAs in other body fluids may also serve as sensitive and specific biomarkers for cancer. The presence of such miRNAs in body fluids represents not only a gold mine of circulating cancer biomarkers for detection, diagnosis and prognosis but also surrogate therapeutic targets. For instance, significantly increased *miR-96* and *miR-183* levels in urine were associated with advancing tumor grade and pathological stage in patients with urothelial carcinoma [157]. Elevation in urine *miR-1236*, *miR-374a*, *miR-767-3p* and absence of *miR-200a* and *miR-891b* were also observed in samples from patients with bladder urothelial cancers [158]. Significantly low levels of *miR-125a* and *miR-200a* in the saliva of cancer patients with oral squamous cell carcinoma compared with matched healthy controls has been reported [159]. In case of HCC, sensitivity of urine *miR-650* levels and the specificity of the combination of *miR-618* and *miR-650* levels for detecting disease were found greatly improved compared with the sensitivity and specificity of the conventionally used AFP-based detection assay [160]. To date, miRNAs have been detected in tears, breast milk, bronchial lavage, colostrum, as well as seminal, amniotic, pleural, peritoneal, and cerebrospinal fluids [158, 161]. Unsupervised hierarchical clustering analysis of the commonly expressed miRNAs showed that the miRNA spectrum in plasma was different from that of most other body fluids [158]. A recent study has also revealed that blood cells contribute to circulating miRNA profiles and that perturbations in blood cell counts and hemolysis can alter plasma miRNA biomarker levels by up to 50-fold [162]. These findings suggest that caution should be exercised in interpreting the circulating micro RNA profiling data since these may sometimes reflect body fluid specific cell based phenomenon rather than variations in the levels of cancer specific biomarkers.

In case of circulating miRNAs reflecting the levels expressed in tumors, the mechanisms underlying their release in circulation remains largely unclear. The extracellular RNA is most likely protected within protein or lipid vesicles, possibly apoptotic bodies. Circulating miRNAs in body fluids and extracellular fluid compartments have been proposed to have hormone-like effects with significant physiological consequences in cells at distant sites [161]. miRNAs may also passively leak from apoptotic or necrotic cells, which has been shown to occur after heart tissue injury [163, 164]. The secretion of miRNAs by tumor cells is associated with their ability to influence the surrounding microenvironment for their own benefit. Pre-miRNA molecules in the cytoplasm can bind to specific proteins responsible for their stability and in association with multivesicular bodies (*MVBs*) and exosomes, can deliver the miRNAs to the recipient cells by the process of endocytosis through inward budding of endosomal membranes, giving rise to intracellular *MVBs* that later fuse with the plasma membrane, releasing the exosomes into the circulating compartments and bloodstream [165]. The exosomes containing mRNA, miRNAs, and angiogenic proteins released by tumor cells have been detected [166]. Secretion

of microvesicle-free miRNAs, selectively packed into microparticles and exosomes, released by cells via shedding of microvesicles could be derived from tumor cells. With the capability to be efficiently transferred between cells, circulating miRNAs, particularly cell-derived microvesicle-contained miRNAs, may be an essential part of the cellular responses to exogenous challenges [167, 168].

## 4.5 miRNAs as Potential Therapeutic Targets in Cancer

As mentioned above, miRNAs are involved in the regulation of cell proliferation, differentiation, apoptosis, and survival, as well as tumor development, metastasis and drug resistance. The loss and gain of miRNA functions contribute to cancer initiation and progression and are, therefore, involved in the pathogenesis of cancer. So, differentially expressing tumor associated miRNAs are being evaluated as potential therapeutic targets for cancer treatment. The strategies for miRNA-based therapies include inhibiting oncogenic miRNAs and restoring tumor-suppressor miRNAs. Direct inhibition of miRNAs can be achieved by using antisense oligonucleotides to bind and sequester the target miRNA. A recent study has shown that chemically engineered oligonucleotides, as antagomirs, which are specific inhibitors of endogenous miRNAs in mice, could be used to silence oncogenic miRNAs, such as the *miR-17-92* cluster, for anticancer therapy [169]. The antagonism of *miR-122* by the LNA oligonucleotide *SPC3649* leads to marked suppression of viremia in chronically HCV-infected chimpanzees [170]. At the same time, chemotherapy in combination with miRNA-based therapies is being investigated. For example, *miR-181a* and *miR-630* play roles in regulating DNA damage-induced apoptosis. *miR-181a* can sensitize cells to cisplatin-induced DNA damage by stimulating *Bax* oligomerization and activating proapoptotic caspases, whereas *miR-630* confers cytoprotection resulting from decreased proliferation coupled to upstream inhibition of the signaling cascades that emanate from damaged DNA [171].

Restoring tumor-suppressor miRNAs by miRNA replacement involves reintroducing synthetic miRNA mimics that rescue the gene expression regulation abrogated in absence of the tumor suppressor miRNA in tumor cells. Designing of *miR-34* tumor suppressor therapy may be among the first treatments using miRNA mimics to reach the clinic [172]. *miR-34a* delivered by *GC4*-targeted nanoparticles significantly downregulated the expression of survivin in the metastatic tumors and reduced tumor load in the lung [173]. Systemic delivery of *miR-34a* mimics also effectively reduced advanced lung tumors in a *KRAS*-activated NSCLC mouse model through inhibition of proliferation and induction of apoptosis [174]. The growth of orthotopic xenografts was reported to be significantly inhibited by restoration of either *miR-34a* or *miR-143/145* through systemic nanovector delivery [175]. It is interesting that those miRNAs with tumor-suppressive effect, such as members of the *miR-200* family, have been shown to prime cancer stem cells for drug sensitivity. In a breast cancer model containing tumor-initiating cells with a *CD44+/CD24−* phenotype, introduction of *miR-128* made the cells more sensitive to doxorubicin treatment compared with the control parental tumor cells [176].

Using polyurethane polyethylenimine (PU-PEI) as the therapeutic delivery vehicle, PU-PEI mediated delivery of *miR-145* inhibited the tumorigenic cancer stem cell (CSC) like properties of *CD133+* glioblastoma cells facilitating their differentiation into non CSC like *CD133−* cells, which showed reduced antiapoptotic potential and sensitivity to temozolomide [177]. These studies point to the usefulness of miRNA mediated therapeutic strategies in inducing differentiation and restoration of drug sensitivity to CSC [178]. A novel epigenetic therapy approach with inhibitors of DNA methylation and *HDAC* to activate miRNAs was demonstrated that involved activation of tumor-suppressor miRNAs such as *miR-127* by chromatin-modifying drugs, which could have an anticancer effect through downregulation of the target oncogenes [112].

For miRNA-based therapies to work, it will be necessary to develop effective delivery mechanisms that will ensure stability of the molecules and their sufficient uptake in target cells while minimizing off-target adverse effects. This would also require optimizing the chemistry and size of the synthetic oligonucleotides to achieve better binding efficiency and increased stability. Delivery mechanisms range from the use of adenoviral vector-based delivery to cationic liposomes to polymer-based nanoparticles [179]. New miRNA-based treatments would require selective and accurate delivery of the agents to the target tumors to increase therapeutic potential and reduce possible side effects [180]. In one study a neutral lipid emulsion did not lead to accumulation of miRNA in liver and displayed excellent delivery to normal lung and orthotopic lung tumors [174]. Targeted miRNA delivery to specific tissues is possible when tumor-specific ligands are linked to nanoparticles that can then be selectively directed to tumor cells.

Finally, targeting an individual gene or a subset of genes with multiple tumor-suppressive miRNAs should enhance the therapeutic effect, and the combined use of two or three miRNAs that target the same gene would decrease the likelihood of mutation-induced resistance [181].

## 4.6 Conclusion

The discovery of differentially expressed tumor associated miRNAs has unraveled novel genetic mechanisms underlying deregulated gene expression in solid tumors. Identification of oncogenic and tumor suppressor miRNAs affecting critical tumor relevant genetic networks and cellular pathways makes it imperative that investigation of miRNA regulatory networks be pursued in depth to elucidate their detailed functional involvement in the initiation and progression of tumors. Findings obtained so far offer compelling evidence in favor of miRNAs being informative cancer biomarkers and their presence in body fluids make it extremely likely that profiling of circulating miRNAs hold the promise of being developed as minimally invasive sensitive and specific biomarker assays for cancer in the near future. Furthermore, preclinical studies strongly suggest that tumor associated miRNAs may prove to be effective therapeutic targets for treatment of cancers.

**Acknowledgements** The biomarker discovery work in Subrata Sen's laboratory is supported by a grant from the Early Detection Research Network of the National Cancer Institute/National Institutes of Health, USA.

## References

1. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75:843–54.
2. Ambros V. The functions of animal microRNAs. Nature. 2004;431:350–5.
3. Gregory RI, Chendrimada TP, Cooch N, et al. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell. 2005;123:631–40.
4. Wang J, Sen S. MicroRNA functional network in pancreatic cancer: from biology to biomarkers of disease. J Biosci. 2011;36:481–91.
5. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism and function. Cell. 2004;116: 281–97.
6. Kusenda B, Mraz M, Mayer J, et al. MicroRNA biogenesis, functionality and cancer relevance. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2006;150:205–15.
7. Höck J, Meister G. The Argonaute protein family. Genome Biol. 2008;9:210.
8. Shantikumar S, Caporali A, Emanueli C. Role of microRNAs in diabetes and its cardiovascular complications. Cardiovasc Res. 2012;93:583–93.
9. Ono K, Kuwabara Y, Han J. MicroRNAs and cardiovascular diseases. FEBS J. 2011;278:1619–33.
10. Kloosterman WP, Plasterk RH. The diverse functions of microRNAs in animal development and disease. Dev Cell. 2006;11:441–50.
11. Mak RH, Cheung WW. MicroRNAs: a new therapeutic frontier for muscle wasting in chronic kidney disease. Kidney Int. 2012;82:373–4.
12. Bushati N, Cohen SM. MicroRNAs in neurodegeneration. Curr Opin Neurobiol. 2008;18:292–6.
13. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of microRNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99:15524–9.
14. Cimmino A, Calin GA, Fabbri M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005;102:13944–9.
15. Calin GA, Cimmino A, Fabbri M, et al. MiR-15a and miR-16-1 cluster functions in human leukemia. Proc Natl Acad Sci U S A. 2008;A105:5166–71.
16. Xia L, Zhang D, Du R, et al. miR-15b and miR-16 modulate multidrug resistance by targeting BCL2 in human gastric cancer cells. Int J Cancer. 2008;123:372–9.
17. Dai L, Wang W, Zhang S, et al. Vector-based miR-15a/16-1 plasmid inhibits colon cancer growth in vivo. Cell Biol Int. 2012;36:765–70.
18. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435:834–8.
19. Volinia S, Calin GA, Liu CG, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006;103:2257–61.
20. Martens-Uzunova ES, Jalava SE, Dits NF, et al. Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. Oncogene. 2012;31:978–91.
21. Farazi TA, Horlings HM, Ten Hoeve JJ, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res. 2011;71:4443–53.
22. Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med. 2008;359:1995–2004.
23. Barshack I, Lithwick-Yanai G, Afek A, et al. MicroRNA expression differentiates between primary lung tumors and metastases to the lung. Pathol Res Pract. 2010;206:578–84.

24. Liu A, Tetzlaff MT, Vanbelle P, et al. MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues. Int J Clin Exp Pathol. 2009;2:519–27.
25. Fridman E, Dotan Z, Barshack I, et al. Accurate molecular classification of renal tumors using microRNA expression. J Mol Diagn. 2010;12:687–96.
26. Tetzlaff MT, Liu A, Xu X, et al. Differential expression of miRNAs in papillary thyroid carcinoma compared to multinodular goiter using formalin fixed paraffin embedded tissues. Endocr Pathol. 2007;18:163–73.
27. Weng L, Wu X, Gao H, et al. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffin-embedded tissue specimens. J Pathol. 2010;222:41–51.
28. Andreasen D, Fog JU, Biggs W, et al. Improved microRNA quantification in total RNA from clinical samples. Methods. 2010;50:S6–9.
29. Yu S, Lu Z, Liu C, et al. miRNA-96 suppresses KRAS and functions as a tumor suppressor gene in pancreatic cancer. Cancer Res. 2010;70:6015–25.
30. Welch C, Chen Y, Stallings RL. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. Oncogene. 2007;26:5017–22.
31. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature. 2007;449:682–8.
32. Selcuklu SD, Donoghue MT, Spillane C. miR-21 as a key regulator of oncogenic processes. Biochem Soc Trans. 2009;37:918–25.
33. Jiang S, Zhang HW, Lu MH, et al. MicroRNA-155 functions as an OncomiR in breast cancer by targeting the suppressor of cytokine signaling 1 gene. Cancer Res. 2010;70:3119–27.
34. Esquela-Kerscher A, Slack FJ. Oncomirs–microRNAs with a role in cancer. Nat Rev Cancer. 2006;6:259–69.
35. Hurst DR, Edmonds MD, Welch DR. Metastamir: the field of metastasis-regulatory microRNA is spreading. Cancer Res. 2009;69:7495–8.
36. Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. Cell. 2005;120:635–47.
37. Takamizawa J, Konishi H, Yanagisawa K, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res. 2004;64:3753–6.
38. Ratner E, Lu L, Boeke M, et al. A KRAS-variant in ovarian cancer acts as a genetic marker of cancer risk. Cancer Res. 2010;70:6509–15.
39. Hollestelle A, Pelletier C, Hooning M, et al. Prevalence of the variant allele rs61764370 T>G in the 3′UTR of KRAS among Dutch BRCA1, BRCA2 and non-BRCA1/BRCA2 breast cancer families. Breast Cancer Res Treat. 2011;128:79–84.
40. Ruzzo A, Graziano F, Vincenzi B, et al. High let-7a microRNA levels in KRAS-mutated colorectal carcinomas may rescue anti-EGFR therapy effects in eatients with chemotherapy-refractory metastatic disease. Oncologist. 2012;17:823–9.
41. Hatley ME, Patrick DM, Garcia MR, et al. Modulation of K-Ras-dependent lung tumorigenesis by microRNA-21. Cancer Cell. 2010;18:282–93.
42. Si ML, Zhu S, Wu H, et al. miR-21-mediated tumor growth. Oncogene. 2007;26:2799–803.
43. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6:857–66.
44. Asangani IA, Rasheed SA, Nikolova DA, et al. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. Oncogene. 2008;27:2128–36.
45. Yang CH, Yue J, Pfeffer SR, et al. MicroRNA miR-21 regulates the metastatic behavior of B16 melanoma cells. J Biol Chem. 2011;286:39172–8.
46. Baumhoer D, Zillmer S, Unger K, et al. MicroRNA profiling with correlation to gene expression revealed the oncogenic miR-17-92 cluster to be up-regulated in osteosarcoma. Cancer Genet. 2012;205:212–9.
47. Xiang J, Wu J. Feud or friend? The role of the miR-17-92 cluster in tumorigenesis. Curr Genomics. 2010;11:129–35.

48. Schmidt H, Bartel F, Kappler M, et al. Gains of 13q are correlated with a poor prognosis in liposarcoma. Mod Pathol. 2005;18:638–44.
49. Yang Y, Li X, Yang Q, et al. The role of microRNA in human lung squamous cell carcinoma. Cancer Genet Cytogenet. 2010;200:127–33.
50. Yao J, Liang L, Huang S, et al. MicroRNA-30d promotes tumor invasion and metastasis by targeting Galphai2 in hepatocellular carcinoma. Hepatology. 2010;51:846–56.
51. Li N, Kaur S, Greshock J, et al. A combined array-based comparative genomic hybridization and functional library screening approach identifies mir-30d as an oncomir in cancer. Cancer Res. 2012;72:154–64.
52. Esposito F, Tornincasa M, Pallante P, et al. Down-regulation of the miR-25 and miR-30d contributes to the development of anaplastic thyroid carcinoma targeting the polycomb protein EZH2. J Clin Endocrinol Metab. 2012;97:E710–8.
53. Martinez I, Cazalla D, Almstead LL, et al. miR-29 and miR-30 regulate B-Myb expression during cellular senescence. Proc Natl Acad Sci U S A. 2011;108:522–7.
54. Kumarswamy R, Mudduluru G, Ceppi P, et al. MicroRNA-30a inhibits epithelial-to-mesenchymal transition by targeting Snai1 and is downregulated in non-small cell lung cancer. Int J Cancer. 2012;130:2044–53.
55. Nass D, Rosenwald S, Meiri E, et al. MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. Brain Pathol. 2009;19:375–83.
56. Khew-Goodall Y, Goodall GJ. Myc-modulated miR-9 makes more metastases. Nat Cell Biol. 2010;12:209–11.
57. Ma L, Young J, Prabhala H, et al. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nat Cell Biol. 2010;12:247–56.
58. Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005;65:7065–70.
59. Sun Y, Wu J, Wu SH, et al. Expression profile of microRNAs in c-Myc induced mouse mammary tumors. Breast Cancer Res Treat. 2009;118:185–96.
60. Tan HX, Wang Q, Chen LZ, et al. MicroRNA-9 reduces cell invasion and E-cadherin secretion in SK-Hep-1 cell. Med Oncol. 2010;27:654–60.
61. Luo H, Zhang H, Zhang Z, et al. Down-regulated miR-9 and miR-433 in human gastric carcinoma. J Exp Clin Cancer Res. 2009;28:82.
62. Wan HY, Guo LM, Liu T, et al. Regulation of the transcription factor NF-κB1 by microRNA-9 in human gastric adenocarcinoma. Mol Cancer. 2010;9:16.
63. Guo LM, Pu Y, Han Z, et al. MicroRNA-9 inhibits ovarian cancer cell growth through regulation of NF-κB1. FEBS J. 2009;276:5537–46.
64. Bandres E, Agirre X, Bitarte N, et al. Epigenetic regulation of microRNA expression in colorectal cancer. Int J Cancer. 2009;125:2737–43.
65. Hildebrandt MA, Gu J, Lin J, et al. Hsa-miR-9 methylation status is associated with cancer development and metastatic recurrence in patients with clear cell renal cell carcinoma. Oncogene. 2010;29:5724–8.
66. Lehmann U, Hasemeier B, Christgen M, et al. Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. J Pathol. 2008;214:17–24.
67. Hsu PY, Deatherage DE, Rodriguez BA, et al. Xenoestrogen-induced epigenetic repression of microRNA-9-3 in breast epithelial cells. Cancer Res. 2009;69:5936–45.
68. Selcuklu SD, Donoghue MT, Rehmet K, et al. MicroRNA-9 inhibition of cell proliferation and identification of novel miR-9 targets by transcriptome profiling in breast cancer cells. J Biol Chem. 2012;287:29516–28.
69. Liu ZL, Wang H, Liu J, et al. MicroRNA-21 (miR-21) expression promotes growth, metastasis, and chemo- or radioresistance in non-small cell lung cancer cells by targeting PTEN. Mol Cell Biochem. 2013;372:35–45. doi:10.1007/s11010-012-1443-3.
70. Segura MF, Hanniford D, Menendez S, et al. Aberrant miR-182 expression promotes melanoma metastasis by repressing FOXO3 and microphthalmia-associated transcription factor. Proc Natl Acad Sci U S A. 2009;106:1814–9.

71. Liu Z, Liu J, Segura MF, et al. MiR-182 overexpression in tumourigenesis of high-grade serous ovarian carcinoma. J Pathol. 2012;228:204–15. doi:10.1002/path.4000.
72. Huynh C, Segura MF, Gaziel-Sovran A, et al. Efficient in vivo microRNA targeting of liver metastasis. Oncogene. 2011;30:1481–8.
73. Mihelich BL, Khramtsova EA, Arva N, et al. miR-183-96-182 cluster is overexpressed in prostate tissue and regulates zinc homeostasis in prostate cells. J Biol Chem. 2011; 286:44503–11.
74. Guttilla IK, White BA. Coordinate regulation of FOXO1 by miR-27a, miR-96, and miR-182 in breast cancer cells. J Biol Chem. 2009;284:23204–16.
75. Myatt SS, Wang J, Monteiro LJ, et al. Definition of microRNAs that repress expression of the tumor suppressor gene FOXO1 in endometrial cancer. Cancer Res. 2010;70:367–77.
76. Valastyan S, Reinhardt F, Benaich N, et al. A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. Cell. 2009;137:1032–46.
77. Sossey-Alaoui K, Downs-Kelly E, Das M, et al. WAVE3, an actin remodeling protein, is regulated by the metastasis suppressor microRNA, miR-31, during the invasion-metastasis cascade. Int J Cancer. 2011;129:1331–43.
78. Schaefer A, Jung M, Mollenkopf HJ, et al. Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. Int J Cancer. 2010;126:1166–76.
79. Creighton CL, Fountain MD, Yu Z, et al. Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. Cancer Res. 2010;70:1906–15.
80. Guo J, Miao Y, Xiao B, et al. Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. J Gastroenterol Hepatol. 2009;24:652–7.
81. Zhang Y, Guo J, Li D, et al. Downregulation of miR-31 expression in gastric cancer tissues and its clinical significance. Med Oncol. 2010;27:685–9.
82. Veerla S, Lindgren D, Kyist A, et al. MiRNA expression in urothelial carcinomas: important roles of miR-10a, miR-222, miR-125b, miR-7 and miR-452 for tumor stage and metastasis, and frequent homozygous losses of miR-31. Int J Cancer. 2009;124:2236–42.
83. Usvasalo A, Ninomiya S, Räty R, et al. Focal 9p instability in hematologic neoplasias revealed by comparative genomic hybridization and single-nucleotide polymorphism microarray analyses. Genes Chromosomes Cancer. 2010;49:309–18.
84. Bhaumik D, Scott GK, Schokrpur S, et al. Expression of microRNA-146 suppresses NF-κB activity with reduction of metastatic potential in breast cancer cells. Oncogene. 2008;42:5643–7.
85. Taganov KD, Boldin MP, Chang KJ, et al. NF-κB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. Proc Natl Acad Sci U S A. 2006;A103:12481–6.
86. Lin SL, Chiang A, Chang D, et al. Loss of mir-146a function in hormone-refractory prostate cancer. RNA. 2008;14:417–24.
87. Tavazoie SF, Alarcon C, Oskarsson T, et al. Endogenous human microRNAs that suppress breast cancer metastasis. Nature. 2008;451:147–52.
88. Vickers MM, Bar J, Gorn-Hondermann I, et al. Stage-dependent differential expression of microRNAs in colorectal cancer: potential role as markers of metastatic disease. Clin Exp Metastasis. 2012;29:123–32.
89. Hur K, Toiyama Y, Takahashi M, et al. MicroRNA-200c modulates epithelial-to-mesenchymal transition (EMT) in human colorectal cancer metastasis. Gut. 2012. doi:10.1136/gutjnl-2011-301846.
90. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer. 2009;9:265–73.
91. Gregory PA, Bert AG, Paterson EL, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008;10:593–601.
92. Chen J, Wang L, Matyunina LV, et al. Overexpression of miR-429 induces mesenchymal-to-epithelial transition (MET) in metastatic ovarian cancer cells. Gynecol Oncol. 2011; 121:200–5.

93. Martello G, Rosato A, Ferrari F, et al. A microRNA targeting Dicer for metastasis control. Cell. 2010;141:1195–207.
94. Qu C, Liang Z, Huang J, et al. MiR-205 determines the radioresistance of human nasopharyngeal carcinoma by directly targeting PTEN. Cell Cycle. 2012;11:785–96.
95. Di Leva G, Gasparini P, Piovan C, et al. MicroRNA cluster 221-222 and estrogen receptor α interactions in breast cancer. J Natl Cancer Inst. 2010;102:706–21.
96. Cochrane DR, Cittelly DM, Howe EN, et al. MicroRNAs link estrogen receptor alpha status and Dicer levels in breast cancer. Horm Cancer. 2011;1:306–19.
97. Stinson S, Lackner MR, Adai AT, et al. miR-221/222 targeting of trichorhinophalangeal 1 (TRPS1) promotes epithelial-to-mesenchymal transition in breast cancer. Sci Signal. 2011;4(186):pt5.
98. Hannafon BN, Sebastiani P, de las Morenas A, et al. Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer. Breast Cancer Res. 2011;13:R24.
99. Chiam K, Ricciardelli C, Bianco-Miotto T. Epigenetic biomarkers in prostate cancer: current and future uses. Cancer Lett 2012 [Epub ahead of print].
100. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol. 2005;45:629–56.
101. Zuo T, Tycko B, Liu TM, et al. Methods in DNA methylation profiling. Epigenomics. 2009;1:331–45.
102. Goel A. DNA methylation-based fecal biomarkers for the noninvasive screening of GI cancers. Future Oncol. 2010;6:333–6.
103. Egger G, Liang G, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429:457–63.
104. Perry AS, Foley R, Woodson K, et al. The emerging roles of DNA methylation in the clinical management of prostate cancer. Endocr Relat Cancer. 2006;13:357–77.
105. Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101:2999–3004.
106. Sevignani C, Calin GA, Nnadi SC, et al. MicroRNA genes are frequently located near mouse cancer susceptibility loci. Proc Natl Acad Sci U S A. 2007;104:8017–22.
107. Laganà A, Russo F, Sismeiro C, et al. Variability in the incidence of miRNAs and genes in fragile sites and the role of repeats and CpG islands in the distribution of genetic material. PLoS One. 2010;5:e11166.
108. Sato F, Tsuchiya S, Meltzer SJ, et al. MicroRNAs and epigenetics. FEBS J. 2011;278:1598–609.
109. Ryan B, Robles A, Harris C. Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer. 2010;10:389–402.
110. Brueckner B, Stresemann C, Kuner R, et al. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. Cancer Res. 2007;67:1419–23.
111. Lujambio A, Calin GA, Villanueva A, et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A. 2008;105:13556–61.
112. Saito Y, Jones P. Epigenetic activation of tumor suppressor microRNAs in human cancer cells. Cell Cycle. 2006;5:2220–2.
113. Saito Y, Liang G, Egger G, et al. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell. 2006;9:435–43.
114. Toyota M, Suzuki H, Sasaki Y, et al. Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. Cancer Res. 2008;68:4123–32.
115. Matzke MA, Birchler JA. RNAi-mediated pathways in the nucleus. Nat Rev Genet. 2005;6:24–35.
116. Fabbri M, Garzon R, Cimmino A, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci U S A. 2007;104:15805–10.

117. Duursma AM, Kedde M, Schrier M, et al. miR-148 targets human DNMT3b protein coding region. RNA. 2008;14:872–7.
118. Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. Cell. 2012;149:515–24.
119. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146:353–8.
120. Tay Y, Kats L, Salmena L, Weiss D, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell. 2011;147:344–57.
121. Kumar MS, Lu J, Mercer KL, et al. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet. 2007;39:673–7.
122. Mayr C, Bartel DP. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009;138:673–84.
123. Ota D, Mimori K, Yokobori T, et al. Identification of recurrence-related microRNAs in the bone marrow of breast cancer patients. Int J Oncol. 2011;38:955–62.
124. Schetter AJ, Leung SY, Sohn JJ, et al. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. JAMA. 2008;299:425–36.
125. Giovannetti E, Funel N, Peters GJ, et al. MicroRNA-21 in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. Cancer Res. 2010;70:4528–38.
126. Ji J, Shi J, Budhu A, et al. MicroRNA expression, survival, and response to interferon in liver cancer. N Engl J Med. 2009;361:1437–47.
127. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105:10513–8.
128. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008;18:997–1006.
129. Lodes MJ, Caraballo M, Suciu D, et al. Detection of cancer with serum miRNAs on an oligonucleotide microarray. PLoS One. 2009;4:e6229.
130. Zhao H, Shen J, Medico L, et al. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. PLoS One. 2010;5:e13735.
131. Silva J, García V, Zaballos Á, et al. Vesicle-related microRNAs in plasma of nonsmall cell lung cancer patients and correlation with survival. Eur Respir J. 2011;37:617–23.
132. Hu Z, Chen X, Zhao Y, et al. Serum microRNA signatures identified in a genome-wide serum microRNA expression profiling predict survival of non-small-cell lung cancer. J Clin Oncol. 2010;28:1721–6.
133. Liu R, Zhang C, Hu Z, et al. A five-microRNA signature identified from genome-wide serum microRNA expression profiling serves as a fingerprint for gastric cancer diagnosis. Eur J Cancer. 2011;47:784–91.
134. Zhang C, Wang C, Chen X, et al. Expression profile of microRNAs in serum: a fingerprint for esophageal squamous cell carcinoma. Clin Chem. 2010;56:1871–9.
135. Shen J, Todd NW, Zhang H, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. Lab Invest. 2011;91:579–87.
136. Liu XG, Zhu WY, Huang YY, et al. High expression of serum miR-21 and tumor miR-200c associated with poor prognosis in patients with lung cancer. Med Oncol. 2012;29:618–26.
137. Resnick KE, Alder H, Hagan JP, et al. The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. Gynecol Oncol. 2009;112:55–9.
138. Wang J, Chen J, Chang P, et al. MicroRNAs in plasma of pancreatic ductal adenocarcinoma patients as novel blood-based biomarkers of disease. Cancer Prev Res (Phila). 2009;2:807–13.
139. Asaga S, Kuo C, Nguyen T, et al. Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. Clin Chem. 2011;57:84–91.
140. Wu Q, Lu Z, Li H, et al. Next-generation sequencing of microRNAs for breast cancer detection. J Biomed Biotechnol. 2011;2011:597145.
141. Wang ZX, Bian HB, Wang JR, et al. Prognostic significance of serum miRNA-21 expression in human non-small cell lung cancer. J Surg Oncol. 2011;104:847–51.

142. Tsujiura M, Ichikawa D, Komatsu S, et al. Circulating microRNAs in plasma of patients with gastric cancers. Br J Cancer. 2010;102:1174–9.
143. Xu J, Wu C, Che X, Wang L, et al. Circulating microRNAs, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis. Mol Carcinog. 2011; 50:136–42.
144. Zhang HL, Yang LF, Zhu Y, et al. Serum miRNA-21: elevated levels in patients with metastatic hormone-refractory prostate cancer and potential predictive factor for the efficacy of docetaxel-based chemotherapy. Prostate. 2011;71:326–31.
145. Komatsu S, Ichikawa D, Takeshita H, et al. Circulating microRNAs in plasma of patients with oesophageal squamous cell carcinoma. Br J Cancer. 2011;105:104–11.
146. Lawrie CH, Gal S, Dunlop HM, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol. 2008;141:672–5.
147. Zhu W, Qin W, Atasoy U, et al. Circulating microRNAs in breast cancer and healthy subjects. BMC Res Notes. 2009;2:89.
148. Roth C, Rack B, Müller V, et al. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. Breast Cancer Res. 2010;12:R90.
149. Rabinowits G, Gerçel-Taylor C, Day JM, et al. Exosomal microRNA: a diagnostic marker for lung cancer. Clin Lung Cancer. 2009;10:42–6.
150. Ho AS, Huang X, Cao H, et al. Circulating miR-210 as a novel hypoxia marker in pancreatic cancer. Transl Oncol. 2010;3:109–13.
151. Ng EK, Chong WW, Jin H, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut. 2009; 58:1375–81.
152. Tanaka M, Oikawa K, Takanashi M, et al. Down-regulation of miR-92 in human plasma is a novel marker for acute leukemia patients. PLoS One. 2009;4:e5532.
153. Shigoka M, Tsuchida A, Matsudo T, et al. Deregulation of miR-92a expression is implicated in hepatocellular carcinoma development. Pathol Int. 2010;60:351–7.
154. Yamamoto Y, Kosaka N, Tanaka M, et al. MicroRNA-500 as a potential diagnostic marker for hepatocellular carcinoma. Biomarkers. 2009;14:529–38.
155. Heneghan HM, Miller N, Lowery AJ, et al. Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. Ann Surg. 2010;251:499–505.
156. Wong TS, Ho WK, Chan JY, et al. Mature miR-184 and squamous cell carcinoma of the tongue. ScientificWorldJournal. 2009;9:130–2.
157. Yamada Y, Enokida H, Kojima S, et al. MiR-96 and miR-183 detection in urine serve as potential tumor markers of urothelial carcinoma: correlation with stage and grade, and comparison with urinary cytology. Cancer Sci. 2011;102:522–9.
158. Weber JA, Baxter DH, Zhang S, et al. The microRNA spectrum in 12 body fluids. Clin Chem. 2010;56:1733–41.
159. Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clin Cancer Res. 2009;15:5473–7.
160. Abdalla MA, Haj-Ahmad Y. Promising candidate urinary microRNA biomarkers for the early detection of hepatocellular carcinoma among high-risk hepatitis C virus Egyptian patients. J Cancer. 2012;3:19–31.
161. Cortez MA, Bueso-Ramos C, Ferdin J, et al. MicroRNAs in body fluids—the mix of hormones and biomarkers. Nat Rev Clin Oncol. 2011;8:467–77.
162. Pritchard CC, Kroh E, Wood B, et al. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila). 2012;5:492–7.
163. Ji X, Takahashi R, Hiura Y, et al. Plasma miR-208 as a biomarker of myocardial injury. Clin Chem. 2009;55:1944–9.
164. El-Hefnawy T, Raja S, Kelly L, et al. Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. Clin Chem. 2004;50:564–73.
165. Thery C, Zitvogel L, Amigorena S. Exosomes: composition, biogenesis and function. Nat Rev Immunol. 2002;2:569–79.

166. Skog J, Würdinger T, van Rijn S, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat. Cell Biol. 2008;10:1470–6.
167. Mostert B, Sieuwerts AM, Martens JW, et al. Diagnostic applications of cell-free and circulating tumor cell-associated miRNAs in cancer patients. Expert Rev Mol Diagn. 2011;11:259–75.
168. Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev. 2012;32:326–48.
169. Krutzfeldt J, Rajewsky N, Braich R, et al. Silencing of microRNAs in vivo with 'antagomirs'. Nature. 2005;438:685–9.
170. Lanford RE, Hildebrandt-Eriksen ES, Petri A, et al. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. Science. 2010;327:198–201.
171. Galluzzi L, Morselli E, Vitale I, et al. miR-181a and miR-630 regulate cisplatin-induced cancer cell death. Cancer Res. 2010;70:1793–803.
172. Bader AG. miR-34 – a microRNA replacement therapy is headed to the clinic. Front Genet. 2012;3:120.
173. Chen Y, Zhu X, Zhang X, et al. Nanoparticles modified with tumor-targeting scFv deliver siRNA and miRNA for cancer therapy. Mol Ther. 2010;18:1650–6.
174. Trang P, Wiggins JF, Daige CL, et al. Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. Mol Ther. 2011;19: 1116–22.
175. Pramanik D, Campbell NR, Karikari C, et al. Restitution of tumor suppressor microRNAs using a systemic nanovector inhibits pancreatic cancer growth in mice. Mol Cancer Ther. 2011;10:1470–80.
176. Zhu Y, Yu F, Jiao Y, et al. Reduced miR-128 in breast tumor-initiating cells induces chemotherapeutic resistance via Bmi-1 and ABCC5. Clin Cancer Res. 2011;17:7105–15.
177. Yang YP, Chien Y, Chiou GY, et al. Inhibition of cancer stem cell-like properties and reduced chemo-radio resistance of glioblastoma using microRNA145 with cationic polyurethane-short branch PEI. Biomaterials. 2012;33:1462–76.
178. Vira D, Basak SK, Veena MS, et al. Cancer stem cells, microRNAs, and therapeutic strategies including natural products. Cancer Metastasis Rev. 2012;31:733–51. doi:10.1007/s10555-012-9382-8.
179. Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. Cancer J. 2012;18:244–52.
180. Kong YW, Ferland-McCollough D, Jackson TJ, et al. microRNAs in cancer management. Lancet Oncol. 2012;13:e249–58.
181. Kasinski AL, Slack FJ. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. Nat Rev Cancer. 2011;11:849–64.

# Chapter 5
# MicroRNAs in Hematologic Malignancies

**Johannes Bloehdorn, Arefeh Rouhi, Pavan Kumar Dhanyamraju, Kathrin Krowiorz, Edith Schneider, Christian Langer, Daniel Mertens, Stephan Stilgenbauer, and Florian Kuchenbauer**

**Abstract** Besides normal hematopoiesis, microRNAs (miRNAs) have been found to be essentially involved in the development of various hematological malignancies. Here, we review the role of miRNAs in lymphoid neoplasias, with focus on lymphomas as well as myeloid malignancies such as acute myeloid leukemia.

**Keywords** miRNA • Cancer • Leukemia • Lymphoma • Oncogene • Tumor suppressor • Therapy • Survival

## 5.1 Introduction

Much evidence implicates miRNAs as contributing factors in the pathogenesis of hematological neoplasias. A provocative observation made by Calin et al. was that a large number of known recurrent genomic alterations involved in cancer are in close proximity to miRNA genes, suggesting that these rearrangements affect the expression of miRNAs with tumor suppressive or oncogenic properties. Indeed, multiple miRNAs have been implied in the pathogenesis of various neoplasias. Especially in chronic lymphocytic leukemia (CLL), the first connection between a frequent loss of chromosomal region 13q14.2 containing two miRNAs (miR-15a and miR-16-1) and the pathogenesis of the disease was discovered. Therefore, the pathophysiological role of miRNAs in lymphomas was recognized even before their regulatory role in normal hematopoiesis.

J. Bloehdorn • A. Rouhi • P.K. Dhanyamraju • K. Krowiorz • E. Schneider • C. Langer
D. Mertens • S. Stilgenbauer • F. Kuchenbauer, M.D., Ph.D. (✉)
Department of Internal Medicine III, University of Ulm, Albert-Einstein Allee 23,
89081 Ulm, Germany
e-mail: florian.kuchenbauer@uni-ulm.de

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_5,

## 5.2 Chronic Lymphocytic Leukemia

CLL is the most frequent form of leukemia in adults in the western world, affecting roughly 4 out of 100,000 people per year with a prevalence in men and a median age at diagnosis of 72 years [1]. People affected by CLL are usually not treated right from the time-point of diagnosis but with developing symptoms or progression to advanced stage [2]. Treatment has changed throughout the last decades switching from single substance treatment (with Fludarabine) to combined chemotherapy protocols (including Fludarabine and Chlorambucil) and with the advent of monoclonal antibodies to chemoimmunotherapy (with Fludarabine, Chlorambucil, and Rituximab) showing significant improvement in complete response (CR) rates [3]. Deletion of chromosome 17p and mutation of p53 have been clearly identified as predictive for refractoriness, but a considerable portion of cases with refractory disease or insufficient response and early relapse might not exhibit these established markers [4, 5]. Characterizing the genetic background and clinical presentation of CLL helped to identify risk groups the clinician can use for models to predict the clinical course, for management of follow up and rational treatment choice [5].

Classical features include recurring genomic aberrations and gene mutations such as TP53 [6] and ATM [7], somatic mutations in the variable regions of the immunoglobulin (Ig) heavy chain (IGHV) genes [8, 9], biased IGHV usage and stereotyped B cell receptors (BCRs) [8, 10, 11]. In about 80 % of all CLL cases chromosomal aberrations can be identified, mostly showing deletion of chromosome 13q14 (55 %) or 11q (18 %), trisomy of chromosome 12 (16 %), or deletion of chromosome 17p (7 %) [12]. Central genes identified at the minimal deleted region are the ataxia teleangiectasia-mutated (ATM) gene spanning the chromosome bands 11q22.3–q23.1 [13], the tumor suppressor gene TP53 at 17p13 [14, 15] and the cluster of two miRNAs named miR-15a and miR-16-1 within the DLEU2 gene on chromosome 13q14.2 [16].

The search for a putative target-gene of these miRNAs identified the BCL-2 gene by sequence-complementarity with the seed regions of both miRNAs. Functional validation confirmed the potential for posttranscriptional repression of BCL2 with an increased rate of apoptosis in MEG-01 cells in vitro and decreased tumorigenicity in xenograft mouse-models upon transfection with miR-15/16 [17, 18]. By generating a mouse model with a deletion of the DLEU2/miR-15a/16-1 cluster, Ulf Klein and colleagues proved the pathogenic effect mediated through the loss of these miRNAs. Mice with loss of the minimal deleted region or with sole miR-15a/16-1 deletion both developed clonal B-cell lymphoproliferation with a slightly pronounced effect when DLEU2 was affected as well. The loss of miR-15a and miR-16-1 affected growth, cell-cycle control and apoptosis, though the effect on BCL2 remains controversial [19]. As the involvement of DLEU2 already implicated additional mechanisms beside the loss of miR-15a and miR-16-1 that foster lymphoma development, Lia and colleagues provided another mouse model in which DLEU7 and RNASEH2B were additionally knocked out. Such mice developed

more aggressive lymphomas and presented with a phenotype of CLL or SLL (small lymphocytic lymphoma) [20]. Most interestingly in the context of these two mouse-model studies is the observation that patients with monoallelic 13q14 deletion tend to express higher miR-15a/16-1 levels than patients with a biallelic 13q14 deletion [21] and the growth kinetics of lymphocytes is slower in patients with a monoallelic compared to biallelic 13q14 deletion [22]. Although this phenotypes suggest differences in the clinical course, current studies draw a heterogeneous picture without a clear prognostic difference [23, 24]. With ongoing research, the functional network of miR-15/miR-16 steadily grows in complexity. Central target genes regulated by these miRNAs are involved in cell-cycle, cell-growth and apoptosis. Moreover, a regulatory loop with TP53 has been unmasked in a recent study by Fabbri and colleagues. While p53 can lead to the induction of miR-15a/miR-16-1 through upstream binding sites, the miRNA themselves specifically target TP53 and reduce protein and mRNA-levels [25]. Beside deletions or rare mutations leading to defective precursor transcript processing [26] epigenetic mechanisms have recently been identified as well. In the study by Sampath et al., epigenetic silencing of miR-15a, miR-16, and miR-29b mediated by histone deacetylases has been found in one-third of all investigated CLL samples. Exposure to histone deacetylases inhibitors led to the induction of all of these miRNAs and was associated with declines in the levels of MCL-1 but not BCL-2 [27]. With respect to the underlying (cyto-) genetic and newly identified epigenetic abnormalities and its clinical presentation, specific miRNA expression patterns help to uncover individual mechanims involved in CLL subgroups and to sharpen prognostic models [28–30]. Although the majority of deregulated miRNAs are not located at the commonly deleted regions in CLL, the regulatory changes seem to converge in similar functional routes as found for the p53-pathway. MiRNAs found to be specifically deregulated in conditions with dysfunctional p53 include miR-151-3p, miR-29c, miR-34a (downregulated) [28], miR-21, miR-155, miR-15a (upregulated) miR-34a, miR-181b (downregulated) [30], miR-34a, miR-29c, miR-17-5p (downregulated) [31]. Changes of miR-34a levels seem crucial, since this miRNA has been identified as a direct target of p53, taking a central part in the DNA-damage response [32–34]. The functional role and clinical relevance in CLL has been confirmed in a recent study showing that irradiation of CLL cells without functional p53 did not lead to induction of miR-34a. Low levels of miR-34a were found in association with fludarabine-refractory disease and impaired DNA-damage response even in cases without 17p-deletion or TP53 mutation.

Profiling studies aiming to discover miRNAs that could be further used as surrogate or prognostic markers were able to identify specific patterns of deregulation.

Initial studies using supervised approaches generated characteristic profiles based on the IGHV mutation status and ZAP-70 expression [26, 35]. Subsequent studies were able to confirm the association of decreased miR-223 levels and members of the miR-29 family with unmutated IGHV genes and disease progression [11, 21, 29, 36]. Shorter treatment free survival and reduced overall survival was shown for cases with low miR-223 and miR-29c. By using a specifically developed score based on the expression levels of these two miRNAs, ZAP-70 and LPL levels, Stamatopoulos et al. were even able to distinguish prognostic subgroups [29].

With the focus on miRNAs that correlate with 17p-deletion in CLL, one study was able to identify two miRNAs with prognostic relevance irrespective of other clinical-pathologic factors. Low miR-181b expression and high miR-21 expression were identified as poor prognostic features and significantly associated with OS and PFS [30]. Confirmation of the prognostic relevance of miR-181b came from a subsequent study that analyzed miRNA expression changes in patients with stable and progressive CLL. By investigating patient-matched and sequentially sampled leukemic cells, miR-181b was found to decrease over time only in samples derived from patients with progressive disease [37]. Putative targets of miR-181b include the myeloid cell leukemia sequence 1 gene (MCL-1) [37], a member of the BCL-2 family with anti-apoptotic function, and the pleomorphic adenoma gene 1 (PLAG1) oncogene [38]. Of note in this context is the observation that miR-181 together with miR-29c have previously been shown to be downregulated in cases with 11q deletion and found to inversely correlate with the TCL1-oncogene [39]. Interestingly, the miR-29 family also downregulates MCL1 [40], which itself is associated with unfavorable prognostic factors and disease course [41]. The example of concomitant downregulation of miR-181 and miR-29 highlights the functional synergism miRNAs can generate in pathogenic circumstances.

## 5.3 Follicular Lymphoma

Follicular lymphoma (FL) represents one of the most common non-Hodgkin lymphomas in the western world. Its incidence approximates 2.6 per 100,000 with an median age ranging between 60 and 70 years at diagnosis with a slight predominance in females: FL typically exhibits the t(14;18)(q32;q21) chromosomal translocation with subsequent proximity of the BCL2 gene on chromosome 18 and the immunoglobulin heavy chain gene locus which result in high levels of the anti-apoptotic protein BCL-2. Despite its frequency, literature covering the relevance of miRNAs in this disease is rare. Roehle et al. investigated the lymphoma specific expression signature in DLBCL, FL and non-neoplastic lymph nodes and developed a classification tree consisting of four miRNAs (miR-330, miR-210, miR-17-5p, and miR-106b) with which most cases were assigned to the correct entity [42].

Though FL usually shows a slow and indolent clinical course transformation to more aggressive DLBCL takes place in a considerable portion of cases [43, 44]. To detect transformation associated changes in miRNA expression levels, Lawrie et al. compared transformed DLBCL cases with de novo DLBCL and FL cases with subsequent transformation to cases without transformation at a median follow-up of 5 years.

Of note, prediction of transformation for FL cases was possible by utilizing six miRNAs (miR-223, miR-217, miR-222, miR-221, and let-7i and let-7b) and therefore highlights the potential as novel prognostic marker. In addition, de novo DLBCL and transformed cases were differentiated based on a 12 miRNA signature

[45]. Despite FL mostly evolves through its characteristic translocation, a subset of approximately 10 % of cases do not exhibit the t(14;18)(q32;q21) and mostly BCL2 is not expressed. Analysis of differences in miRNA levels of t(14;18)-positive and t(14;18)-negative FL identified 17 miRNAs to be downregulated in FL lacking t(14;18). Using a highly sophisticated approach, the authors succeeded to correlate five downregulated miRNAs (miR-16, miR-26a, miR-101, miR-29c, and miR-138) in the t(14;18)-negative FL subset with significant mRNA expression changes of their predicted targets. Validation of these miRNAs using qPCR confirmed the downregulation of miR-26a, miR-29c, miR-138, and most significantly miR-16. The investigation of putative miR-16 targets like CHEK1 and CDK6 by immunohistochemistry revealed an inverse correlation with the respective protein expression levels [46].

## 5.4 Mantle Cell Lymphoma

Mantle cell lymphomas (MCL) represent a highly aggressive form of lymphoma and account for 3–10 % of all non-Hodgkin lymphomas that manifest in advanced stage with predominance in male patients at higher age [47, 48]. Outcome with conventional therapies is poor and median survival ranges around 3–4 years. However, novel approaches with intensive regimen, combining immunochemotherapies followed by autologous stem-cell-transplantation currently lead the way to considerable improvement [49]. The central pathogenic event in MCL is t(11;14)(q13;q32) which links the cyclin D1 gene (CCND1) to the immunoglobulin heavy chain promoter and leads to consecutive overexpression of cyclin D1 with subsequent deregulation of the cell cycle control. In addition, cell-cycle regulators like BMI1, INK4a, ARF, CDK4, and RB1 and DNA-damage-pathway members like ATM, CHK2, or p53 are frequently affected by chromosomal alteration or mutation [50, 51].

Cyclin D1 mRNA transcripts without a full-length 3′ UTR are often detected in more aggressive MCL variants with a high proliferation rate and short clinical course [52, 53]. Highly interesting in this context is the role of miR-15/miR-16 which regulate Cyclin D1 expression by targeting its 3′ UTR. Truncation of the gene therefore renders MCL cases unresponsive to the posttranscriptional regulation of CCND1 [54, 55]. Moreover, other cell-cycle regulators like CDK6 have been identified as miRNA targets. CDK6 belongs to the family of cyclin-dependent protein kinases and, as the name suggests, acts together with Cyclin D1 (and CDK4) to accelerate proliferation. One profiling-study identified the miR-29 family as regulators of CDK6 and found that patients with significant downregulated miR-29 levels had a less favorable clinical course than those with higher expression of miR-29 [56]. Another study performed by Navarro et al., investigated the expression of 86 miRNAs that are located at commonly disrupted genomic regions in MCL. Two clusters characterized by different mutational status of the immunoglobulin genes, proliferation signature, and number of genomic alterations were detected based on

an unsupervised analysis of the generated miRNA expression profiles. Like in DLBCL, overexpression of miR-17-5p/miR-20a was found in aggressive tumors with high MYC levels [57]. Another profiling study, using 23 primary MCL samples and 8 MCL cell lines identified a MCL miRNA-signature consisting of 117 miRNAs when compared to that of 11 reactive lymphoid tissues and CD19þ/IgDþ/CD27 lymph node-sorted B cells, respectively. MiR-617, miR-370 and miR-654 were found among the most significantly upregulated miRNAs whereas miR-31, miR-148a and miR-27b ranged among the most significantly downregulated ones. A subsequent bioinformatic approach by combining regular gene expression and miRNA-profiles with target-prediction databases hinted to several overactive routes including CD40, NF-kB, and mitogen-activated protein kinase (MAPK) pathways. Moreover, the authors provide evidence that downregulation of miR-26a leads to NF-kB activation potentially by targeting MAP3K2. They show that induced expression of miR-26a leads to abrogation of the nuclear translocation of RelA. Similar to other studies on different cancer types, the authors here identify miR-20b as significantly associated with prognosis in MCL. Patients exhibiting low levels of miR-20b had a higher probability for longer survival than patients with high miR-20b levels [58].

## 5.5 Diffuse Large B-Cell Lymphoma

Diffuse large B-cell lymphoma (DLBCL) belongs to the most frequent aggressive B-cell neoplasms with an incidence of approx. 25,000 cases per year in the US [59]. Significant progress in treatment efficacy has been achieved through the addition of rituximab to CHOP-like regimen [60–63]. However, treatment outcome remains variable due to the heterogeneity DLBCL presents with regards to clinical aspects, biology and pathogenesis. By the use of gene-expression profiling it was possible to link the molecular phenotype of biologically distinct groups to the clinical presentation. DLBCL can be split in three major subtypes, namely, the germinal-center B-cell-like (GCB) DLBCL, the activated B-cell-like (ABC) DLBCL, and the primary mediastinal DLBCL (PMBCL) with survival rates of 59, 30, and 64 % after 5-years [64–66]. Molecular characteristics of ABC-DLBCL include overexpression of BCL2 and amplification of its locus, deletion of the INK4A–ARF locus, trisomy 3 with consecutive upregulation of FOXP1 and a constant activation of the nuclear factor (NF)-κB pathway. GCB-DLBCL show recurrent t(14;18), TP53 mutations, loss of PTEN and amplification of the oncogenic miR-17-92 cluster as well as the proto-oncogene REL. Similar to the ABC-subtype, PMBCL present with overactive NF-κB signaling, in addition PMBCL show frequent amplification of a chromosome region on 9p24 encoding JAK2 and loss or mutation of its suppressor SOCS1 [67, 68]. Extending the molecular characterization of DLBCL to the miRNA level helped to identify further pathogenic mechanisms. The lymphoma specific relevance of the BIC-locus transcript has been known since the end of the 1980s [69]. However, suggestions for the classification and a putative role as noncoding RNA

[70, 71] as well as the observation that miR-155 is encoded in the BIC-transcript were published several years later [72]. Consequently miR-155 has been classified as an oncogenic acting miRNA due to its frequent overexpression in a variety of B-cell neoplasms [72, 73]. Following experiments confirmed its oncogenic potential in transgenic mice that overexpress miR-155 and develop a pre-B-lymphoproliferative disease with consecutive progression to high-grade B-cell neoplasms [74]. Mechanistically, the downregulation of SHIP, a negative regulator of PI3K-signaling, and C/EBPβ have been attributed to the lymphoma mediating effects of this miRNA [75, 76]. MiR-155 shows higher levels in PMBCL and ABC- than in GCB-DLBCL which may be attributed to the constitutive activation of the (NF-κB) pathway in these DLBCL subtypes [66, 77, 78], beside AP1 and MYB, NF-κB has been identified as central regulator of miR-155 expression [79, 80] and sustained upregulation of miR-155 was found to happen in response to autocrine stimulation by the tumor necrosis factor α (TNFα) [76]. GCB-DLBCL have lower levels of miR-155, however, the GCB-type specific loss of PTEN [81, 82] or amplification of the miR-17-92 cluster [82] which targets PTEN [83] point to the observed importance of maintaining PI3K-signaling in DLBCL [84]. Overexpression of the miR-17-92 cluster in lymphomas [85] and a MYC driven overexpression in DLBCL has been confirmed in independent studies [82, 86]. By using an integrative approach through combining the results of miR specific array CGH and microarray based miRNA expression profiling, Li et al. were able to generate a detailed map of commonly disrupted miRNA-loci in the DLBCL genome [86]. Hierarchical clustering of the investigated miRNAs separated the analyzed DLBCLs in three subsets independent of the hitherto identified DLBCL subclasses but with respect to the transcriptional level of MYC. Generated subgroup profiles were associated with transcriptional levels of MYC, influenced by genomic abnormalities and showed significant overlap with the discriminating miRNA profiles of B-cell subsets [86]. The central role for malignant transformation has been shown in a mouse B-cell lymphoma model where enforced expression of the miR-17-92 cluster acted with MYC expression to accelerate tumor development [87]. This miRNA cluster was identified to regulate the cell cycle and inhibit apoptosis that takes place at a higher rate if the miR-17-92 cluster is not expressed. Essentially, this effect was attributed to miR-17 and miR-20a by targeting the E2F family members [88–90]. With respect to the seed sequence homology the miR-17-92 cluster consists of the families miR-17/miR-20a, miR-18a, miR-19a/1miR-19b, and miR-92a. The selective deletion or overexpression of single family members of this cluster helped to identify the miR-19 family as most relevant for MYC-induced lymphomagenesis with PTEN as its main target [91, 92]. Moreover the miR-17-92 cluster was found to target the cyclin-dependent kinase inhibitor CDKN1A/p21 with consecutive increase in cell growth and to regulate the proapoptotic protein Bim, leading to overexpression of BCL-2 [93]. Uncontrolled expression of the clusters miR-17-92/miR-106b-25 and/or other family members in the miR-106a-363 cluster may also guide malignant cells to escape from TGFβ-dependent cell cycle arrest and apoptosis as previously exemplified [94]. Since miR-155 targets SMAD5 and renders DLBCL resistant to growth inhibitory effects that are mediated through cytokines of the bone morphogenetic

protein (BMP) family and transforming growth factor (TGF)-β [95], the cooperative action of these miRNAs represents an exemplary model on how different molecular routes may converge in pathogenic effects. Clinical relevance was found by the observation that DLBCL cases with upregulation of miR-17-92 and its paralog miR-106a-363 had the worst overall survival (OS) in one study [86].

## 5.6 Primary CNS Lymphoma

Beside the frequent occurrence of DLBCL without involvement of the central nervous system, a rare subtype of lymphoma presents with isolated manifestation in the brain, spinal cord or related structures and is therefore called primary CNS lymphoma (PCNSL). The majority of these cases can be classified as DLBCL and show an invariably poor outcome [96]. Data on differential regulation of miRNAs in this lymphoma-subtype and a specific pathogenic role is scarce due to the rare occurrence and difficult sampling procedure. A recent study investigated the miRNA expression in 11 samples of PCNSL compared to 10 samples of nodal DLBCL. 18 miRNAs turned out to be differentially regulated. Upregulated miRNAs in PCNSL were associated with the Myc-pathway (miR-17-5p, miR-20a, miR-9), blocking of terminal B-cell differentiation (miR-9, miR-30b/c) and inflammatory cytokines (miR-155), whereas downregulated miRNAs were found to involve miRNAs with ascribed tumor-suppressor function (miR-199a, miR-214, miR-193b, miR-145). Amongst brain specific miRNAs, only miR-9 was found to be upregulated in PCNSL cases [97]. Another study analyzed the differential expression of 15 miRNAs (miR-15a, miR-15b, miR-16, miR-17-3p, miR-17-5p, miR-18a, miR-19a, miR-19b, miR-20a, miR-21, miR-92, miR-127, miR-155, miR-181a, and miR-221) in 19 nodal cases without extranodal dissemination, 9 cases of PCNSL, 11 cases of primary testicular and 11 cases of other primary extranodal DLBCL and identified a significantly higher expression level of miR-17-5p in the cases with CNS manifestation [98]. Both studies did not identify differential expression between germinal and non-germinal center DLBCL cases of any of the investigated miRNAs [97, 98].

Investigations on the prognostic value of miRNAs in DLBCL identified several miRNAs that may help to predict the clinical course. Montes-Moreno et al. retrospectively analyzed a series of 258 de novo cases of DLBCL treated with standard protocols (243 Patients of this cohort were treated with R-CHOP) and identified a set of 9 miRNAs with prognostic relevance. Seven of these miRNAs (miR-221, miR-222,miR-331, miR-451, miR-28, miR-151, and miR-148a) were identified with respect to the putative cellular origin as previously described [99]. Two additional, independent prognostic miRNAs were miR-93 and miR-491. The authors succeeded to identify a high-risk group of patients with a 2-year OS and a progression free survival (PFS) probability of <50 % by applying a combined model including the IPI score [100]. Lawrie et al. who evaluated three miRNAs (miR-21, miR-155, or miR-221) which were differentially expressed between ABC and GCB-DLBCL identified miR-21 as independent prognostic marker. In this

retrospective series of 49 de novo DLBCL, high miR-21 expression was associated with longer relapse free survival [101]. The specific evaluation of three miRNAs (miR-21, miR-155, and miR-222) in another study on 106 DLBCL cases uniformly treated with R-CHOP therapy found a correlation of high levels of miR-222 and shorter OS and PFS [102]. 11 miRNAs that had been previously identified as variably expressed in DLBCL were investigated in an independent study in patients uniformly treated with R-CHOP, high levels of miR-18a were correlated with shorter OS, high levels of miR-181a were associated with longer PFS and increase expression miR-222 with shorter PFS [103].

## 5.7 Burkitt's Lymphoma

Burkitt's Lymphoma (BL) represents a high-grade B-cell neoplasm belonging to the class of Non-Hodgkin's lymphoma. BL has a highly aggressive clinical phenotype and is one of the most rapidly dividing tumors in humans with an approximate doubling time of 24 h [104]. With respect to its epidemiological presentation BL is subdivided in three different groups, namely, endemic BL, HIV-associated BL, and sporadic BL. Epstein–Barr Virus genome (EBV) is found in nearly all cases with endemic BL with its geographic hot-spot found in equatorial Africa, whereas the minority of sporadic BL and less than half of the HIV-associated BL are tested positive for EBV. Endemic BL in equatorial Africa is prevalent in children of the younger age, has distinct predilection sites like the jaw and shows an incidence that is 50 times higher than in the US. Unlike endemic BL sporadic BL mostly manifests with abdominal bulks in young adults and, with an incidence of 1–2 %, belongs to the less frequent form of lymphoma in Western Europe and the United States. Mortality shows a close correlation with age, rising consistently from pediatric patients to older adults [105]. The current basis of BL treatment consists of high intensity, brief-duration regimens, with which 65–100 % of adults achieve a CR and 47–86 % of patients maintaining these remissions at least 1 year following therapy (reviewed in [106]). The combination of high intensity regimen with the monoclonal antibody Rituximab has increased response rates in BL patients [107]. The major pathogenic event in BL is the translocation of the MYC gene to the immunoglobulin (Ig) heavy-IgH t(8;14) or light-chain (Ig-κ, Ig-λ) t(2;8) or t(8;22) locus with subsequent overexpression of MYC [108]. Similar to mRNA-based gene expression profiling that can reliably classify classic BL and might even more reliably distinguish borderline cases [109], miRNA based approaches succeed to accurately classify cases with DLBCL or BL regardless to the sometimes overlapping morphology [110]. However, subtypes of BL show a rather homogenous miRNA profile and might therefore be more similar with respect to underlying pathogenic mechanisms than one would assume from the epidemiologic distribution [110]. Transcriptional activity of Myc is central to differences in miRNA profiles between BL and other lymphoma entities [110, 111]. Especially members of the miR-17-92 cluster were identified as targets of Myc with a significant overexpression in Myc driven

lymphomas [87]. In contrast, activation of Myc may also leads to a widespread repression of miRNAs with tumor suppressive function. MiR-22, miR-26a, miR-29c, miR-30e, miR-146a, let-7, miR-15a, miR-29a, miR-34a, miR-195, and miR-150 were identified in one study by microarray screens as specifically downregulated by Myc. Enforced expression of these miRNAs reduced the tumorigenic potential of lymphoma cells [112]. Myc-dependent deregulation of miR-26a, miR-181a, and miR-16 has been associated with altered cell proliferation and loss of miR-26a leads to overexpression of its targeted oncogene EZH2 [113]. In addition let-7a, which targets and suppresses Myc is similarly downregulated in this genetic entity and therefore represents a blocked autoregulatory mechanism for the control of Myc [114]. However, deregulated miRNA expression is not only induced on the basis of the transcriptional activity of Myc itself. The closer analysis of t(8;14) negative BL cases and translocation-positive BL provides evidence for a specific role of miR-9* in this context. Of note is the finding of significant downregulation of miR-9* in cases without Myc-translocation and a strong methylation of the miR-9-1 gene. Mechanistically, miR-9* can target E2F1 (which itself is able to induce Myc) and by this indirectly leads to changes in Myc expression levels [115].

## 5.8 Multiple Myeloma

Multiple myeloma is characterized by uncontrolled clonal expansion of malignant plasma cells. Usually, the clinically defined stages in the supposed sequential development of myeloma include monoclonal gammopathy of undetermined clinical significance (MGUS), consecutive progression to smoldering myeloma and at the end to symptomatic myeloma [116]. The approximate incidence per year is 5–6 cases per 100000 persons, with a median age of 70 years. Post-germinal-center B-cells with constant proliferation are considered as basis for further transformation triggered by genetic and microenvironmental changes [116, 117]. On the molecular level multiple myeloma is usually characterized by a complex karyotype. Recurrent genomic changes are regularly found and include hyperdiploidy, deletion of chromosome 13, gain of chromosome 1q, translocations with IgH on chromosome 14q, deletion of chromosome 17p and indicate variable clinical course [116, 118–120]. Translocations of 14q involve specific partner regions with consecutively deregulated genes on 11q13 (CCND1) [121, 122] and 6p21 (CCND3) [123], 4p16 (MMSET and FGFR3) [124–126], 16q23 (MAF) [126, 127], and 20q11 (MAFB) and are additionally used as subgroup discriminators [119]. Beside cytogenetic changes, NRAS and KRAS have been found mutated in approximately one-third of investigated MM cases [128] and differentiate between MM and MGUS to a certain extend as mutations are found only in 5 % in MGUS [116, 129, 130]. Profiling of miRNAs with respect to the underlying cytogenetic changes in MM identified expression patterns mainly linked with the major IGH translocations [131, 132]. Especially in cases with t(4;14) the miR-let-7e, miR-125a-5p, and miR-99b from the cluster region at 19q13.33 were specifically overexpressed. Allelic imbalances

and LOH was associated with changes in expression levels of miRNAs located in the affected regions and included miR-let-7b (22q13.31) and miR-140-3p (16q22) [132]. A global assessment aiming to detect differential miRNA expression between samples from MM and MGUS patients compared to healthy donors identified several discriminating miRNAs with specific regulatory function [133]. Most upregulated miRNAs in MGUS and MM samples involved miR-181a/b, cluster miR-106b-25, and miR-21. MiRNAs predominantly upregulated in MM involved miR-32 and the miR-17-92 cluster. Functional investigations on these miRNAs revealed a potential role in controlling p53 activity by targeting the p300-CBP-Associated Factor (PCAF). MiR-19a and b were almost exclusively upregulated in MM and shown to target SOCS-1 which suggests a functional interconnection to the Il-6R/STAT-3 pathway. Moreover the miR-17-92 cluster was confirmed as specific regulator of the pro-apoptotic gene BIM [133]. MiRNA profiles derived from samples of relapsed or refractory patients exhibit an overexpression of miR-222, miR-221, miR-382, miR-181a and b and decrease of miR-15a and miR-16. Functional investigation of miR-15a and miR-16 implicated these two miRNAs as regulators of growth and proliferation by interacting with BCL2 (which has been identified in CLL [18]) and AKT3, the ribosomal protein S6 and MAP kinases. Specific inhibition has further been demonstrated on the NF-kB pathway, probably by regulating TAB3 [134] which has previously been found to activate NF-kB signaling [135, 136]. Myeloma cell surrounding osteoclasts, bone marrow stromal cells or osteoblasts secrete numerous factors including IL-6 and MM cells with amplifications of chromosome 1q21 show overexpression of the IL-6 receptor [137]. Both mechanisms therefore build the ground for steady or enhanced upregulation of IL-6 dependent genes as was shown for Stat3 mediated and dependent miR-21 expression [138]. Increase of miR-21 levels in the absence of IL-6 significantly reduce apoptosis in myeloma cells [138]. Like miRNA expression can be influenced by factors derived from surrounding cells as reported for IL-6 [138], miRNAs themselves can shape the microenvironment by regulating the release of mediators like vascular endothelial growth factor (VEGF) [134] which induces neo-angiogenesis and again IL-6 secretion [139]. Overexpression of miR-15a and miR-16-1 can diminish VEGF release by direct interaction and decreases consecutive capillary formation in vivo and in vitro [134]. Increased angiogenesis has been shown to promote disease progression and to render MM cells more resistant to conventional therapeutic approaches [140–143]. Decrease or loss of miR-15a and miR-16-1, as found in 13q deleted cases, therefore confers advantage in MM and represents an example on how pathogenic axes can be interconnected on different levels. However, the exact role of miR-15a and miR-16-1 in MM currently remains unclear. While miR-15a has been found to be upregulated in newly diagnosed cases [133, 144], other authors report low levels of 15a in advanced stages [134]. Similar to MM with deletion 13q, cases with 17p deletion or TP53 mutation show a very poor outcome. Inactivation of the p53 pathway by deletion or mutation is usually found at advanced stages [128] though, p53 inactivation has been suggested in the context of MDM2 overexpression in MM [145, 146]. A recent study identified a p53 dependent miRNA feedback-loop that regulates the expression levels of MDM2 expression. As previously

shown for miR-34a (reviewed in [147]) the miRNA cluster miR-194-2-192 and miR-194-1-215 are direct p53 targets. Activation of p53 induces upregulation of these miRNAs with a subsequent downregulation of their mutual target MDM2 and finally leads to cell-cycle arrest or apoptosis in a p53-dependent manner [148]. Moreover the authors identify the miR-192 and miR-215 dependent inhibitory effect on the MDM2 mediated ubiquitination of IGF-1R and its consecutive influence on the IGF-1 dependent mobility and invasion of MM cells as previously described [148–150]. Of note is the observation that these miRNAs show different expression levels in plasma-cells (highest), samples derived from MGUS-patients and MM cells (lowest). Aberrant promoter methylation of the miR-194-2-192 cluster, which has been identified in MM cell lines, might serve in part as explanation for this deregulated miR-expression and consecutive clonal selection [148].

## 5.9 Hodgkin Lymphoma

Hodgkin Lymphoma (HL) is a hematologic malignancy with an approximate incidence in the western world of 3 new cases per 100,000 persons and year. Effectiveness of treatment has consistently increased over the last years and therapy with current protocols achieves 5-year survival rates for patients with early-stage Hodgkin's lymphoma of at least 90 %. With regard to this development, reduction of long-term complications affecting the heart or lung and prevention from secondary malignancies gains importance [151]. The main groups of HL consist of the frequently diagnosed classical Hodgkin's lymphoma (cHL) and the rare nodular lymphocyte-predominant Hodgkin's lymphoma (NLPHL), which account for 95 % and 5 % of all HL cases. The classification is based on differences in the morphology, specific aspects of lymphoma cells and cellular infiltration pattern. With the diagnosis of cHL, subclassification into nodular sclerosis, which accounts for the majority of cHL cases, mixed cellularity, lymphocyte depletion, and lymphocyte-rich HL is applied. The hallmark of HL is the typical presentation of Hodgkin and Reed–Sternberg (HRS) cells and the extensive surrounding of these cells with a reactive, inflammatory environment consisting of putatively nonmalignant B- and T-lymphocytes, eosinophils, and plasma cells [152].

Though research currently considers HRS cells as derived from germinal center (GC) or post-GC-B-cells, it seems controversial that HRS cells have lost most of their B-cell-specific gene expression [152]. HRS cells show constitutive activation of NF-κB, JAK–STAT, PI3K–AKT, ERK, AP1, and NOTCH1 signaling and several recurrent genetic lesions have been found to majorly involve members of the NF-κB and Jak–Stat signaling pathways [152]. Moreover, the *TNFAIP3* tumor suppressor gene, a negative regulator of NF-κB signaling, has been identified to be affected in up to 40 % by inactivating mutations [153, 154]. HRS cells are latently infected by Epstein–Barr virus (EBV) in about 40 %. The pathogenic relevance for this finding is attributed to two latent membrane proteins expressed by EBV, which are able to mimic active BCR and CD40 receptors [155, 156]. Revelations from miRNA

mediated pathogenic effects perfectly fit into the hitherto delineated pathogenic traits that have been identified in HL. Studying microdissected HRS cells from cHL patients revealed a specific miR expression pattern when compared to CD77+ GC-B-cells. Putative targets of these miRNAs (overexpressed: miR-20a, miR-21, miR-9, miR-155, miR-16, miR-140, miR-18a, miR-30b, miR-30a-5p, miR-196a, miR-374, miR-186 and downregulated: miR-520a, miR-614, miR-200a) include numerous members of the SOCS family [157] which might lead to inactivation of SOCS with consecutively activated JAK/STAT signaling as previously described [158]. In addition, this miRNA expression pattern includes multiple well known miRNAs from studies on HL [73, 159, 160] or other lymphoma entities and different cancers as previously outlined. Beside the activation of the JAK/STAT-signaling pathway other miRNA mediated mechanisms have been revealed that may have a role in HL. Amongst these is the regulation of PRDM1 that is centrally involved in the process of plasma cell differentiation and is a target of miR-9 which itself is overexpressed in HL [159]. In addition high levels were found for miR-155 [73] which have been attributed to its specific induction through NF-κB, based on the EBV mediated expression of LMP1 [161]. In contrast, the study of Navarro et al. could not identify miR-155 as significantly upregulated based on the EBV status. Ten miRNAs (miR-96, miR-128a, miR-128b, miR-129, and miR-205 (low levels), miR-28, miR-130b, miR-132, miR-140, and miR-330 (high levels)) were differentially expressed in EBV+ cHL compared with EBV− cHL [162].

Amongst the signature based on 25-miRNAs that could be used to differentiate between classic HL and reactive lymph nodes, miR-138 expression levels were found in relation with the Ann Arbor stage of investigated Hodgkin cases. Interestingly, chromosomal mapping locates differentially expressed miRNAs to regions of frequent loss or gain in cHL that in part explains the observed deregulation. Exemplarily, gains were found for 17q harboring miR-21, 2p and miR-216, 22q and miR-185 and 14q with miR-134. Losses are found for 4q with miR-302a, miR-302b, and miR-302c and 3p with miR-135a [162]. A subsequent analysis by the same group showed that low levels of miR-135a expression were associated with shorter disease-free survival and more frequent relapse in patients with cHL. Transfection experiments of cell lines with pre-miR-135 increased apoptosis and reduced proliferation. The specific evaluation of the predicted target JAK2 confirmed its downregulation when miR-135a was overexpressed and in addition revealed that JAK2 is coordinatively downregulated together with the antiapoptotic protein BCL-XL [163].

## 5.10 Chronic Myelogenous Leukemia

Chronic myelogenous leukemia (CML) is a clonal myeloproliferative disorder of the pluripotent stem cell and affects approximately 1–2 per 100,000 persons and year with a slight predominance in men at a median age around 65 years. Usually the disease manifests with a sudden onset of symptoms caused by progressive

splenomegaly, marrow hypercellularity and resulting anemia, thrombocytopenia, and leukocytosis. The pathogenic reason is cytogenetically identified by the presence of the Philadelphia (Ph) chromosome that is caused by the reciprocal translocation t(9;22)(q34;q11) in 90–95 % of CML patients.

The involved genes ABL1 and BCR are coupled to the fusion gene BCR–ABL with sustained kinase activation of ABL which leads to the uncontrolled expansion of the malignant clones. If untreated, the disease typically follows a biphasic or triphasic course with an initial chronic phase changing to the accelerated phase after an average of 5–5.5 years and finally to the so called blast crisis [164, 165].

Approaches using miRNA-profiling have integrated the miRNAs into the concepts of pathogenesis and disease progression of CML. Similar to other hematological malignancies the miR-17-92 cluster was identified as Myc dependent in CML and in addition has been found to be regulated by BCR–ABL. Modes of specific inhibition of BCR–ABL through either using Imatinib or RNA interference resulted in decreased expression of miRNAs encoded in the miR-17-92-cluster. Coordinate signaling through a BCR–ABL–MYC–miR-17-92 pathway has therefore been suggested for enhanced miRNA expression in early chronic phase in CML [166]. Similar to this, another independent study reported about the relevance of another miRNA involved in the regulation of the BCR–ABL pathway. Initially miR-203 was identified by characterizing the fragile region on the mouse chromosome 12 that harbored about 12 % of the known miRNAs and was mostly lost in γ-radiation-induced T cell lymphomas. Though, downregulation of this miRNA was caused in a considerable portion of cases not through deletion of the chromosomal region but through the hypermethylation of the corresponding promoter region of Ph+ malignancies like B-ALL and CML. Most significantly, the authors identified ABL as the specific target of miR-203 and showed that expression of miR-203 leads to the inhibition of proliferation in malignant cells [167]. Comparative analysis of cells from healthy donors and newly diagnosed CML patients identified miR-96 as overexpressed and miR-10a, miR-150, and miR-151 as selectively downregulated in CML samples. BCR–ABL independent downregulation of miR-10a was found to correlate with upregulation of the putative target gene upstream stimulatory factor 2 (USF2), which itself leads to increased cell growth upon overexpression [168].

Progenitors of CML blast crisis have been shown to lose their ability for differentiation by suppression of the transcription factor CEBPα, which controls myeloid differentiation. Interestingly CEBPα protein levels are regulated by hnRNP E2 through interacting with the UTR of CEBPα, correspondingly upregulation of hnRNP E2 diminishes CEBPα protein levels. Of note is the observation that changes in posttranscriptional gene regulation induced by hnRNP E2 have been found as central mechanism for the transition to blast crisis in CML [164, 169–171]. Investigations on the role of miRNAs this context identified miR-328 as specifically modulated through the MAPK-hnRNPE2 pathway with decreased miR-328 levels in blast crisis CML [169, 170]. Analysis of the molecular architecture of miR-328 uncovered a high similarity between this miRNA and the binding site for hnRNP E2 contained in the CEBPα mRNA region. As expected, functional investigation identified miR-328 as competitive target and therefore decreases hnRNP E2 binding and

association of hnRNP E2 to CEBPα, respectively. Reconstitution of miR-328 expression was found to recover the ability of maturation in BCR–ABL positive cells with subsequently higher rates of apoptosis [167].

## 5.11 Acute Lymphoblastic Leukemia

Acute lymphoblastic leukemia (ALL) is the most common childhood leukemia with a peak incidence at 2–5 years of age. Cure is a realistic goal, as ≥94 % of children have continuous disease-free survival for 5 years and appear cured [172]. In contrast, only 25 % of adults in the age group of 45–54 have continuous disease-free survival for 5 years [172]. Recurring genetic abnormalities with prognostic and therapeutic relevance involve hyperdiploidy, MLL and BCR–ABL translocations, HOX genes as well as PAX5 and IKZF1 [173]. The potential of miRNAs to distinguish between related hematological diseases was shown in the study by Mi et al., demonstrating that four miRNAs, including miR-223, miR-128a, miR-128b (miR-128a and miR-128b were later found to be identical) and let-7b were the most discriminatory between acute myeloid leukemia (AML) and ALL, regardless of leukemia subtype [174]. A combination of any two of these miRNAs could discriminate ALL from AML cases with an overall diagnostic accuracy of 97–99 %. However, it is not clear if this study was performed on solely on ALL and AML samples from adults. Comparing exclusively pediatric AML and ALL, Zhang et al. found miR-100, miR-125b and miR-335 more abundant in AML samples compared to the healthy donors [175]. Subsequently, Fulci et al. identified a three-miRNA signature of miR-148, miR-151, and miR-424 as discriminative of adult T-lineage versus B-lineage ALL [176]. Furthermore, the authors described a set of six miRNAs, miR-425-5p, miR-191, miR-146b, miR-128, miR-629, and miR-126, that distinguished between B-ALL subgroups harboring distinct molecular lesions such as BCR-ABL, MLL-AF4, and E2A-PBX1 fusions [176]. These findings were extended by Schotte et al., who quantified 397 miRNAs in pediatric precursor B-ALL patients, demonstrating that miRNA expression profiles vary between leukemic cells, normal bone marrow, and sorted $CD34^+$ cells [177]. Based on ALL cytogenetics, Schotte et al. were also able to identify characteristic miRNA expression signatures [177]. Differences were found for 11q23/MLL-rearranged precursor B-ALL cases, which exhibited a downregulation of miR-708 and increased levels of miR-196b as well as t(12;21)/TEL–AML1-positive precursor B-ALL cases, which displayed an upregulation of miR-383, miR-99a, miR-100, and miR-125b [177]. Signatures that associate with prognosis, include high expression of miR-33, miR-215, miR-369-5p, miR-496, miR-518d, and miR-599 for a worse outcome and high abundance of miRNAs such as miR-10a, miR-134, miR-214, miR-484, miR-572, miR-580, miR-624, and miR-627 with a more favorable prognosis in pediatric ALL [177]. Central nervous system (CNS) involvement is a common and prognostic relevant feature of ALL. Therefore, Zhang et al. have identified a blood miRNA signature in pediatric ALL complicated by central nervous system (CNS) relapse [175]. They found

significant upregulation of miR-7, miR-198, and miR-633 and a downregulation of miR-126, miR-345, miR-222, and miR-551a in ALL patients with CNS relapse versus non-CNS relapsed ALL. In contrast, Kaddar et al. found miR-16 to be of prognostic relevance for ALL patients [178]. The authors describe that high miR-16 levels were associated with hyperleukocytosis and poor cytogenetic subgroups. Disease-free survival (DFS) was shown to be significantly shorter for miR-16 levels above the 75th quartile in all analyzed B-cell ALL samples. Considering the prominent role of miR-16 in CLL, these findings might point towards a dual role of miR-16 in both diseases and even a shared mechanism in their pathogenesis.

## 5.12 Acute Myeloid Leukemia

The incidence of acute myeloid leukemia (AML) is 3–5 cases/100,000 [179]. Its prognosis and subtypes are mainly determined by cytogenetics and molecular genetics as reflected in the WHO classification [180]. Before the first comprehensive miRNA profiling studies in AML were published, the role of individual miRNAs has preferentially been studied based on their expression in normal hematopoiesis [181, 182]. Therefore, miR-223 became one of the most investigated miRNAs in myelopoiesis (and AML) due to its specific expression in differentiated myeloid cells [181]. Although, its role in the pathogenesis of AML is not clear, profiling miRNA expression in hematopoietic subpopulations as well as in a human APL cell line (NB4) upon differentiation with ATRA revealed miR-223 to be expressed at low levels in the stem cell compartment with increasing expression throughout myeloid differentiation [182, 183]. Lentiviral overexpression of miR-223 in an AML cell line as well as in AML patient samples induced myeloid differentiation [183, 184], demonstrating that changes in the miRNA transcriptome can promote reprogramming of AML cells.

Gain and loss of miR-223 was shown to have distinct effects, as genetic depletion of miR-223 led to a significant increase of myeloid progenitor cells as well as hyper-mature circulating neutrophils [185]. However, AML profiling studies did not connect miR-223 expression to a particular leukemia subtype [186–189 {Marcucci, 2008 #272}]. Contradicting the prevailing view that only one strand of the miRNA:miRNA* duplex is actively silencing genes, it was further shown that both strands miR-223 and miR-223* are functionally relevant in myeloid cells [190].

Multiple miRNA expression studies of AML patient samples have been performed using different methodological approaches and different patient subgroups [186–188, 191, 192]. Recently, Garzon and colleagues applied custom DNA microarrays to quantify miRNA expression levels in 240 AML patient samples with intermediate and poor cytogenetics. Based on this approach, miRNA signatures associated with 11q23 translocations`, trisomy 8 and FLT3 mutations (FLT3-ITD) were identified, demonstrating that cytogenetics drive miRNA profiles [191]. The same group further investigated the role of miRNAs in AML carrying NPM1 and FLT3-ITD mutations, the two most frequent molecular aberrations in AML [187]. A signature

distinguishing mutated *NPM1* from wildtype cases included the upregulation of miR-10a, miR-10b as well as let-7 and miR-29 family members. The correlation of the presence of FLT3-ITD and miR-155 upregulation was further confirmed by several works [187, 188, 191, 193–195] although FLT3 inhibitor studies showed that the upregulation of miR-155 was independent from FLT3 signaling [187]. However, only little overlap was found between the different miRNA signatures published for NPM1+ AML, which might depend on the applied technology, patient samples and the comparisons made within the study. For example, Jongen-Lavrencic et al. and Cammarata et al. used a similar multiplexing RT-PCR approach, whereas Garzon et al. profiled their AML samples with a custom made miRNA microarray [188, 191, 192]. In each study a different number of patients, ranging from 9 to 68 was analyzed and different comparisons were made: Garzon et al. profiled cNPM1+ and cNPM1− NK-AML, whereas Cammarata et al. compared their AML miRNA profiles to $CD34^+$ bone marrow cells. Jongen-Lavrencic et al. applied unsupervised ordering to create subgroups with similar expression patterns of miRNAs. A total of three miRNAs, miR-10a, miR-10b, and miR-9 were consistently deregulated in all three studies. This indicates that they might not only be important for the pathogenesis of cNPM1+ AML, but also reflect the minimum phenotype of this AML subgroup.

In addition, multiple studies exploring the expression of miRNAs by quantitative RT-PCR in AML patient cohorts could also associate miRNA expression patterns with cytogenetic and molecular subtypes [188, 196]. However, for a miRNA based prediction of AML subtypes the necessary number of miRNAs varied drastically. In the study of Jongen-Lavrencic et al. a class predictor of only ten miRNAs predicts AML with t(8;21) and a set of seven miRNAs AML with t(15;17). In contrast, a predictor comprising 72 miRNAs was necessary for AML with inv(16), thereby suggesting that not all cytogenetic aberrations might have a quite unique miRNA expression pattern and indeed some of the heterogeneity may stem from the fact that current subgroups may need to be further subdivided. In addition, miRNA expression levels might mainly be influenced by the differentiation stage of the leukemic cells, and thus for example inv(16) might be hard to distinguish form other inv(16)-negative AML cases with an identical morphology. In contrast to the complexity of AML subgroups, a signature of only two miRNAs (miR-128 and miR-223) is highly discriminate between AML and acute lymphoblastic leukemia [174].

Similarly, smaller genome-wide miRNA expression studies using bead-based miRNA profiling approaches, microarrays and quantitative RT-PCR confirmed miRNA expression patterns characteristic of cytogenetic subgroups such as t(15;17), t(11q23), t(8;21) and inv(16) [192, 197], as well as molecular subtypes like AML with *CEBPA* and *NPM1* mutations or deregulated *MN1* expression [198]. Interestingly, almost all studies pointed towards the deregulation of miRNAs located in the *HOX* gene cluster, including miR-10a/b and miR-196a/miR-196b, as well as miR-221, let-7 family members, miR-155, miR-29, miR-125b, miR-181 and members of the miR-17-92 cluster in AML. This suggests that a defined group of miRNAs, possibly associated with normal hematopoietic stem cells might be involved in leukemogenic processes such as impaired differentiation and increased self-renewal.

Being correlated with altered gene expression and cytogenetic and molecular genetic aberrations, miRNA expression signatures have been shown to also confer prognostic information. While the study of Garzon and colleagues could identified two miRNAs, miR-191 and miR-199a to be significantly correlated with overall (OS) and disease-free survival (DFS) [191], Sun et al. could show that miR-212 expression correlates with OS, DFS and relapse-free survival independent of cytogenetic subgroup in AML [199]. Within cytogenetically normally AML (CN-AML), Schwind et al. described an association of high miR-181a levels with favorable OS, especially CN-AML with FLT3-ITD and NPM1 wildtype [200]. The same group also found that the combination of high BAALC expression and a BAALC-hosted miRNA, miR-3151 identified CN-AML patient subset with a poor outcome [201]. However, in the future additional studies are needed to determine the impact of miRNAs as reliable biomarkers for diagnosis as well as prognosis in AML.

So far, only few studies investigated the role of miRNAs in leukemic stem cells (LSC) as well as leukemia development. Recently, Wong et al. showed that the miR-17-92 polycistron regulates LSC activity through p21 in a murine MLL model [202]. In a more direct approach, Han and colleagues showed that retroviral overexpression of miR-29a can induce AML in mice [203], and O'Connell and colleagues demonstrated similar findings through overexpression of miR-155 in primitive hematopoietic cells that led to a myeloproliferative syndrome [204]. In contrast to miRNAs that can function as proto-oncogenes, there also has been evidence that distinct miRNAs can function as tumor-suppressor in AML. For example, recently Garzon and colleagues highlighted the potential of miR-29b as tumor suppressor by inducing apoptosis and reducing tumorgenicity in a xenograft AML model [205]. Furthermore, miRNAs have also been shown to represent both targets and effectors of the epigenetic machinery. In accordance to other genes, miRNA expression can be affected by DNA promoter methylation and histone modifications. As mentioned above AML1-ETO can lead to heterochromatic silencing of the miR-223 genomic region and demethylation can restore miR-223 expression followed by differentiation of leukemic blasts [184]. On the other hand, miRNA expression can impact the epigenetic modifications as for example miR-29b targets DNA methyltransferases in AML [205].

## 5.13 Summary

Here, we show that our knowledge about miRNAs in aberrant hematopoiesis has dramatically advanced since their discovery in CLL. However, most of the studies are descriptive and the functional relevance for many potential oncogenic miRNAs is still questionable. Functional approaches, especially animal models might significantly add to the understanding how miRNAs contribute to the development of cancer, by revealing novel miRNAs, miRNA isoforms, mutations, and absolute sequence counts, thereby highlighting additional miRNAs that might serve as future therapeutic targets in hematological neoplasias.

## References


1. Dores GM, Anderson WF, Curtis RE, Landgren O, Ostroumova E, Bluhm EC, et al. Chronic lymphocytic leukaemia and small lymphocytic lymphoma: overview of the descriptive epidemiology. Br J Haematol. 2007;139(5):809–19.
2. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood. 2008;111(12):5446–56.
3. Hallek M, Fischer K, Fingerle-Rowson G, Fink AM, Busch R, Mayer J, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. Lancet. 2010;376(9747):1164–74.
4. Stilgenbauer S, Zenz T. Understanding and managing ultra high-risk chronic lymphocytic leukemia. Hematology Am Soc Hematol Educ Program. 2010;2010:481–8.
5. Zenz T, Gribben JG, Hallek M, Dohner H, Keating MJ, Stilgenbauer S. Risk categories and refractory CLL in the era of chemoimmunotherapy. Blood. 2012;119(18):4101–7.
6. Dohner H, Fischer K, Bentz M, Hansen K, Benner A, Cabot G, et al. p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. Blood. 1995;85(6):1580–9.
7. Schaffner C, Stilgenbauer S, Rappold GA, Dohner H, Lichter P. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. Blood. 1999;94(2):748–53.
8. Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest. 1998;102(8):1515–25.
9. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94(6):1848–54.
10. Tobin G, Thunberg U, Johnson A, Eriksson I, Soderberg O, Karlsson K, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood. 2003;101(12):4952–7.
11. Stamatopoulos K, Belessi C, Moreno C, Boudjograh M, Guida G, Smilevska T, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. Blood. 2007;109(1):259–70.
12. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med. 2000;343(26):1910–6.
13. Stilgenbauer S, Liebisch P, James MR, Schroder M, Schlegelberger B, Fischer K, et al. Molecular cytogenetic delineation of a novel critical genomic region in chromosome bands 11q22.3-923.1 in lymphoproliferative disorders. Proc Natl Acad Sci U S A. 1996;93(21):11837–41.
14. McBride OW, Merry D, Givol D. The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). Proc Natl Acad Sci U S A. 1986;83(1):130–4.
15. Isobe M, Emanuel BS, Givol D, Oren M, Croce CM. Localization of gene for human p53 tumour antigen to band 17p13. Nature. 1986;320(6057):84–5.
16. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524–9.
17. Calin GA, Cimmino A, Fabbri M, Ferracin M, Wojcik SE, Shimizu M, et al. MiR-15a and miR-16-1 cluster functions in human leukemia. Proc Natl Acad Sci U S A. 2008;105(13):5166–71.
18. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005;102(39):13944–9.

19. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell. 2010;17(1):28–40.
20. Lia M, Carette A, Tang H, Shen Q, Mo T, Bhagat G, et al. Functional dissection of the chromosome 13q14 tumor-suppressor locus using transgenic mouse lines. Blood. 2012;119(13): 2981–90.
21. Fulci V, Chiaretti S, Goldoni M, Azzalin G, Carucci N, Tavolaro S, et al. Quantitative technologies establish a novel microRNA profile of chronic lymphocytic leukemia. Blood. 2007; 109(11):4944–51.
22. Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, et al. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood. 2007;109(3):1202–10.
23. Garg R, Wierda W, Ferrajoli A, Abruzzo L, Pierce S, Lerner S, et al. The prognostic difference of monoallelic versus biallelic deletion of 13q in chronic lymphocytic leukemia. Cancer. 2012;118(14):3531–7.
24. Chena C, Avalos JS, Bezares RF, Arrossagaray G, Turdo K, Bistmans A, et al. Biallelic deletion 13q14.3 in patients with chronic lymphocytic leukemia: cytogenetic, FISH and clinical studies. Eur J Haematol. 2008;81(2):94–9.
25. Fabbri M, Bottoni A, Shimizu M, Spizzo R, Nicoloso MS, Rossi S, et al. Association of a microRNA/TP53 feedback circuitry with pathogenesis and outcome of B-cell chronic lymphocytic leukemia. JAMA. 2011;305(1):59–67.
26. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med. 2005;353(17):1793–801.
27. Sampath D, Liu C, Vasan K, Sulda M, Puduvalli VK, Wierda WG, et al. Histone deacetylases mediate the silencing of miR-15a, miR-16, and miR-29b in chronic lymphocytic leukemia. Blood. 2012;119(5):1162–72.
28. Visone R, Rassenti LZ, Veronese A, Taccioli C, Costinean S, Aguda BD, et al. Karyotype-specific microRNA signature in chronic lymphocytic leukemia. Blood. 2009;114(18): 3872–9.
29. Stamatopoulos B, Meuleman N, Haibe-Kains B, Saussoy P, Van Den Neste E, Michaux L, et al. microRNA-29c and microRNA-223 down-regulation has in vivo significance in chronic lymphocytic leukemia and improves disease risk stratification. Blood. 2009;113(21):5237–45.
30. Rossi S, Shimizu M, Barbarotto E, Nicoloso MS, Dimitri F, Sampath D, et al. microRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival. Blood. 2010;116(6):945–52.
31. Mraz M, Malinova K, Kotaskova J, Pavlova S, Tichy B, Malcikova J, et al. miR-34a, miR-29c and miR-17-5p are downregulated in CLL patients with TP53 abnormalities. Leukemia. 2009;23(6):1159–63.
32. Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH, et al. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. Mol Cell. 2007;26(5):745–52.
33. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, et al. A microRNA component of the p53 tumour suppressor network. Nature. 2007;447(7148):1130–4.
34. Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N, et al. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. Mol Cell. 2007;26(5):731–43.
35. Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, Dumitru CD, et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc Natl Acad Sci U S A. 2004;101(32):11755–60.
36. Marton S, Garcia MR, Robello C, Persson H, Trajtenberg F, Pritsch O, et al. Small RNAs analysis in CLL reveals a deregulation of miRNA expression and novel miRNA candidates of putative relevance in CLL pathogenesis. Leukemia. 2008;22(2):330–8.

37. Visone R, Veronese A, Rassenti LZ, Balatti V, Pearl DK, Acunzo M, et al. miR-181b is a biomarker of disease progression in chronic lymphocytic leukemia. Blood. 2011;118(11): 3072–9.
38. Pallasch CP, Patz M, Park YJ, Hagist S, Eggle D, Claus R, et al. miRNA deregulation by epigenetic silencing disrupts suppression of the oncogene PLAG1 in chronic lymphocytic leukemia. Blood. 2009;114(15):3255–64.
39. Pekarsky Y, Santanam U, Cimmino A, Palamarchuk A, Efanov A, Maximov V, et al. Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. Cancer Res. 2006;66(24):11590–3.
40. Mott JL, Kobayashi S, Bronk SF, Gores GJ. mir-29 regulates Mcl-1 protein expression and apoptosis. Oncogene. 2007;26(42):6133–40.
41. Pepper C, Lin TT, Pratt G, Hewamana S, Brennan P, Hiller L, et al. Mcl-1 expression has in vitro and in vivo significance in chronic lymphocytic leukemia and is associated with other poor prognostic markers. Blood. 2008;112(9):3807–17.
42. Roehle A, Hoefig KP, Repsilber D, Thorns C, Ziepert M, Wesche KO, et al. MicroRNA signatures characterize diffuse large B-cell lymphomas and follicular lymphomas. Br J Haematol. 2008;142(5):732–44.
43. Lossos IS, Levy R. Higher grade transformation of follicular lymphoma: phenotypic tumor progression associated with diverse genetic lesions. Semin Cancer Biol. 2003;13(3): 191–202.
44. Montoto S, Davies AJ, Matthews J, Calaminici M, Norton AJ, Amess J, et al. Risk and clinical implications of transformation of follicular lymphoma to diffuse large B-cell lymphoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(17):2426–33.
45. Lawrie CH, Chi J, Taylor S, Tramonti D, Ballabio E, Palazzo S, et al. Expression of microRNAs in diffuse large B cell lymphoma is associated with immunophenotype, survival and transformation from follicular lymphoma. J Cell Mol Med. 2009;13(7):1248–60.
46. Leich E, Zamo A, Horn H, Haralambieva E, Puppe B, Gascoyne RD, et al. MicroRNA profiles of t(14;18)-negative follicular lymphoma support a late germinal center B-cell phenotype. Blood. 2011;118(20):5550–8.
47. Anderson JR, Armitage JO, Weisenburger DD. Epidemiology of the non-Hodgkin's lymphomas: distributions of the major subtypes differ by geographic locations. Non-Hodgkin's Lymphoma Classification Project. Ann Oncol. 1998;9(7):717–20.
48. Zhou Y, Wang H, Fang W, Romaguer JE, Zhang Y, Delasalle KB, et al. Incidence trends of mantle cell lymphoma in the United States between 1992 and 2004. Cancer. 2008;113(4): 791–8.
49. Dreyling M, Hiddemann W. Current treatment standards and emerging strategies in mantle cell lymphoma. Hematology Am Soc Hematol Educ Program 2009: 542–51.
50. Jares P, Colomer D, Campo E. Genetic and molecular pathogenesis of mantle cell lymphoma: perspectives for new targeted therapeutics. Nat Rev Cancer. 2007;7(10):750–62.
51. Perez-Galan P, Dreyling M, Wiestner A. Mantle cell lymphoma: biology, pathogenesis, and the molecular basis of treatment in the genomic era. Blood. 2011;117(1):26–38.
52. Wiestner A, Tehrani M, Chiorazzi M, Wright G, Gibellini F, Nakayama K, et al. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. Blood. 2007;109(11): 4599–606.
53. Rosenwald A, Wright G, Wiestner A, Chan WC, Connors JM, Campo E, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell. 2003;3(2):185–97.
54. Chen RW, Bemis LT, Amato CM, Myint H, Tran H, Birks DK, et al. Truncation in CCND1 mRNA alters miR-16-1 regulation in mantle cell lymphoma. Blood. 2008;112(3):822–9.
55. Deshpande A, Pastore A, Deshpande AJ, Zimmermann Y, Hutter G, Weinkauf M, et al. 3'UTR mediated regulation of the cyclin D1 proto-oncogene. Cell Cycle. 2009;8(21): 3584–92.

56. Zhao JJ, Lin J, Lwin T, Yang H, Guo J, Kong W, et al. microRNA expression profile and identification of miR-29 as a prognostic marker and pathogenetic factor by targeting CDK6 in mantle cell lymphoma. Blood. 2010;115(13):2630–9.
57. Navarro A, Bea S, Fernandez V, Prieto M, Salaverria I, Jares P, et al. MicroRNA expression, chromosomal alterations, and immunoglobulin variable heavy chain hypermutations in Mantle cell lymphomas. Cancer Res. 2009;69(17):7071–8.
58. Di Lisio L, Gomez-Lopez G, Sanchez-Beato M, Gomez-Abad C, Rodriguez ME, Villuendas R, et al. Mantle cell lymphoma: transcriptional regulation by microRNAs. Leukemia. 2010;24(7):1335–42.
59. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. Blood. 1997;89(11):3909–18.
60. Coiffier B, Lepage E, Briere J, Herbrecht R, Tilly H, Bouabdallah R, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. N Engl J Med. 2002;346(4):235–42.
61. Feugier P, Van Hoof A, Sebban C, Solal-Celigny P, Bouabdallah R, Ferme C, et al. Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d'Etude des Lymphomes de l'Adulte. J Clin Oncol. 2005;23(18):4117–26.
62. Habermann TM, Weller EA, Morrison VA, Gascoyne RD, Cassileth PA, Cohn JB, et al. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. J Clin Oncol. 2006;24(19):3121–7.
63. Pfreundschuh M, Trumper L, Osterborg A, Pettengell R, Trneny M, Imrie K, et al. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. Lancet Oncol. 2006;7(5):379–91.
64. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403(6769):503–11.
65. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002;346(25):1937–47.
66. Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med. 2003;198(6):851–62.
67. Lenz G, Staudt LM. Aggressive lymphomas. N Engl J Med. 2010;362(15):1417–29.
68. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, et al. Stromal gene signatures in large-B-cell lymphomas. N Engl J Med. 2008;359(22):2313–23.
69. Clurman BE, Hayward WS. Multiple proto-oncogene activations in avian leukosis virus-induced lymphomas: evidence for stage-specific events. Mol Cell Biol. 1989;9(6):2657–64.
70. Tam W. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene. 2001;274(1–2):157–67.
71. Tam W, Ben-Yehuda D, Hayward WS. bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. Mol Cell Biol. 1997;17(3):1490–502.
72. Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, et al. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A. 2005;102(10):3627–32.
73. Kluiver J, Poppema S, de Jong D, Blokzijl T, Harms G, Jacobs S, et al. BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas. J Pathol. 2005;207(2):243–9.
74. Costinean S, Zanesi N, Pekarsky Y, Tili E, Volinia S, Heerema N, et al. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. Proc Natl Acad Sci U S A. 2006;103(18):7024–9.

75. Costinean S, Sandhu SK, Pedersen IM, Tili E, Trotta R, Perrotti D, et al. Src homology 2 domain-containing inositol-5-phosphatase and CCAAT enhancer-binding protein beta are targeted by miR-155 in B cells of Emicro-MiR-155 transgenic mice. Blood. 2009;114(7): 1374–82.
76. Pedersen IM, Otero D, Kao E, Miletic AV, Hother C, Ralfkiaer E, et al. Onco-miR-155 targets SHIP1 to promote TNFalpha-dependent growth of B cell lymphomas. EMBO Mol Med. 2009;1(5):288–95.
77. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. J Exp Med. 2001;194(12):1861–74.
78. Savage KJ, Monti S, Kutok JL, Cattoretti G, Neuberg D, De Leval L, et al. The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. Blood. 2003;102(12): 3871–9.
79. Vargova K, Curik N, Burda P, Basova P, Kulvait V, Pospisil V, et al. MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. Blood. 2011;117(14): 3816–25.
80. Yin Q, Wang X, McBride J, Fewell C, Flemington E. B-cell receptor activation induces BIC/ miR-155 expression through a conserved AP-1 element. J Biol Chem. 2008;283(5): 2654–62.
81. Siebert R, Gesk S, Harder S, Plotz S, Matthiesen P, Grote W, et al. Deletions in the long arm of chromosome 10 in lymphomas with t(14;18): a pathogenetic role of the tumor supressor genes PTEN/MMAC1 and MXI1? Blood. 1998;92(11):4487–9.
82. Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci U S A. 2008;105(36):13520–5.
83. Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol. 2008;9(4):405–14.
84. Kloo B, Nagel D, Pfeifer M, Grau M, Duwel M, Vincendeau M, et al. Critical role of PI3K signaling for NF-kappaB-dependent survival in a subset of activated B-cell-like diffuse large B-cell lymphoma cells. Proc Natl Acad Sci U S A. 2011;108(1):272–7.
85. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res. 2004;64(9):3087–95.
86. Li C, Kim SW, Rai D, Bolla AR, Adhvaryu S, Kinney MC, et al. Copy number abnormalities, MYC activity, and the genetic fingerprint of normal B cells mechanistically define the microRNA profile of diffuse large B-cell lymphoma. Blood. 2009;113(26):6681–90.
87. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043):828–33.
88. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839–43.
89. Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, et al. An E2F/miR-20a autoregulatory feedback loop. J Biol Chem. 2007;282(4):2135–43.
90. Woods K, Thomson JM, Hammond SM. Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. J Biol Chem. 2007;282(4):2130–4.
91. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, et al. Genetic dissection of the miR-17 92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes Dev. 2009; 23(24):2806–11.
92. Olive V, Bennett MJ, Walker JC, Ma C, Jiang I, Cordon-Cardo C, et al. miR-19 is a key oncogenic component of mir-17-92. Genes Dev. 2009;23(24):2839–49.
93. Inomata M, Tagawa H, Guo YM, Kameoka Y, Takahashi N, Sawada K. MicroRNA-17-92 down-regulates expression of distinct targets in different B-cell lymphoma subtypes. Blood. 2009;113(2):396–402.

94. Petrocca F, Vecchione A, Croce CM. Emerging role of miR-106b-25/miR-17-92 clusters in the control of transforming growth factor beta signaling. Cancer Res. 2008;68(20):8191–4.
95. Rai D, Kim SW, McKeller MR, Dahia PL, Aguiar RC. Targeting of SMAD5 links microRNA-155 to the TGF-beta pathway and lymphomagenesis. Proc Natl Acad Sci U S A. 2010;107(7):3111–6.
96. Ferreri AJ. How I, treat primary CNS lymphoma. Blood. 2011;118(3):510–22.
97. Fischer L, Hummel M, Korfel A, Lenze D, Joehrens K, Thiel E. Differential micro-RNA expression in primary CNS and nodal diffuse large B-cell lymphomas. Neuro Oncol. 2011;13(10):1090–8.
98. Robertus JL, Harms G, Blokzijl T, Booman M, de Jong D, van Imhoff G, et al. Specific expression of miR-17-5p and miR-127 in testicular and central nervous system diffuse large B-cell lymphoma. Mod Pathol. 2009;22(4):547–55.
99. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A. 2003;100(17):9991–6.
100. Montes-Moreno S, Martinez N, Sanchez-Espiridion B, Diaz Uriarte R, Rodriguez ME, Saez A, et al. miRNA expression in diffuse large B-cell lymphoma treated with chemoimmunotherapy. Blood. 2011;118(4):1034–40.
101. Lawrie CH, Soneji S, Marafioti T, Cooper CD, Palazzo S, Paterson JC, et al. MicroRNA expression distinguishes between germinal center B cell-like and activated B cell-like subtypes of diffuse large B cell lymphoma. Int J Cancer. 2007;121(5):1156–61.
102. Malumbres R, Sarosiek KA, Cubedo E, Ruiz JW, Jiang X, Gascoyne RD, et al. Differentiation stage-specific expression of microRNAs in B lymphocytes and diffuse large B-cell lymphomas. Blood. 2009;113(16):3754–64.
103. Alencar AJ, Malumbres R, Kozloski GA, Advani R, Talreja N, Chinichian S, et al. MicroRNAs are independent predictors of outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. Clin Cancer Res. 2011;17(12):4125–35.
104. de Leval L, Hasserjian RP. Diffuse large B-cell lymphomas and Burkitt lymphoma. Hematol Oncol Clin North Am. 2009;23(4):791–827.
105. Mbulaiteye SM, Anderson WF, Bhatia K, Rosenberg PS, Linet MS, Devesa SS. Trimodal age-specific incidence patterns for Burkitt lymphoma in the United States, 1973–2005. Int J Cancer. 2010;126(7):1732–9.
106. Blum KA, Lozanski G, Byrd JC. Adult Burkitt leukemia and lymphoma. Blood. 2004;104(10):3009–20.
107. Thomas DA, Faderl S, O'Brien S, Bueso-Ramos C, Cortes J, Garcia-Manero G, et al. Chemoimmunotherapy with hyper-CVAD plus rituximab for the treatment of adult Burkitt and Burkitt-type lymphoma or acute lymphoblastic leukemia. Cancer. 2006;106(7):1569–80.
108. Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, et al. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A. 1982;79(24):7837–41.
109. Dave SS, Fu K, Wright GW, Lam LT, Kluin P, Boerma EJ, et al. Molecular diagnosis of Burkitt's lymphoma. N Engl J Med. 2006;354(23):2431–42.
110. Lenze D, Leoncini L, Hummel M, Volinia S, Liu CG, Amato T, et al. The different epidemiologic subtypes of Burkitt lymphoma share a homogenous micro RNA profile distinct from diffuse large B-cell lymphoma. Leukemia. 2011;25(12):1869–76.
111. Robertus JL, Kluiver J, Weggemans C, Harms G, Reijmers RM, Swart Y, et al. MiRNA profiling in B non-Hodgkin lymphoma: a MYC-related miRNA profile characterizes Burkitt lymphoma. Br J Haematol. 2010;149(6):896–9.
112. Chang TC, Yu D, Lee YS, Wentzel EA, Arking DE, West KM, et al. Widespread microRNA repression by Myc contributes to tumorigenesis. Nat Genet. 2008;40(1):43–50.
113. Sander S, Bullinger L, Klapproth K, Fiedler K, Kestler HA, Barth TF, et al. MYC stimulates EZH2 expression by repression of its negative regulator miR-26a. Blood. 2008;112(10):4202–12.

114. Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res. 2007;67(20):9762–70.
115. Onnis A, De Falco G, Antonicelli G, Onorati M, Bellan C, Sherman O, et al. Alteration of microRNAs regulated by c-Myc in Burkitt lymphoma. PLoS One. 2010;5(9):e12960.
116. Kuehl WM, Bergsagel PL. Multiple myeloma: evolving genetic events and host interactions. Nat Rev Cancer. 2002;2(3):175–87.
117. Palumbo A, Anderson K. Multiple myeloma. N Engl J Med. 2011;364(11):1046–60.
118. Avet-Loiseau H, Attal M, Moreau P, Charbonnel C, Garban F, Hulin C, et al. Genetic abnormalities and survival in multiple myeloma: the experience of the Intergroupe Francophone du Myelome. Blood. 2007;109(8):3489–95.
119. Bergsagel PL, Kuehl WM. Molecular pathogenesis and a consequent classification of multiple myeloma. J Clin Oncol. 2005;23(26):6333–8.
120. Calasanz MJ, Cigudosa JC, Odero MD, Ferreira C, Ardanaz MT, Fraile A, et al. Cytogenetic analysis of 280 patients with multiple myeloma and related disorders: primary breakpoints and clinical correlations. Genes Chromosomes Cancer. 1997;18(2):84–93.
121. Chesi M, Bergsagel PL, Brents LA, Smith CM, Gerhard DS, Kuehl WM. Dysregulation of cyclin D1 by translocation into an IgH gamma switch region in two multiple myeloma cell lines. Blood. 1996;88(2):674–81.
122. Gabrea A, Bergsagel PL, Chesi M, Shou Y, Kuehl WM. Insertion of excised IgH switch sequences causes overexpression of cyclin D1 in a myeloma tumor cell. Mol Cell. 1999;3(1):119–23.
123. Shaughnessy Jr J, Gabrea A, Qi Y, Brents L, Zhan F, Tian E, et al. Cyclin D3 at 6p21 is dysregulated by recurrent chromosomal translocations to immunoglobulin loci in multiple myeloma. Blood. 2001;98(1):217–23.
124. Chesi M, Nardini E, Brents LA, Schrock E, Ried T, Kuehl WM, et al. Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. Nat Genet. 1997;16(3):260–4.
125. Chesi M, Nardini E, Lim RS, Smith KD, Kuehl WM, Bergsagel PL. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood. 1998;92(9):3025–34.
126. Chesi M, Bergsagel PL, Shonukan OO, Martelli ML, Brents LA, Chen T, et al. Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. Blood. 1998;91(12):4457–63.
127. Hurt EM, Wiestner A, Rosenwald A, Shaffer AL, Campo E, Grogan T, et al. Overexpression of c-maf is a frequent oncogenic event in multiple myeloma that promotes proliferation and pathological interactions with bone marrow stroma. Cancer Cell. 2004;5(2):191–9.
128. Chng WJ, Glebov O, Bergsagel PL, Kuehl WM. Genetic events in the pathogenesis of multiple myeloma. Best Pract Res Clin Haematol. 2007;20(4):571–96.
129. Corradini P, Ladetto M, Voena C, Palumbo A, Inghirami G, Knowles DM, et al. Mutational activation of N- and K-ras oncogenes in plasma cell dyscrasias. Blood. 1993;81(10):2708–13.
130. Bezieau S, Devilder MC, Avet-Loiseau H, Mellerin MP, Puthier D, Pennarun E, et al. High incidence of N and K-Ras activating mutations in multiple myeloma and primary plasma cell leukemia at diagnosis. Hum Mutat. 2001;18(3):212–24.
131. Chi J, Ballabio E, Chen XH, Kusec R, Taylor S, Hay D, et al. MicroRNA expression in multiple myeloma is associated with genetic subtype, isotype and survival. Biol Direct. 2011;6:23.
132. Lionetti M, Biasiolo M, Agnelli L, Todoerti K, Mosca L, Fabris S, et al. Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma. Blood. 2009;114(25):e20–6.
133. Pichiorri F, Suh SS, Ladetto M, Kuehl M, Palumbo T, Drandi D, et al. MicroRNAs regulate critical genes associated with multiple myeloma pathogenesis. Proc Natl Acad Sci U S A. 2008;105(35):12885–90.

134. Roccaro AM, Sacco A, Thompson B, Leleu X, Azab AK, Azab F, et al. MicroRNAs 15a and 16 regulate tumor proliferation in multiple myeloma. Blood. 2009;113(26):6669–80.
135. Kanayama A, Seth RB, Sun L, Ea CK, Hong M, Shaito A, et al. TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains. Mol Cell. 2004;15(4): 535–48.
136. Jin G, Klika A, Callahan M, Faga B, Danzig J, Jiang Z, et al. Identification of a human NF-kappaB-activating protein, TAB3. Proc Natl Acad Sci U S A. 2004;101(7):2028–33.
137. Hideshima T, Mitsiades C, Tonon G, Richardson PG, Anderson KC. Understanding multiple myeloma pathogenesis in the bone marrow to identify new therapeutic targets. Nat Rev Cancer. 2007;7(8):585–98.
138. Loffler D, Brocke-Heidrich K, Pfeifer G, Stocsits C, Hackermuller J, Kretzschmar AK, et al. Interleukin-6 dependent survival of multiple myeloma cells involves the Stat3-mediated induction of microRNA-21 through a highly conserved enhancer. Blood. 2007;110(4):1330–3.
139. Dankbar B, Padro T, Leo R, Feldmann B, Kropff M, Mesters RM, et al. Vascular endothelial growth factor and interleukin-6 in paracrine tumor-stromal cell interactions in multiple myeloma. Blood. 2000;95(8):2630–6.
140. Vacca A, Ribatti D, Roncali L, Ranieri G, Serio G, Silvestris F, et al. Bone marrow angiogenesis and progression in multiple myeloma. Br J Haematol. 1994;87(3):503–8.
141. Singhal S, Mehta J, Desikan R, Ayers D, Roberson P, Eddlemon P, et al. Antitumor activity of thalidomide in refractory multiple myeloma. N Engl J Med. 1999;341(21):1565–71.
142. Hayashi T, Hideshima T, Anderson KC. Novel therapies for multiple myeloma. Br J Haematol. 2003;120(1):10–7.
143. Hideshima T, Chauhan D, Podar K, Schlossman RL, Richardson P, Anderson KC. Novel therapies targeting the myeloma cell and its bone marrow microenvironment. Semin Oncol. 2001;28(6):607–12.
144. Zhou Y, Chen L, Barlogie B, Stephens O, Wu X, Williams DR, et al. High-risk myeloma is associated with global elevation of miRNAs and overexpression of EIF2C2/AGO2. Proc Natl Acad Sci U S A. 2010;107(17):7904–9.
145. Quesnel B, Preudhomme C, Oscier D, Lepelley P, Collyn-d'Hooghe M, Facon T, et al. Overexpression of the MDM2 gene is found in some cases of haematological malignancies. Br J Haematol. 1994;88(2):415–8.
146. Teoh G, Urashima M, Ogata A, Chauhan D, DeCaprio JA, Treon SP, et al. MDM2 protein overexpression promotes proliferation and survival of multiple myeloma cells. Blood. 1997;90(5):1982–92.
147. Hermeking H. The miR-34 family in cancer and apoptosis. Cell Death Differ. 2010;17(2): 193–9.
148. Pichiorri F, Suh SS, Rocci A, De Luca L, Taccioli C, Santhanam R, et al. Downregulation of p53-inducible microRNAs 192, 194, and 215 impairs the p53/MDM2 autoregulatory loop in multiple myeloma development. Cancer Cell. 2010;18(4):367–81.
149. Girnita L, Girnita A, Larsson O. Mdm2-dependent ubiquitination and degradation of the insulin-like growth factor 1 receptor. Proc Natl Acad Sci U S A. 2003;100(14):8247–52.
150. Froment P, Dupont J, Christophe-Marine J. Mdm2 exerts pro-apoptotic activities by antagonizing insulin-like growth factor-I-mediated survival. Cell Cycle. 2008;7(19):3098–103.
151. Armitage JO. Early-stage Hodgkin's lymphoma. N Engl J Med. 2010;363(7):653–62.
152. Kuppers R. Molecular biology of Hodgkin lymphoma. Hematology Am Soc Hematol Educ Program 2009: 491–6.
153. Kato M, Sanada M, Kato I, Sato Y, Takita J, Takeuchi K, et al. Frequent inactivation of A20 in B-cell lymphomas. Nature. 2009;459(7247):712–6.
154. Schmitz R, Hansmann ML, Bohle V, Martin-Subero JI, Hartmann S, Mechtersheimer G, et al. TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. J Exp Med. 2009;206(5):981–9.
155. Kilger E, Kieser A, Baumann M, Hammerschmidt W. Epstein-Barr virus-mediated B-cell proliferation is dependent upon latent membrane protein 1, which simulates an activated CD40 receptor. EMBO J. 1998;17(6):1700–9.

156. Mancao C, Hammerschmidt W. Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. Blood. 2007;110(10):3715–21.
157. Van Vlierberghe P, De Weer A, Mestdagh P, Feys T, De Preter K, De Paepe P, et al. Comparison of miRNA profiles of microdissected Hodgkin/Reed-Sternberg cells and Hodgkin cell lines versus CD77+ B-cells reveals a distinct subset of differentially expressed miRNAs. Br J Haematol. 2009;147(5):686–90.
158. Mottok A, Renne C, Willenbrock K, Hansmann ML, Brauninger A. Somatic hypermutation of SOCS1 in lymphocyte-predominant Hodgkin lymphoma is accompanied by high JAK2 expression and activation of STAT6. Blood. 2007;110(9):3387–90.
159. Nie K, Gomez M, Landgraf P, Garcia JF, Liu Y, Tan LH, et al. MicroRNA-mediated down-regulation of PRDM1/Blimp-1 in Hodgkin/Reed-Sternberg cells: a potential pathogenetic lesion in Hodgkin lymphomas. Am J Pathol. 2008;173(1):242–52.
160. Gibcus JH, Tan LP, Harms G, Schakel RN, de Jong D, Blokzijl T, et al. Hodgkin lymphoma cell lines are characterized by a specific miRNA expression profile. Neoplasia. 2009;11(2): 167–76.
161. Gatto G, Rossi A, Rossi D, Kroening S, Bonatti S, Mallardo M. Epstein-Barr virus latent membrane protein 1 trans-activates miR-155 transcription through the NF-kappaB pathway. Nucleic Acids Res. 2008;36(20):6608–19.
162. Navarro A, Gaya A, Martinez A, Urbano-Ispizua A, Pons A, Balague O, et al. MicroRNA expression profiling in classic Hodgkin lymphoma. Blood. 2008;111(5):2825–32.
163. Navarro A, Diaz T, Martinez A, Gaya A, Pons A, Gel B, et al. Regulation of JAK2 by miR-135a: prognostic impact in classic Hodgkin lymphoma. Blood. 2009;114(14):2945–51.
164. Melo JV, Barnes DJ. Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nat Rev Cancer. 2007;7(6):441–53.
165. Sawyers CL. Chronic myeloid leukemia. N Engl J Med. 1999;340(17):1330–40.
166. Venturini L, Battmer K, Castoldi M, Schultheis B, Hochhaus A, Muckenthaler MU, et al. Expression of the miR-17-92 polycistron in chronic myeloid leukemia (CML) CD34+ cells. Blood. 2007;109(10):4399–405.
167. Bueno MJ, Perez de Castro I, Gomez de Cedron M, Santos J, Calin GA, Cigudosa JC, et al. Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. Cancer Cell. 2008;13(6):496–506.
168. Agirre X, Jimenez-Velasco A, San Jose-Eneriz E, Garate L, Bandres E, Cordeu L, et al. Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34+ cells increases USF2-mediated cell growth. Mol Cancer Res. 2008;6(12):1830–40.
169. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat Genet. 2002;30(1): 48–58.
170. Perrotti D, Neviani P. From mRNA metabolism to cancer therapy: chronic myelogenous leukemia shows the way. Clin Cancer Res. 2007;13(6):1638–42.
171. Chang JS, Santhanam R, Trotta R, Neviani P, Eiring AM, Briercheck E, et al. High levels of the BCR/ABL oncoprotein are required for the MAPK-hnRNP-E2 dependent suppression of C/EBPalpha-driven myeloid differentiation. Blood. 2007;110(3):994–1003.
172. Gaynon PS. Childhood acute lymphoblastic leukaemia and relapse. Br J Haematol. 2005; 131(5):579–87.
173. Mullighan CG. Molecular genetics of B-precursor acute lymphoblastic leukemia. J Clin Invest. 2012;122(10):3407–15.
174. Mi S, Lu J, Sun M, Li Z, Zhang H, Neilly MB, et al. MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. Proc Natl Acad Sci U S A. 2007;104(50):19971–6.
175. Zhang H, Luo XQ, Zhang P, Huang LB, Zheng YS, Wu J, et al. MicroRNA patterns associated with clinical prognostic parameters and CNS relapse prediction in pediatric acute leukemia. PLoS One. 2009;4(11):e7826.
176. Fulci V, Colombo T, Chiaretti S, Messina M, Citarella F, Tavolaro S, et al. Characterization of B- and T-lineage acute lymphoblastic leukemia by integrated analysis of MicroRNA and mRNA expression profiles. Genes Chromosomes Cancer. 2009;48(12):1069–82.

177. Schotte D, Chau JC, Sylvester G, Liu G, Chen C, van der Velden VH, et al. Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia. Leukemia. 2009;23(2):313–22.
178. Kaddar T, Chien WW, Bertrand Y, Pages MP, Rouault JP, Salles G, et al. Prognostic value of miR-16 expression in childhood acute lymphoblastic leukemia relationships to normal and malignant lymphocyte proliferation. Leuk Res. 2009;33(9):1217–23.
179. Bloomfield CD, Marcucci G, Dohner K, Dohner H. Acute myeloid leukemia. Introduction. Semin Oncol. 2008;35(4):324–5.
180. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937–51.
181. Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science. 2004;303(5654):83–6.
182. Petriv OI, Kuchenbauer F, Delaney AD, Lecault V, White A, Kent D, et al. Comprehensive microRNA expression profiling of the hematopoietic hierarchy. Proc Natl Acad Sci U S A. 2010;107(35):15443–8.
183. Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, et al. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. Cell. 2005;123(5):819–31.
184. Fazi F, Racanicchi S, Zardo G, Starnes LM, Mancini M, Travaglini L, et al. Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/ETO oncoprotein. Cancer Cell. 2007;12(5):457–66.
185. Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, Kirak O, et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature. 2008;451(7182):1125–9.
186. Isken F, Steffen B, Merk S, Dugas M, Markus B, Tidow N, et al. Identification of acute myeloid leukaemia associated microRNA expression patterns. Br J Haematol. 2008;140(2):153–61.
187. Garzon R, Garofalo M, Martelli MP, Briesewitz R, Wang L, Fernandez-Cymering C, et al. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. Proc Natl Acad Sci U S A. 2008;105(10):3945–50.
188. Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. Blood. 2008;111(10):5078–85.
189. Debernardi S, Skoulakis S, Molloy G, Chaplin T, Dixon-McIver A, Young BD. MicroRNA miR-181a correlates with morphological sub-class of acute myeloid leukaemia and the expression of its target genes in global genome-wide analysis. Leukemia. 2007;21(5):912–6.
190. Kuchenbauer F, Mah SM, Heuser M, McPherson A, Ruschmann J, Rouhi A, et al. Comprehensive analysis of mammalian miRNA* species and their role in myeloid cells. Blood. 2011;118(12):3350–8.
191. Garzon R, Volinia S, Liu CG, Fernandez-Cymering C, Palumbo T, Pichiorri F, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. Blood. 2008;111(6):3183–9.
192. Cammarata G, Augugliaro L, Salemi D, Agueli C, La Rosa M, Dagnino L, et al. Differential expression of specific microRNA and their targets in acute myeloid leukemia. Am J Hematol. 2010;85(5):331–9.
193. Faraoni I, Laterza S, Ardiri D, Ciardi C, Fazi F, Lo-Coco F. MiR-424 and miR-155 deregulated expression in cytogenetically normal acute myeloid leukaemia: correlation with NPM1 and FLT3 mutation status. J Hematol Oncol. 2012;5(1):26.
194. Danen-van Oorschot AA, Kuipers JE, Arentsen-Peters S, Schotte D, de Haas V, Trka J, et al. Differentially expressed miRNAs in cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. Pediatr Blood Cancer. 2012;58(5):715–21.
195. Whitman SP, Maharry K, Radmacher MD, Becker H, Mrozek K, Margeson D, et al. FLT3 internal tandem duplication associates with adverse outcome and gene- and microRNA-

expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. Blood. 2010;116(18): 3622–6.
196. Zhu YD, Wang L, Sun C, Fan L, Zhu DX, Fang C, et al. Distinctive microRNA signature is associated with the diagnosis and prognosis of acute leukemia. Med Oncol. 2012;29(4): 2323–31.
197. Li Z, Lu J, Sun M, Mi S, Zhang H, Luo RT, et al. Distinct microRNA expression profiles in acute myeloid leukemia with common translocations. Proc Natl Acad Sci U S A. 2008; 105(40):15535–40.
198. Langer C, Marcucci G, Holland KB, Radmacher MD, Maharry K, Paschka P, et al. Prognostic importance of MN1 transcript levels, and biologic insights from MN1-associated gene and microRNA expression signatures in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol. 2009;27(19):3198–204.
199. Sun SM, Rockova V, Bullinger L, Dijkstra MK, Dohner H, Lowenberg B, et al. The prognostic relevance of miR-212 expression with survival in cytogenetically and molecularly heterogeneous AML. Leukemia. 2013;27(1):100–6.
200. Schwind S, Maharry K, Radmacher MD, Mrozek K, Holland KB, Margeson D, et al. Prognostic significance of expression of a single microRNA, miR-181a, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol. 2010;28(36):5257–64.
201. Eisfeld AK, Marcucci G, Maharry K, Schwind S, Radmacher MD, Nicolet D, et al. miR-3151 interplays with its host gene BAALC and independently affects outcome of patients with cytogenetically normal acute myeloid leukemia. Blood. 2012;120(2):249–58.
202. Wong P, Iwasaki M, Somervaille TC, Ficara F, Carico C, Arnold C, et al. The miR-17-92 microRNA polycistron regulates MLL leukemia stem cell potential by modulating p21 expression. Cancer Res. 2010;70(9):3833–42.
203. Han YC, Park CY, Bhagat G, Zhang J, Wang Y, Fan JB, et al. microRNA-29a induces aberrant self-renewal capacity in hematopoietic progenitors, biased myeloid development, and acute myeloid leukemia. J Exp Med. 2010;207(3):475–89.
204. O'Connell RM, Rao DS, Chaudhuri AA, Boldin MP, Taganov KD, Nicoll J, et al. Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder. J Exp Med. 2008;205(3):585–94.
205. Garzon R, Liu S, Fabbri M, Liu Z, Heaphy CE, Callegari E, et al. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. Blood. 2009;113(25): 6411–8.

# Chapter 6
# miRNAs as Cancer Biomarkers

**Edward R. Sauter**

**Abstract** In this chapter we discuss the attributes of a good cancer biomarker, and more specifically the attributes of a good micro (mi)RNA biomarker for cancer. In discussing miRNAs biomarkers in different cancers, we (1) provide evidence supporting the potential use of circulating (serum and plasma), urinary, salivary, and pleural effusion miRNAs as diagnostic markers, (2) discuss the advantages of their use, and (3) address challenges (e.g., validation of cancer biomarkers) and possible solutions in their assessment as biomarkers for clinical application.

**Keywords** Cancer • microRNAs • Biomarkers

## 6.1 Introduction

Early diagnosis is critical to optimize survival from cancer. Current cancer diagnostic procedures are usually invasive, unpleasant, and inconvenient, which limits their application [1, 2]. Therefore, the identification of noninvasive approaches for early cancer detection is urgently needed. RNA is known to exist both in cells and freely in body fluids including serum, plasma, saliva, tears, urine, amniotic fluid, colostrum, breast milk, nipple aspirate fluid, bronchial lavage, cerebrospinal fluid, peritoneal fluid, pleural fluid, and seminal fluid [3]. miRNAs have been observed in the body fluids of patients with cancer as well as in healthy controls. Because of the importance of miRNAs in carcinogenesis, body fluid miRNA biomarkers may prove useful for early cancer detection.

E.R. Sauter (✉)
Department of Surgery, University of North Dakota School of Medicine and Health Sciences, 501 N. Columbia Rd, Grand Forks, ND 58202, USA
e-mail: edward.sauter@med.und.edu

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_6,

## 6.2 Description of a Cancer Biomarker

A cancer biologic marker (biomarker) may be a molecular, cellular, tissue, or process-based alteration that predicts the presence of cancer using tissue, cells or body fluids [4]. The marker is present in or produced by a tumor and is differentially expressed (either at higher or lower levels) in tumor compared to benign samples. Perhaps the earliest identification of a cancer biomarker was reported in 1847 by Henry Bence-Jones who noted the presence of abnormal proteins in the urine of a patient with multiple myeloma [5, 6]. Although many biomarkers have been discovered since then, only a few are approved by the Food and Drug Administration (FDA) for cancer diagnosis or prognosis (Table 6.1).

## 6.3 Types and Uses of Cancer Biomarkers

Cancer biomarkers are cells or cell components, such as DNA, mRNA, proteins, metabolites, or processes such as apoptosis, angiogenesis or proliferation, that are useful in disease diagnosis, prognosis, or prediction of response to treatment [4]. Methylated DNA of genes found in cancer cells and circulating DNA in body fluids can also be used as potential biomarkers for clinical application in the risk assessment, diagnosis, and prognosis of cancer. The biomarkers are produced either by the tumor itself or by other tissues, in response to the presence of cancer or other associated conditions, e.g., inflammation. Potential uses of cancer markers include the following: (1) screening the general population, (2) differential diagnosis in symptomatic patients, (3) clinical staging of cancer, (4) evaluation of response to treatment, (5) assessment of disease recurrence through monitoring, (6) as prognostic indicators of disease progression, or (7) estimation of tumor volume [7].

## 6.4 FDA Approved Cancer Biomarkers

Several markers have been approved by the FDA for use as an aid in diagnosis, prognosis or management of cancer in individuals with risk factors or who have symptoms of cancer (Table 6.1). Such markers include breast cancer (e.g., cancer antigen (CA)15-3, CA27-29, cytokeratins, estrogen and progesterone receptor: ER and PR, heregulin (Her)-2/neu, mammaprint, oncotype DX); ovarian cancer marker (e.g., CA125), cervical cancer marker (e.g., pap smear); prostate cancer marker (e.g., prostate-specific antigen, PSA); testicular cancer marker (e.g., human chorionic gonadotropin (hCG)β, α-fetoprotein, AFP); germ-cell hepatoma marker (e.g., AFP); colon cancer marker (e.g., carcinoembryonic antigen: CEA, epidermal growth factor receptor); gastrointestinal stromal tumor marker (e.g., KIT); pancreatic cancer marker (e.g., CA19-9); thyroid cancer marker (e.g., thyroglobulin) and bladder cancer marker (e.g., chromosomes 3, 7, 9, 17; NMP22; fibrin/FDP; BTA, high molecular weight CEA and mucin).

**Table 6.1** FDA-approved biomarkers for various cancers

| S/N | Types of cancer | Name of cancer marker | Type of cancer marker | Source of cancer marker | Clinical use of cancer marker | References |
|---|---|---|---|---|---|---|
| I | Breast cancer | | | | | |
| 1 | | CA15-3 | Glycoprotein | Serum | Monitoring | [8, 9] |
| 2 | | CA27-29 | Glycoprotein | Serum | Monitoring | [70, 71] |
| 3 | | Cytokeratins | Protein (IHC) | Tumor | Prognosis | [70, 72] |
| 4 | | ER and PR | Protein (IHC) | Tumor | Selection for hormonal therapy | [73] |
| 5 | | HER2/neu | Protein | Serum | Monitoring | [73–75] |
| 6 | | HER2/neu | Protein | Tumor | Prognosis and selection for therapy | [73–75] |
| 7 | | HER2/neu | DNA (FISH) | Tumor | Prognosis and selection for therapy | [73–75] |
| 8 | | Mammaprint | Multigene | Tumor | Prognosis | [76, 77] |
| 9 | | Oncotype DX | Multigene | Tumor | Prognosis and selection for therapy | [78] |
| II | Ovarian cancer | | | | | |
| 10 | | CA125 | Glycoprotein | Serum | Monitoring | [14] |
| III | Cervical cancer | | | | | |
| 11 | | Pap smear | Cervical smear | Cervix | Screening | [15] |
| IV | Prostate cancer | | | | | |
| 12 | | PSA (total) | Protein | Serum | Screening and monitoring | [19–21] |
| 13 | | PSA (complex) | Protein | Serum | Screening and monitoring | [19–21] |
| 14 | | Free PSA (%) | Protein | Serum | Benign prostatic hyperplasia versus cancer diagnosis | [19–21] |
| V | Testicular cancer | | | | | |
| 15 | | hCGβ | Glycoprotein | Serum | Staging | [23, 24] |
| 16 | | AFP | Glycoprotein | Serum | Staging | [28–30, 79] |
| VI | Hepatocellular carcinoma | | | | | |

(continued)

**Table 6.1** (continued)

| S/N | Types of cancer | Name of cancer marker | Type of cancer marker | Source of cancer marker | Clinical use of cancer marker | References |
|---|---|---|---|---|---|---|
| 17 | | AFP | Glycoprotein | Serum | Staging Diagnosis and Monitoring therapy | [31–33] |
| VII | Colon cancer | | | | | |
| 18 | | CEA | Protein | Serum | Monitoring | [40, 42] |
| 19 | | EGFR | Protein | Colon | Selection of therapy | [43] |
| VIII | Gist | | | | | |
| 20 | | KIT | Protein (IHC) | Gastrointestinal tumor | Diagnosis and selection of therapy | [44] |
| IX | Pancreatic cancer | | | | | |
| 21 | | CA19-9 | Carbohydrate | Serum | Monitoring | [52, 53] |
| X | Thyroid cancer | | | | | |
| 22 | | Tg | Protein | Serum | Monitoring | [55, 56] |
| XI | Bladder cancer | | | | | |
| 23 | | Chromosomes 3, 7, 9 and 17 | DNA (FISH) | Urine | Screening and monitoring | [62] |
| 24 | | NMP22 | Protein | Urine | Screening and monitoring | [80] |
| 25 | | Fibrin/FDP | Protein | Urine | Monitoring | [68] |
| 26 | | BTA | Protein | Urine | Monitoring | [81, 82] |
| 27 | | High molecular weight CEA and mucin | Protein (Immunofluorescence) | Urine | Monitoring | [69] |

*IHC* immunohistochemistry

## 6.5 Examples of Biomarkers in Clinical Use

### *6.5.1 Breast Cancer: CA15-3, CA27-29*

The expression of CA15-3 and CA27-29 is highly associated with breast cancer prognosis [8, 9]. The two serum markers have been evaluated for diagnosis, but have not been found useful [10, 11]. Neither marker is specific to breast cancer, as they are expressed in other malignancies, and both can be found in patients with benign disorders of the liver, breast and kidney, as well as in patients with ovarian cysts [8].

### *6.5.2 Ovarian Cancer: CA125*

This antigen was first identified in 1981 by Bast and coworkers [12]. It is a high molecular weight glycoprotein which is elevated in approximately 90 % of patients with advanced epithelial ovarian cancer [13]. CA125 is more useful in following response to treatment than in diagnosis, as it can be elevated in benign conditions [14].

### *6.5.3 Cervical Cancer: Papanicolaou (Pap) Test, Liquid Based Cytology*

The Pap test represents the most cost-effective cancer screening test developed to date [15]. The test allows the diagnosis and treatment of precancerous lesions, thereby decreasing the incidence of cervical cancer [16]. The second major advance in cervical cancer screening was liquid-based cytology (LBC), which allows improved detection of abnormal cervical cells [17].

### *6.5.4 Prostate Cancer: PSA*

PSA is a glycoprotein present in the epithelial cells of the prostatic ducts and acini that is released into the bloodstream in men with prostate cancer [18]. PSA is useful in monitoring patients with advanced stages of prostatic cancer [19], and the pretreatment value of PSA in men with prostate cancer is prognostic of patient survival [20, 21].

### *6.5.5 Testicular Cancer: hCG and AFP*

hCG is expressed at very high concentrations by the placenta and is a tumor marker in trophoblastic diseases such as molar pregnancy, choriocarcinoma, and testicular

cancer [22]. Due to the fact that seminomas may produce only hCGß and not intact hCG, assays measuring hCG and hCGβ together are recommended to monitor testicular cancer [23, 24].

AFP is a normal fetal serum protein, initially produced by the yolk sac and later by the fetal liver and gastrointestinal tract during pregnancy [25]. Following birth, it clears rapidly from the circulation, falling to adult levels (less than 20 ng/ml in the serum), at 8–10 months of age [26, 27]. AFP is rarely elevated in healthy persons, and a rise is seen in only a few disease states, such as nonseminomatous germ cell tumors (NSGCT) [28–30] and hepatocellular carcinoma [31–33].

### *6.5.6 Hepatocellular Carcinoma (HCC): AFP*

A highly elevated level of AFP (greater than 500 ng/mL) in an HCC high incidence area and in the appropriate clinical setting is almost diagnostic of HCC, whereas modestly elevated levels of AFP (10–500 ng/mL) may also be found in benign nonmalignant chronic liver diseases [31, 34, 35]. This problem is compounded since most HCCs arise in patients with coexisting chronic liver disease [36, 37].

### *6.5.7 Colon Cancer: CEA, Epidermal Growth Factor Receptor (EGFR)*

CEA is a glycoprotein present in the fetal colon and colon adenocarcinoma but generally absent in a healthy adult colon [38, 39]. While CEA is expressed in certain healthy tissues including colon, stomach, tongue, esophagus, cervix, sweat glands, and the prostate [40], concentrations in colorectal cancer average 60-fold higher than in nonmalignant tissues [41]. Because of the presence of false positives, CEA is more beneficial in following individuals with colorectal cancer, especially in the early detection of cancer spread to the liver, rather than in cancer diagnosis [40, 42]. EGFR is one of the tumor markers that are overexpressed in colorectal cancer. Its expression is associated with advanced disease stage [43].

### *6.5.8 Gastrointestinal Stromal Tumor (GIST): c-KIT (CD117)*

c-KIT is an immunohistochemical marker used to aid in the diagnosis of GIST, the most common mesenchymal tumor in the gastrointestinal tract. Mesenchymal tumors arising in the digestive tract are rare compared with epithelial neoplasms. A high level of total c-KIT expression is a characteristic feature of GIST [44] and the immunohistochemical demonstration of KIT in addition to light microscopy histology has been used for the pathological diagnosis of GIST [45–48].

### 6.5.9 Pancreatic Cancer: CA19-9

CA19-9 (sialylated Lewis (a) antigen) is a tumor marker commonly associated with pancreatic adenocarcinoma [49–51]. Because of false positive and negative results in pancreatic cancer diagnosis [52], serum CA19-9 is used mainly for monitoring [53] individuals with known disease. High postoperative CA 19-9 levels have been associated with poor survival and may identify those patients who should receive alternative systemic therapy [53].

### 6.5.10 Thyroid Cancer: Thyroglobulin (Tg)

Tg is a prohormone of thyroxine (T4) and triiodothyronine (T3). The high specificity of Tg is based on the fact that thyroid (or thyroid cancer) cells are the only source of Tg in the human body [54]. Postoperative measurements of serum Tg provide information about the presence or absence of residual, recurrent, or metastatic disease in patients with differentiated thyroid cancer [55, 56].

### 6.5.11 Bladder Cancer

Fluorescence in situ hybridization (*FISH) test, nuclear matrix protein (NMP)22, fibrin/fibrinogen degradation products (FDP), bladder tumor antigen (BTA), ImmunoCyt.* UroVysion (Vysis-Abbot Laboratories, Downers Grove, IL, USA) is a commercially available FISH test [57] that detects aneuploidy in chromosomes 3, 7, 17, and loss of the 9p21 locus in malignant urothelial cells with a reported sensitivity of 73–92 % and a specificity of 89–96 % [58–63]. NMP22 is a noninvasive urine marker that uses a specific nuclear matrix protein to detect a bladder tumor via voided urine. The test had an overall sensitivity of 70–80 % for the detection of recurrent transitional cell carcinoma [64–66]. Martínez-Piñeiro et al. [67] found an increase in fibrin/FDP in 75 % of bladder cancer patients. The diagnostic ability of FDP could be superior to NMP22 [68]. The immunocytology test (ImmunoCyt/Ucyt+ test), [69] uses fluorescence immunohistochemistry with antibodies to a mucin glycoprotein and CEA.

## 6.6 Characteristics of an Ideal Cancer Biomarker

The ideal biomarker should: (1) be easily accessible such that it can be sampled relatively noninvasively, such as through the collection of a body fluid, and (2) be highly predictive (sensitive and specific) in the diagnosis or prognosis of a given

cancer(s) [7]. Unfortunately, to date there are no ideal biomarker(s) for any cancer [83, 84]. Nevertheless, cancer markers play an important role in disease detection and in assessing response to therapy in select groups of patients [85].

It is especially difficult to find an ideal biomarker for the early diagnosis of cancer, since most biomarkers found to be elevated in cancer are not cancer specific [86]. For this reason, there are more reliable biomarkers to follow individuals whose tumor expresses a given marker than there are biomarkers to identify cancers that have not been diagnosed. For example, two biomarkers found in the circulation, CEA and CA15–3, have shown promise in the management of cancers that express the proteins, but are not useful in early cancer detection, for many tumors do not express them, and benign processes can lead to their expression [87–89]. Even PSA, which is FDA approved to screen for prostate cancer, is often falsely elevated [90], and does not differentiate individuals with indolent disease, which may be best left untreated, from aggressive disease, which should be treated.

A promising area of biomarker research is in the evaluation of miRNAs, which are far more stable that messenger (m)RNA and therefore more readily detected in body fluids and fixed tissue. Tumor cells release miRNAs into body fluids [91] and miRNAs in body fluids are often altered in disease states including cancer [84, 91–94], suggesting their potential use to screen body fluids for differential expression of miRNAs for cancer risk assessment and early disease detection.

## 6.7 Attributes of miRNA Biomarker

### *6.7.1 Overview*

miRNAs are small noncoding RNAs that have shown potential as cancer biomarkers [95]. miRNAs are naturally occurring, evolutionarily conserved, single stranded molecules ~22 nucleotides in length which are generated by the RNase-III-type enzyme Dicer from an endogenous transcript that contains a local hairpin structure [96]. miRNAs regulate gene expression by either repressing the translation or causing degradation of mRNAs [96, 97].

### *6.7.2 Organization and Biogenesis*

miRNAs can be transcribed independently or from within their host genes, sharing the same promoter regions. The majority (about 70 %) of miRNA genes are intragenic (localized in introns or exons), the remainder being situated in intergenic regions as independent transcription units. miRNA genes represent about 1 % of the genome in different species, with about 30 % of the protein-coding genes in the human genome regulated by miRNAs [98].

### *6.7.3 Functions and Mechanisms of Mature miRNAs*

miRNAs regulate the expression of their target genes [99] by catalyzing the cleavage of mRNA [100–102] or repressing mRNA translation [103–105]. A notable characteristic of miRNA is that it only requires partial complementarity with its target gene sequence. Based on this fact, a miRNA could potentially target multiple genes. In addition, it is possible to predict which genes might be targeted by a miRNA based on the sequence characteristics of known miRNA–target pairs [106], and as such several hundred thousand miRNA–target pairs have been predicted in human, rat, and mouse [107, 108]. On the other hand, since only partial sequence complementarities are needed, it is more challenging to ascertain the presence of a miRNA–target pair [109]. Although several thousand has been predicted, only a few dozen miRNA–target pairs have been experimentally validated in human, rat, or mouse. Experimental approaches required to validate the presence and relevance of a microRNA–target pair are laborious and time-consuming.

More than 2000 miRNA sequences have been described in animals, plants, flies, worms and viruses. However, the specific functions of only a few of these miRNAs have been experimentally determined. For the few miRNAs whose function has been uncovered, they are important regulators of various aspects of biological processes in both plants and animals. miRNAs can play specific roles in diverse regulatory pathways, including the regulation of cell proliferation and metabolism [110, 111], tissue differentiation and maintenance [112], developmental timing [99, 111], cell death, hematopoiesis [113], neuron development [114], human carcinogenesis [115, 116], DNA methylation, and chromatin modification [117].

## 6.8 Evidence Supporting the Use of miRNAs as Cancer Biomarkers

The discovery of miRNAs in serum/plasma [118] has opened up the possibility of using miRNAs in body fluids to detect cancer. In addition to blood, urine, saliva, and pleural fluid have been found to contain miRNAs which are differentially expressed in individuals compared to healthy individuals (Patel and Sauter, 2011). Two hypotheses advanced to explain how miRNAs enter body fluids include: (1) passive release from tumor cells which undergo lysis as part of their rapid turnover, and (2) transfer from microvesicles shed by the cell membrane to a body fluid such as the blood [118].

Many studies suggest the potential use of miRNAs as body fluid biomarkers for cancer detection based on the following observations: (1) miRNA expression is frequently dysregulated in cancer [119, 120], (2) expression patterns of miRNAs in human cancer appear to be tissue-specific [121], and (3) miRNAs are stable in formalin-fixed tissues [122–124], and are present in cell-free form in body fluids [91].

Circulating miRNAs may be present only in healthy individuals, only in individuals with cancer, or both [92]. miRNA expression in matched serum and plasma are highly correlated, implying that both sample types are suitable for investigations of miRNAs as blood-based biomarkers [91]. The expression of specific circulating miRNAs is a good indicator of tumor miRNA burden [91].

## 6.9 miRNA Identification and Quantification Techniques

There are various methods used to detect circulating miRNA molecules in body fluids. Large scale screening for miRNAs can be performed with miRNA arrays. Northern blotting is currently considered the gold standard for miRNA detection [99], but is hampered by the difficulties of achieving stringent hybridization conditions, since these are very tiny molecules. Quantification of miRNA expression is generally performed by use of quantitative real-time PCR (qRT-PCR). A modified invader assay [125] and an RNA-primed, array-based Klenow enzyme assay (RAKE) [123] have also been described for quantification of miRNA. Lu et al. [126] reported the identification of simple and robust methods to detect, quantify, and distinguish miRNAs, particularly when cellular RNA is limited.

## 6.10 Advantages in the Use of miRNAs as Biomarkers

There are several advantages to using body fluid miRNAs as biomarkers, since miRNA expression is frequently altered in diseases or during damage to an organ and many miRNAs are highly tissue specific, stable, and can be used in high-throughput analysis. The availability of high-throughput experimental approaches for miRNA characterization and simple, universally applicable assays for quantitation (e.g., RT-qPCR) suggests that the discovery and validation of miRNA biomarkers will be more efficient than traditional proteomic biomarker discovery and validation, which usually encounter difficulties at the point of antibody generation.

## 6.11 miRNA Expression in Body Fluids of Patients with Different Human Cancers

Specific miRNAs have been found to have altered expression in malignant compared to benign tissue. Such tissue is generally available for individuals with newly diagnosed cancer, but less so for at risk individuals. Specimen collection for at risk individuals should be, if at all possible, noninvasive. miRNAs originating in cancer tissues are also transported outside the cell and remain stable in blood (serum or plasma), and other body fluids [127]. Numerous lines of evidence support miRNAs

as powerful biomarker candidates: miRNA pattern expression seems to be specific for different types of malignancies, their expression in body fluids, including serum, can be easily detected by PCR-related assays, and, compared to other biomarkers, miRNAs are remarkably stable and therefore could be reproducibly extracted and assayed in bodily fluids [91]. The detection of miRNAs in the systemic circulation and other body fluids suggests the potential usefulness of bodily fluid miRNA assessment for noninvasive cancer screening, diagnosis, prognosis, and therapeutics (Table 6.2) [128].

## *6.11.1 miRNA Expression in Body Fluids of Patients with Breast Cancer*

### 6.11.1.1 miRNA Expression in Blood

A study by Zhu et al. [129] on circulating microRNAs in breast cancer and healthy subjects showed that miR-155 was overexpressed in the serum of women with PR positive compared to women with PR negative disease. In another study, it is found that circulating miR-195 and let-7a [130] expression were increased in women with breast cancer compared to controls. A separate study by Asaga et al. [131] revealed the development of a direct blood PCR assay for detection of circulating miR-21 and its ability to assess early stage breast cancer in the serum of women. Previous studies have shown that expression of miR-425 was altered in cancer cell lines and tumor tissues [161, 162], and was significantly increased in the serum of breast cancer patients compared to healthy controls [132]. In addition, Zhao et al. [132] showed that not only upregulated but also downregulated miRNAs can discriminate patients with breast cancer from healthy controls with reasonable sensitivity and specificity. Specifically, it was reported that the levels of circulating let-7c were significantly higher in healthy controls than in women with breast cancer. This finding is consistent with the notion that let-7 family of miRNA acts as tumor suppressor genes in breast tumors [132]. It has been shown that advanced stage breast cancer patients had significantly more miR-34a in their blood than patients with early stage disease, and further demonstrated that changes in serum levels of miR-10b, miR-34a and miR-155 correlated with the presence of metastases [163].

## *6.11.2 miRNA Expression in Body Fluids of Patients with Colorectal Cancer*

### 6.11.2.1 miRNA Expression in Blood

Studies have shown that the expression of some miRNAs are significantly upregulated or downregulated in the plasma of colorectal cancer (CRC) patients. miR-92 expression is increased in both tissue and plasma of subjects with colorectal cancer

**Table 6.2** miRNA expression in body fluids

| miRNA category | Body fluid | Cancer type | miRNA type | Reference |
|---|---|---|---|---|
| Diagnostic miRNAs | Increased expression in cancer | | | |
| | Blood | Breast | miR-155 | [129] |
| | | | miR-195, let-7a | [130] |
| | | | miR-21 | [131] |
| | | | miR-302b, miR-425 | [132] |
| | | Colorectal | miR-92, miR-17-3p | [133] |
| | | | miR-29, miR-92 | [134] |
| | | Gastric | miR-17-5p, miR-21, miR-106a, miR-106b | [135] |
| | | | miR-32, miR-182, miR-143 | [136] |
| | | HCC | miR-500 | [137] |
| | | | miR-16, miR-195, miR-199a | [138] |
| | | Lung | miR-25, miR-223 | [92] |
| | | | miR-21, miR-126, miR-486-5p | [139] |
| | | | miR-155, miR-197, miR-182 | [140] |
| | | Ovarian | miR-21, -141, -200a, -200b, -200c, -203, -205, and -214 | [141] |
| | | | miR-21, -92, -93, -126, -29a, -155, -127, and -99b | [142] |
| | | Pancreas | miR-200a, miR-200b | [143] |
| | | | miR-210 | [144] |
| | | | miR-155 | [145] |
| | | | miR-18a | [146] |
| | | Prostate | miR-141 | [91] |
| | | OSCC | miR-31 | [91] |
| | | SSC of tongue | miR-184 | [147] |
| | | ESCC | miR-10a, miR-22, miR-100, miR-148b, miR-223, miR-133a, miR-127-3p | [148] |
| | | Melanoma | let-7c | [149] |
| | | Rhabdomyosarcoma | miR-206 | [150] |
| | Sputum | Lung | miR-21 | [151] |
| | | | miR-21, miR-182, miR-375, miR-200b | [152] |

(continued)

**Table 6.2** (continued)

| miRNA category | Body fluid | Cancer type | miRNA type | Reference |
|---|---|---|---|---|
| | | | miR-205, miR-210 miR-708 | [153] |
| | Pleural effusion | Lung | miR-24, miR-26a, miR-30d | [154] |
| | | Gastric | miR-24, miR-26a, miR-30d | [154] |
| | Urine | Bladder | miR-126, miR-182, miR-199a | [155] |
| | Saliva | OSCC | miR-31 | [156] |
| | Decreased expression in cancer | | | |
| | Blood | Breast | Let -7c | [132] |
| | | Gastric | let-7a | [135] |
| | | Melanoma | miR-125 | [149] |
| | Sputum | Lung | miR-486, miR-126, miR-145 | [152] |
| | Saliva | OSCC | miR-200a, miR-125a | [157] |
| | Lower ratio in cancer | | | |
| | Blood | HCC | miR-92a/638 | [158] |
| | Urine | Bladder | miR-126/152, miR-182/152 | [155] |
| Prognostic miRNAs | Blood | Lung | miR-1, miR-30d, miR-486, miR-499 | [159] |
| | | Prostate | miR-375, miR-141 | [160] |
| Predictive miRNAs | Blood | DLBCL | miR-21, miR-155, miR-210 | [93] |

*DLBCL* diffuse large B-cell lymphoma, *HCC* hepatocellular carcinoma, *OSCC* oral squamous cell carcinoma, *SSC* of tongue Squamous cell carcinoma of tongue, *ESCC* esophageal squamous cell carcinoma

compared to healthy controls [133]. Levels of miR-92 and miR-17-3p are increased in the plasma of subjects with colorectal cancer compared to subjects with other gastric cancer, inflammatory bowel disease and healthy controls [133]. In a separate study, plasma miR-29a and -92 were upregulated in individuals with adenomas (precursor lesions) and in those with colorectal cancer (CRC) [134]. These observations suggest that miR-92 and perhaps miR-17-3 and -29a, have potential as biomarkers for the detection of colorectal cancer.

### *6.11.3 miRNA Expression in Body Fluids of Patients with Gastric Cancer*

#### 6.11.3.1 miRNA Expression in Blood

Plasma concentrations of miR-17-5p, miR-21, miR-106a, and miR-106b were upregulated in patients with gastric cancer in comparison to healthy controls [135]. The expression levels of miRNAs miR-21 and miR-106b decreased after surgical removal of the disease, confirming the tumor release of these circulating miRNAs [135]. The researchers also discovered that the concentration of let-7a in plasma was significantly lower in gastric cancer patients than in controls, contrary to their expectation [135]. Li et al [136] showed that miR-32, -182 and -143 may be potential diagnostic biomarkers for intestinal-type gastric cancers.

#### 6.11.3.2 miRNA Expression in Pleural Effusions

Xie et al. [154] demonstrated that miR-24, -26a, and -30d were elevated in the cell free pleural fluid of individuals with lung or gastric cancer who had malignant effusions when compared to individuals without cancer who had benign effusions.

### *6.11.4 miRNA Expression in Body Fluids of Patients with Liver Cancer*

#### 6.11.4.1 miRNA Expression in Blood

Elevated levels of miR-500 [137] were observed in the sera of HCC patients, and its levels in sera returned to normal after surgical removal of the tumor. Qu et al. [138] reported the use of serum miR-16, -195, and -199a, alone or in combination with conventional serum markers, to differentiate HCC from chronic liver disease. Levels of miR-885-5p were significantly higher [164] in the sera of patients with HCC, liver cirrhosis and chronic hepatitis B than in healthy controls or gastric cancer patients. The ratio of miR-92a/miR-638 was found to be lower in individuals with HCC than in healthy controls [158].

### *6.11.5 miRNA Expression in Body Fluids of Patients with Lung Cancer*

#### 6.11.5.1 miRNA Expression in Blood

Bianchi et al. [165] reported the development of an miRNA diagnostic test in serum to identify asymptomatic high-risk individuals with early stage lung cancer. Zheng et al.

[140] found that circulating miR-155, -197, and -182 can be potential noninvasive biomarkers for early detection of lung cancer. By using both Solexa sequencing and quantitative reverse transcription (qRT)-PCR, it was determined that miR-25 and -223 were highly expressed in lung cancer compared to normal sera [92]. Shen et al. [139] reported that plasma miR-21, -126, and -486-5p are potential circulating biomarkers for the noninvasive diagnosis of lung cancer among individuals with solitary pulmonary nodules.

#### 6.11.5.2 miRNA Expression in Sputum

Xie et al. [151] found that miR-21 overexpression in the sputum of individuals with lung cancer. In a separate study, Yu et al. [152] found that miR-486, -126, and -145 were expressed at lower levels, whereas miR-21, miR-182, -375, and -200b displayed higher expression in cancer patients' sputum compared to sputum of cancer-free individuals. Three (miR-205, -210, and -708) miRNAs [153] used in combination best distinguished lung squamous cell carcinoma patients from normal subjects (73 % sensitivity and 96 % specificity).

#### 6.11.5.3 miRNA Expression in Pleural Effusions

Xie et al. [154] reported that miR-24,- 26a and -30d were increased in the cell free pleural fluid of individuals with lung or gastric cancer who had malignant effusions in comparison to individuals without cancer who had benign effusions. Eleven serum miRNAs were altered more than fivefold between in longer vs. shorter survival groups with NSCLC [159], and levels of four miRNAs (i.e., miR-486, -30d, -1, and -499) were associated with overall survival [159].

### *6.11.6 miRNA Expression in Body Fluids of Patients with Ovarian Cancer*

#### 6.11.6.1 miRNA Expression in Blood

Eight blood miRs (-21, -141, -200a, -200b, -200c, -203, -205, and -214) were reported to be markers of ovarian cancer [141]. Serum miR-21, -92, -93, -126, -29a, -155, -127, and -99b were found to be altered in women with ovarian cancer who had normal CA-125 levels [142].

## *6.11.7 miRNA Expression in Body Fluids of Patients with Pancreatic Cancer*

### 6.11.7.1 miRNA Expression in Blood

Increased levels of miR-200a and miR-200b in the serum of most patients with pancreatic cancer suggest their use as diagnostic pancreatic cancer biomarkers [143]. Plasma miR-210 expression from patients with newly diagnosed locally advanced pancreatic adenocarcinomas was highly increased compared to age-matched controls [144]. Wang et al. [18] identified plasma miR-155 as a candidate biomarker of early pancreatic neoplasia, whereas increased expression of miR196a correlated with disease progression. In a recent study, Morimura et al. [146] found that plasma concentrations of miR-18a were significantly higher in pancreatic cancer patients than in controls.

## *6.11.8 miRNA Expression in Body Fluids of Patients with Prostate Cancer*

### 6.11.8.1 miRNA Expression in Blood

miRNAs are stable and quantifiable in human plasma [91, 92]. Mitchell and coworkers [91] showed that miRNAs originating from human prostate cancer xenografts enter the circulation and are easily measured in plasma. Serum levels of miR-141 distinguished patients with prostate cancer from healthy controls. Brase et al. [160] demonstrated that expression levels of circulating miRNAs-375 and -141, are enhanced in prostate cancer and correlate with tumor progression.

## *6.11.9 miRNA Expression in Body Fluids of Patients with Oral Squamous Cell Cancer (OSCC)*

### 6.11.9.1 miRNA Expression in Blood

Based on findings of increased miR-31 expression in OSCC tissue, the plasma of 43 individuals with OSCC and 21 controls was evaluated [166]. miR-31 expression was elevated in those with cancer, providing a receiver operating characteristic curve (ROC) area of 0.82.

#### 6.11.9.2 miRNA Expression In Saliva

Two miRNAs, -125a and -200a, were present in significantly lower levels in the saliva of OSCC patients than in controls [157]. Salivary levels of miR-31 were higher in subjects with OSCC compared to controls [156] and in OSCC compared to leukoplakia. They also found that that miR-31 was concentrated in saliva compared to plasma, which suggests that salivary assessment of miR-31 was preferred. After excision of oral carcinoma, salivary levels of miR-31 markedly decreased, suggesting salivary miR-31 as a biomarker for early detection and postoperative follow-up of OSCC [156].

### *6.11.10 miRNA Expression in Body Fluids of Patients with SCC of Tongue*

#### 6.11.10.1 miRNA Expression in Blood

Plasma miR-184 levels were significantly higher in the tissue of patients with SCC compared to healthy individuals, and miR-184 levels significantly decreased after surgical treatment [147].

### *6.11.11 miRNA Expression in Body Fluids of Patients with Esophageal SCC*

#### 6.11.11.1 miRNA Expression in Blood

A panel of 7 serum miRNAs (miR-10a, -22, -100, -148b, -223, -133a, and -127-3p) were identified as blood-based biomarkers for the diagnosis of esophageal SCC [148].

### *6.11.12 miRNA Expression in Body Fluids of Patients with Bladder Cancer (BC)*

#### 6.11.12.1 miRNA Expression in Urine

In subjects with bladder cancer, urine levels of miR-126, -182, and -199a were more abundant than in controls [155], and the ratio of miR-126 and miR-182 in urine samples could be used to detect bladder cancer. Particularly, the researchers demonstrated that the ratio of miR-126/miR-152 and miR-182/miR-152 was higher in individuals with bladder cancer compared to either controls or individuals with urinary tract infections [155].

## *6.11.13 miRNA Expression in Body Fluids of Patients with Melanoma*

### 6.11.13.1 miRNA Expression in Blood

In one study, about 900 human miRNAs were screened to detect miRNAs that are deregulated in their expression in the blood of melanoma patients, using a microarray based approach. The researchers found that 21 miRNAs were downregulated and 30 upregulated, concluding that miRNA expression signatures are useful biomarkers for melanoma [167]. Holst et al. [149] investigated miRNAs that would better differentiate atypical nevi (AN) from common acquired nevi (CN). AN, elevated melanocyte lesions more than 5 mm in diameter, are independent predictors of malignant melanoma. miR-125 was downregulated and miR-let-7c upregulated in AN compared to CN [149]. In another study, miR-125b was reported as a prognostic marker of metastatic melanoma [168]. These researchers found that miR-125b was downregulated in primary cutaneous melanomas that produced early metastases (T2, N1, M0) compared to sentinel lymph node-negative (T2, N0, M0) melanomas [168].

## *6.11.14 miRNA Expression in Body Fluids of Patients with Rhabdomyosarcoma*

### 6.11.14.1 miRNA Expression in Blood

Miyachi et al. [150] investigated the possibility of using muscle-specific miRNAs (miR-1, -133a, -133b, and -206) as biomarkers of rhabdomyosarcoma (RMS). Their results indicated that expression of muscle-specific miRNAs, especially miR-206, was significantly higher in RMS cell lines than in other tumor cell lines, as well as in rhabdomyosarcoma tumor specimens. It was further demonstrated that serum levels of muscle-specific miRNAs were significantly higher in patients with RMS tumors in comparison to patients with without RMS. Their findings suggest the possibility of using circulating muscle-specific miRNAs, especially miR-206, as biomarkers for RMS [150].

## *6.11.15 miRNA Expression in Body Fluids of Patients with Diffuse Large B-Cell Lymphoma (DLBCL)*

### 6.11.15.1 miRNA Expression in Blood

The presence of miRNAs in the serum of patients with DLBCL was first described in 2008 [93]. The high expression levels of tumor-associated circulating miRNAs (miR-21, -155, and -210) in serum of patients with DLBCL compared to health

subjects suggests that miRNAs can serve as potential noninvasive diagnostic markers for DLBCL [93]. It was further demonstrated that high expression levels of miR-21 was associated with relapse-free survival [93].

## 6.12 Challenges in the Use of miRNAs as Biomarkers and Possible Solutions

Advances in molecular medicine, genomics and proteomics have resulted in the discovery of many candidate biomarkers with potential clinical value [169], including miRNAs. In addition, such advances, specifically in the areas of gene expression and proteomics have made it possible to correlate disease stages with molecular marker profiles. Although hundreds of miRNA cancer markers have been identified to date, only few candidate biomarkers are universally recognized as superior prognostic or diagnostic tools, and even fewer have been validated and approved by FDA. Multiple factors might have contributed to the slow pace of advance in cancer biomarkers. miRNAs, as mRNA regulators, could serve as efficient diagnostic and prognostic candidates, and potential therapeutic targets. However, there are still a lot of challenges regarding miRNAs in body fluids that need to be confronted, in order to produce an ideal cancer biomarker.

One of the several challenges of using plasma or serum samples for miRNA analysis is that no specific miRNA is available to use as an endogenous control to normalize the relative quantity of miRNA in the plasma or serum. Another limitation is that other classes of RNAs or mRNAs which could be used as a standard for qRT-PCR for measuring circulating miRNAs (e.g., as "housekeeping" serum miRNA/small RNA) are not stable in serum. A third challenge are the questions concerning the biological half-life of the circulating miRNAs, which requires more evaluation, as it may have important implications in clinical applications [152]. A fourth challenge involves the specificity of miRNAs; for instance, a single miRNA can distinguish different cancers which have the same serum miRNA, e.g., miR-21 in DLBCL and pancreatic cancer. A fifth reason is that the technologies available have limited power to detect low-abundance cancer biomarkers against the background of high-abundance plasma proteins, which implies that many of the best markers may be missed until discovery technology improves. A sixth reason is the standardization of miRNAs: the preparation of serum/plasma will need be standardized in order to generalize findings from different labs, patients or groups. A seventh challenge is that the capacity to verify and validate existing candidate biomarkers (through the approach of rigorous testing in large sample sets from many diseases) is limited, suggesting that it is possible the required biomarkers may have already been "discovered" but not yet validated [170].

Possible solutions to the challenges facing the use of miRNAs in body fluids as cancer biomarkers has been suggested by many researchers. Patell and Sauter [94] note that most body fluid studies have been performed on samples from the circulation. These researchers suggest that organ specific body fluids will be more

predictive of disease, as they are not diluted by the contribution from other organs in the body. They further suggested that additional work needs to be done looking at organ specific body fluids for cancer risk assessment and early disease detection [94]. Overall, although important progress has been made in the field of miRNAs as cancer biomarkers, and many researchers have focused on the prediction and identification of novel miRNAs, much work remains to validate, annotate, and characterize identified miRNAs. It is our hope that with continuous improvement in detection techniques, more progress will be made in the application of circulating miRNA in clinical diagnosis and management of cancer.

## 6.13 Conclusion

The detection of miRNA in body fluids is a new field of study that has shown promise as useful markers for early cancer detection. This book chapter has summarized the use of miRNAs as biomarkers for cancer, with an emphasis on the attributes of a good miRNA biomarker, evidence supporting the potential use of circulating (serum and plasma) and urinary miRNAs as diagnostic markers, the advantages of their use, challenges (e.g., validation of cancer biomarkers etc.) and possible solutions in their assessment as biomarkers for clinical diagnostic application. Numerous studies are being conducted to identify novel biomarkers, as well as with validating promising candidates, once identified. Although detecting cancer at an early stage is very attractive, factors such as the probability of getting false positive/negative results must be considered when determining the usefulness of the marker.

## References

1. Duffy MJ. Role of tumor markers in patients with solid cancers: a critical review. Eur J Intern Med. 2007;18(3):175–84.
2. Roulston JE. Limitations of tumour markers in screening. Br J Surg. 1990;77(9):961–2.
3. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, Galas DJ, Wang K. The microRNA spectrum in 12 body fluids. Clin Chem. 2010;56(11):1733–41.
4. Hayes DF, Bast RC, Desch CE, Fritsche Jr H, Kemeny NE, Jessup JM, Locker GY, Macdonald JS, Mennel RG, Norton L, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst. 1996;88(20):1456–66.
5. Bence-Jones H. Papers on chemical pathology. Lecture III. Lancet. 1847;ii:269–72.
6. Kyle RA. Multiple myeloma: how did it begin? Mayo Clin Proc. 1994;69(7):680–3.
7. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol. 2008;5(10):588–99.
8. Chan DW, Beveridge RA, Muss H, Fritsche HA, Hortobagyi G, Theriault R, Kiang D, Kennedy BJ, Evelegh M. Use of Truquant BR radioimmunoassay for early detection of breast cancer recurrence in patients with stage II and stage III disease. J Clin Oncol. 1997;15(6):2322–8.
9. Gion M, Mione R, Leon AE, Dittadi R. Comparison of the diagnostic accuracy of CA27.29 and CA15.3 in primary breast cancer. Clin Chem. 1999;45(5):630–7.

10. Brooks M. Breast cancer screening and biomarkers. Methods Mol Biol. 2009;472:307–21.
11. Chourin S, Veyret C, Chevrier A, Loeb A, Gray C, Basuyau J. Routine use of serial plasmatic CA 15-3 determinations during the follow-up of patients treated for breast cancer. Evaluation as factor of early diagnosis of recurrence. Ann Biol Clin (Paris). 2008;66(4):385–92.
12. Bast Jr RC, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest. 1981;68(5):1331–7.
13. Bast Jr RC, Klug TL, St John E, Jenison E, Niloff JM, Lazarus H, Berkowitz RS, Leavitt T, Griffiths CT, Parker L, et al. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med. 1983;309(15):883–7.
14. Hogdall E. Cancer antigen 125 and prognosis. Curr Opin Obstet Gynecol. 2008;20(1):4–8.
15. Greenberg MD, Sedlacek TV, Campion MJ. Cervical neoplasia: are adjunctive tests to cervical cytology worthwhile? Clin Obstet Gynecol. 1995;38(3):600–9.
16. Eddy DM. Screening for cervical cancer. Ann Intern Med. 1990;113(3):214–26.
17. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, Raab S, Sherman M, Wilbur D, Wright Jr T, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA. 2002;287(16):2114–9.
18. Wang MC, Valenzuela LA, Murphy GP, Chu TM. Purification of a human prostate specific antigen. Invest Urol. 1979;17(2):159–63.
19. Kuriyama M, Wang MC, Lee CI, Papsidero LD, Killian CS, Inaji H, Slack NH, Nishiura T, Murphy GP, Chu TM. Use of human prostate-specific antigen in monitoring prostate cancer. Cancer Res. 1981;41(10):3874–6.
20. Killian CS, Emrich LJ, Vargas FP, Yang N, Wang MC, Priore RL, Murphy GP, Chu TM. Relative reliability of five serially measured markers for prognosis of progression in prostate cancer. J Natl Cancer Inst. 1986;76(2):179–85.
21. Killian CS, Yang N, Emrich LJ, Vargas FP, Kuriyama M, Wang MC, Slack NH, Papsidero LD, Murphy GP, Chu TM. Prognostic importance of prostate-specific antigen for monitoring patients with stages B2 to D1 prostate cancer. Cancer Res. 1985;45(2):886–91.
22. Mizuchi A, Kitagawa N, Miyachi Y, Iio M. Monoclonal antibodies to human chorionic gonadotropin and their application to two-site sandwich radioimmunoassay. J Immunol Methods. 1984;74(2):369–74.
23. Laguna MP, Pizzocaro G, Klepp O, Algaba F, Kisbenedek L, Leiva O. EAU guidelines on testicular cancer. Eur Urol. 2001;40(2):102–10.
24. Mann K, Saller B, Hoermann R. Clinical use of HCG and hCG beta determinations. Scand J Clin Lab Invest Suppl. 1993;216:97–104.
25. Abelev GI. Alpha-fetoprotein as a marker of embryo-specific differentiations in normal and tumor tissues. Transplant Rev. 1974;20:3–37.
26. Blohm ME, Vesterling-Horner D, Calaminus G, Gobel U. Alpha 1-fetoprotein (AFP) reference values in infants up to 2 years of age. Pediatr Hematol Oncol. 1998;15(2):135–42.
27. Brewer JA, Tank ES. Yolk sac tumors and alpha-fetoprotein in first year of life. Urology. 1993;42(1):79–80.
28. Javadpour N. The role of biologic tumor markers in testicular cancer. Cancer. 1980;45(7 Suppl):1755–61.
29. Lange PH, Fraley EE. Serum alpha-fetoprotein and human chorionic gonadotropin in the treatment of patients with testicular tumors. Urol Clin North Am. 1977;4(3):393–406.
30. Nichols CR, Saxman S, Williams SD, Loehrer PJ, Miller ME, Wright C, Einhorn LH. Primary mediastinal nonseminomatous germ cell tumors. A modern single institution experience. Cancer. 1990;65(7):1641–6.
31. Johnson PJ, Portmann B, Williams R. Alpha-fetoprotein concentrations measured by radioimmunoassay in diagnosing and excluding hepatocellular carcinoma. Br Med J. 1978;2(6138):661–3.
32. Nomura F, Ohnishi K, Tanabe Y. Clinical features and prognosis of hepatocellular carcinoma with reference to serum alpha-fetoprotein levels. Analysis of 606 patients. Cancer. 1989;64(8):1700–7.

33. Sheu JC, Sung JL, Chen DS, Yang PM, Lai MY, Lee CS, Hsu HC, Chuang CN, Yang PC, Wang TH, et al. Growth rate of asymptomatic hepatocellular carcinoma and its clinical implications. Gastroenterology. 1985;89(2):259–66.
34. Lok AS, Lai CL. Alpha-Fetoprotein monitoring in Chinese patients with chronic hepatitis B virus infection: role in the early detection of hepatocellular carcinoma. Hepatology. 1989;9(1):110–5.
35. Okuda K. Early recognition of hepatocellular carcinoma. Hepatology. 1986;6(4):729–38.
36. Johnson PJ, Williams R. Cirrhosis and the aetiology of hepatocellular carcinoma. J Hepatol. 1987;4(1):140–7.
37. Kew MC, Popper H. Relationship between hepatocellular carcinoma and cirrhosis. Semin Liver Dis. 1984;4(2):136–46.
38. Gold P, Freedman SO. Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. J Exp Med. 1965;121:439–62.
39. Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. J Exp Med. 1965;122(3):467–81.
40. Hammarstrom S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Semin Cancer Biol. 1999;9(2):67–81.
41. Boucher D, Cournoyer D, Stanners CP, Fuks A. Studies on the control of gene expression of the carcinoembryonic antigen family in human tissue. Cancer Res. 1989;49(4):847–52.
42. Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? Clin Chem. 2001;47(4):624–30.
43. Gross ME, Zorbas MA, Danels YJ, Garcia R, Gallick GE, Olive M, Brattain MG, Boman BM, Yeoman LC. Cellular growth response to epidermal growth factor in colon carcinoma cells with an amplified epidermal growth factor receptor derived from a familial adenomatous polyposis patient. Cancer Res. 1991;51(5):1452–9.
44. Rubin BP, Singer S, Tsao C, Duensing A, Lux ML, Ruiz R, Hibbard MK, Chen CJ, Xiao S, Tuveson DA, et al. KIT activation is a ubiquitous feature of gastrointestinal stromal tumors. Cancer Res. 2001;61(22):8118–21.
45. Fletcher CD, Berman JJ, Corless C, Gorstein F, Lasota J, Longley BJ, Miettinen M, O'Leary TJ, Remotti H, Rubin BP, et al. Diagnosis of gastrointestinal stromal tumors: a consensus approach. Hum Pathol. 2002;33(5):459–65.
46. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, Ishiguro S, Kawano K, Hanada M, Kurata A, Takeda M, et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. Science. 1998;279(5350):577–80.
47. Miettinen M, Lasota J. Gastrointestinal stromal tumors–definition, clinical, histological, immunohistochemical, and molecular genetic features and differential diagnosis. Virchows Arch. 2001;438(1):1–12.
48. Sarlomo-Rikala M, Kovatich AJ, Barusevicius A, Miettinen M. CD117: a sensitive marker for gastrointestinal stromal tumors that is more specific than CD34. Mod Pathol. 1998;11(8):728–34.
49. Duffy MJ. CA 19-9 as a marker for gastrointestinal cancers: a review. Ann Clin Biochem. 1998;35(Pt 3):364–70.
50. Goggins M. Molecular markers of early pancreatic cancer. J Clin Oncol. 2005; 23(20):4524–31.
51. Lamerz R. Role of tumour markers, cytogenetics. Ann Oncol. 1999;10 Suppl 4:145–9.
52. Klapdor R, Aronsson AC, Duffy MJ. Tumor markers in gastrointestinal cancers, EGTM recommendations. Anticancer Res. 1999;19:2785–820.
53. Ferrone CR, Finkelstein DM, Thayer SP, Muzikansky A, Fernandez-delCastillo C, Warshaw AL. Perioperative CA19-9 levels can predict stage and survival in patients with resectable pancreatic adenocarcinoma. J Clin Oncol. 2006;24(18):2897–902.
54. Van Herle AJ, Vassart G, Dumont JE. Control of thyroglobulin synthesis and secretion. (First of two parts). N Engl J Med. 1979;301(5):239–49.
55. Holt E. Controversies in the surveillance of patients with well differentiated thyroid cancer. Curr Opin Oncol. 2007;19(1):6–10.

56. Mazzaferri EL, Robbins RJ, Spencer CA, Braverman LE, Pacini F, Wartofsky L, Haugen BR, Sherman SI, Cooper DS, Braunstein GD, et al. A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. J Clin Endocrinol Metab. 2003;88(4):1433–41.
57. Halling KC. Vysis UroVysion for the detection of urothelial carcinoma. Expert Rev Mol Diagn. 2003;3(4):507–19.
58. Bergman J, Reznichek RC, Rajfer J. Surveillance of patients with bladder carcinoma using fluorescent in-situ hybridization on bladder washings. BJU Int. 2008;101(1):26–9.
59. Bollmann M, Heller H, Bankfalvi A, Griefingholt H, Bollmann R. Quantitative molecular urinary cytology by fluorescence in situ hybridization: a tool for tailoring surveillance of patients with superficial bladder cancer? BJU Int. 2005;95(9):1219–25.
60. Bubendorf L, Grilli B, Sauter G, Mihatsch MJ, Gasser TC, Dalquen P. Multiprobe FISH for enhanced detection of bladder cancer in voided urine specimens and bladder washings. Am J Clin Pathol. 2001;116(1):79–86.
61. Daniely M, Rona R, Kaplan T, Olsfanger S, Elboim L, Freiberger A, Lew S, Leibovitch I. Combined morphologic and fluorescence in situ hybridization analysis of voided urine samples for the detection and follow-up of bladder cancer in patients with benign urine cytology. Cancer. 2007;111(6):517–24.
62. Halling KC, King W, Sokolova IA, Meyer RG, Burkhardt HM, Halling AC, Cheville JC, Sebo TJ, Ramakumar S, Stewart CS, et al. A comparison of cytology and fluorescence in situ hybridization for the detection of urothelial carcinoma. J Urol. 2000;164(5):1768–75.
63. Skacel M, Fahmy M, Brainard JA, Pettay JD, Biscotti CV, Liou LS, Procop GW, Jones JS, Ulchaker J, Zippe CD, et al. Multitarget fluorescence in situ hybridization assay detects transitional cell carcinoma in the majority of patients with bladder cancer and atypical or negative urine cytology. J Urol. 2003;169(6):2101–5.
64. Landman J, Chang Y, Kavaler E, Droller MJ, Liu BC. Sensitivity and specificity of NMP-22, telomerase, and BTA in the detection of human bladder cancer. Urology. 1998;52(3): 398–402.
65. Miyanaga N, Akaza H, Ishikawa S, Ohtani M, Noguchi R, Kawai K, Koiso K, Kobayashi M, Koyama A, Takahashi T. Clinical evaluation of nuclear matrix protein 22 (NMP22) in urine as a novel marker for urothelial cancer. Eur Urol. 1997;31(2):163–8.
66. Soloway MS, Briggman V, Carpinito GA, Chodak GW, Church PA, Lamm DL, Lange P, Messing E, Pasciak RM, Reservitz GB, et al. Use of a new tumor marker, urinary NMP22, in the detection of occult or rapidly recurring transitional cell carcinoma of the urinary tract following surgical treatment. J Urol. 1996;156(2 Pt 1):363–7.
67. Martinez-Pineiro JA, Pertusa C, Maganto E, Zancajo VG, Magallon M, Losada G, Ortega F. Urinary fibrinogen degradation products (FDP) in bladder cancer. Eur Urol. 1978;4(5): 348–50.
68. Oeda T, Manabe D. The usefulness of urinary FDP in the diagnosis of bladder cancer: comparison with NMP22, BTA and cytology. Nihon Hinyokika Gakkai Zasshi. 2001;92(1):1–5.
69. Fradet Y, Lockhard C. Performance characteristics of a new monoclonal antibody test for bladder cancer: ImmunoCyt trade mark. Can J Urol. 1997;4(3):400–5.
70. Moll R, Franke WW, Schiller DL, Geiger B, Krepler R. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell. 1982;31(1): 11–24.
71. Price MR, Rye PD, Petrakou E, Murray A, Brady K, Imai S, Haga S, Kiyozuka Y, Schol D, Meulenbroek MF, et al. Summary report on the ISOBM TD-4 Workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin. San Diego, Calif., November 17-23, 1996. Tumour Biol. 1998;19 Suppl 1:1–20.
72. Chu PG, Weiss LM. Keratin expression in human tissues and neoplasms. Histopathology. 2002;40(5):403–39.
73. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. Mod Pathol. 2008;21 Suppl 2:S8–15.

74. Diaz NM. Laboratory testing for HER2/neu in breast carcinoma: an evolving strategy to predict response to targeted therapy. Cancer Control. 2001;8(5):415–8.
75. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. 1989;244(4905):707–12.
76. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002;347(25):1999–2009.
77. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530–6.
78. Ross JS, Hatzis C, Symmans WF, Pusztai L, Hortobagyi GN. Commercialized multigene predictors of clinical outcome for breast cancer. Oncologist. 2008;13(5):477–93.
79. Kay PH, Wells FC, Goldstraw P. A multidisciplinary approach to primary nonseminomatous germ cell tumors of the mediastinum. Ann Thorac Surg. 1987;44(6):578–82.
80. Jamshidian H, Kor K, Djalali M. Urine concentration of nuclear matrix protein 22 for diagnosis of transitional cell carcinoma of bladder. Urol J. 2008;5(4):243–7.
81. Schroeder GL, Lorenzo-Gomez MF, Hautmann SH, Friedrich MG, Ekici S, Huland H, Lokeshwar V. A side by side comparison of cytology and biomarkers for bladder cancer detection. J Urol. 2004;172(3):1123–6.
82. Toma MI, Friedrich MG, Hautmann SH, Jakel KT, Erbersdobler A, Hellstern A, Huland H. Comparison of the ImmunoCyt test and urinary cytology with other urine tests in the detection and surveillance of bladder cancer. World J Urol. 2004;22(2):145–9.
83. Taplin S, Abraham L, Barlow WE, Fenton JJ, Berns EA, Carney PA, Cutter GR, Sickles EA, Carl D, Elmore JG. Mammography facility characteristics associated with interpretive accuracy of screening mammography. J Natl Cancer Inst. 2008;100(12):876–87.
84. Thompson A, Brennan K, Cox A, Gee J, Harcourt D, Harris A, Harvie M, Holen I, Howell A, Nicholson R, et al. Evaluation of the current knowledge limitations in breast cancer research: a gap analysis. Breast Cancer Res. 2008;10(2):R26.
85. Perkins GL, Slater ED, Sanders GK, Prichard JG. Serum tumor markers. Am Fam Physician. 2003;68(6):1075–82.
86. Strauss GM. Measuring effectiveness of lung cancer screening: from consensus to controversy and back. Chest. 1997;112(4 Suppl):216S–28.
87. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast Jr RC. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. 2007;25(33):5287–312.
88. O'Hanlon DM, Kerin MJ, Kent P, Maher D, Grimes H, Given HF. An evaluation of preoperative CA 15-3 measurement in primary breast carcinoma. Br J Cancer. 1995;71(6):1288–91.
89. Uehara M, Kinoshita T, Hojo T, Akashi-Tanaka S, Iwamoto E, Fukutomi T. Long-term prognostic study of carcinoembryonic antigen (CEA) and carbohydrate antigen 15-3 (CA 15-3) in breast cancer. Int J Clin Oncol. 2008;13(5):447–51.
90. Smith DS, Bullock AD, Catalona WJ. Racial differences in operating characteristics of prostate cancer screening tests. J Urol. 1997;158(5):1861–5. discussion 1865-1866.
91. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008;105(30):10513–8.
92. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008;18(10):997–1006.
93. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, Banham AH, Pezzella F, Boultwood J, Wainscoat JS, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol. 2008;141(5):672–5.

94. Patel N, Sauter ER. Body fluid micro(mi)RNAs as biomarkers for human cancer. J Nucleic Acids Invest 2011, 2(e1).
95. Garzon R, Marcucci G, Croce CM. Targeting microRNAs in cancer: rationale, strategies and challenges. Nat Rev Drug Discov. 2010;9(10):775–89.
96. Ambros V. The functions of animal microRNAs. Nature. 2004;431(7006):350–5.
97. Gregory RI, Shiekhattar R. MicroRNA biogenesis and cancer. Cancer Res. 2005; 65(9):3509–12.
98. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116(2):281–97.
99. Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell. 2003;113(6):673–6.
100. Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. Genes Dev. 2004;18(5):504–11.
101. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science. 2002;297(5589):2056–60.
102. Zhang H, Kolb FA, Brondani V, Billy E, Filipowicz W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. 2002;21(21):5875–85.
103. Doench JG, Petersen CP. Sharp PA: siRNAs can function as miRNAs. Genes Dev. 2003; 17(4):438–42.
104. Olsen PH, Ambros V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol. 1999;216(2):671–80.
105. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. 1993;75(5):855–62.
106. Rajewsky N. microRNA target predictions in animals. Nat Genet. 2006;38(Suppl):S8–13.
107. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. PLoS Biol. 2004;2(11):e363.
108. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15–20.
109. Krutzfeldt J, Poy MN, Stoffel M. Strategies to determine the biological function of microRNAs. Nat Genet. 2006;38(Suppl):S14–9.
110. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell. 2003;113(1):25–36.
111. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901–6.
112. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. An analysis of human microRNA and disease associations. PLoS One. 2008;3(10):e3420.
113. Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. Science. 2004;303(5654):83–6.
114. Johnston RJ, Hobert O. A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature. 2003;426(6968):845–9.
115. Croce CM, Calin GA. miRNAs, cancer, and stem cell division. Cell. 2005;122(1):6–7.
116. Michael MZ, O'Connor SM, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res. 2003; 1(12):882–91.
117. Bao N, Lye KW, Barton MK. MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev Cell. 2004; 7(5):653–62.
118. Brase JC, Wuttig D, Kuner R, Sultmann H. Serum microRNAs as non-invasive biomarkers for cancer. Mol Cancer. 2010;9:306.
119. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006; 6(11):857–66.

120. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer. 2006;6(4):259–69.
121. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834–8.
122. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, Guenther SM, O'Leary JJ, Sheils O. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol. 2007;7:36.
123. Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. Microarray-based, high-throughput gene expression profiling of microRNAs. Nat Methods. 2004; 1(2):155–61.
124. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, Ju J. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. RNA. 2007;13(10):1668–74.
125. Allawi HT, Dahlberg JE, Olson S, Lund E, Olson M, Ma WP, Takova T, Neri BP, Lyamichev VI. Quantitation of microRNAs using a modified Invader assay. RNA. 2004;10(7):1153–61.
126. Lu DP, Read RL, Humphreys DT, Battah FM, Martin DI, Rasko JE. PCR-based expression analysis and identification of microRNAs. J RNAi Gene Silencing. 2005;1(1):44–9.
127. Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids–the mix of hormones and biomarkers. Nat Rev Clin Oncol. 2011; 8(8):467–77.
128. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio MV, Visone R, Sever NI, Fabbri M, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med. 2005;353(17):1793–801.
129. Zhu W, Qin W, Atasoy U, Sauter ER. Circulating microRNAs in breast cancer and healthy subjects. BMC Res Notes. 2009;2:89.
130. Heneghan HM, Miller N, Kelly R, Newell J, Kerin MJ. Systemic miRNA-195 differentiates breast cancer from other malignancies and is a potential biomarker for detecting noninvasive and early stage disease. Oncologist. 2010;15(7):673–82.
131. Asaga S, Kuo C, Nguyen T, Terpenning M, Giuliano AE, Hoon DS. Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. Clin Chem. 2011; 57(1):84–91.
132. Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. PLoS One. 2010;5(10):e13735.
133. Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, Poon TC, Ng SS, Sung JJ. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut. 2009;58(10):1375–81.
134. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. Int J Cancer. 2010;127(1):118–26.
135. Tsujiura M, Ichikawa D, Komatsu S, Shiozaki A, Takeshita H, Kosuga T, Konishi H, Morimura R, Deguchi K, Fujiwara H, et al. Circulating microRNAs in plasma of patients with gastric cancers. Br J Cancer. 2010;102(7):1174–9.
136. Li X, Luo F, Li Q, Xu M, Feng D, Zhang G, Wu W. Identification of new aberrantly expressed miRNAs in intestinal-type. Oncol Rep. 2011;26(6):1431–9.
137. Yamamoto Y, Kosaka N, Tanaka M, Koizumi F, Kanai Y, Mizutani T, Murakami Y, Kuroda M, Miyajima A, Kato T, et al. MicroRNA-500 as a potential diagnostic marker for hepatocellular carcinoma. Biomarkers. 2009;14(7):529–38.
138. Qu KZ, Zhang K, Li H, Afdhal NH, Albitar M. Circulating microRNAs as biomarkers for hepatocellular carcinoma. J Clin Gastroenterol. 2011;45(4):355–60.
139. Shen J, Liu Z, Todd NW, Zhang H, Liao J, Yu L, Guarnera MA, Li R, Cai L, Zhan M, et al. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. BMC Cancer. 2011;11:374.

140. Zheng D, Haddadin S, Wang Y, Gu LQ, Perry MC, Freter CE, Wang MX. Plasma microRNAs as novel biomarkers for early detection of lung cancer. Int J Clin Exp Pathol. 2011;4(6):575–86.
141. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. Gynecol Oncol. 2008;110(1):13–21.
142. Resnick KE, Alder H, Hagan JP, Richardson DL, Croce CM, Cohn DE. The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. Gynecol Oncol. 2009;112(1):55–9.
143. Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M, Borges M, Goggins M. Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. Cancer Res. 2010;70(13):5226–37.
144. Ho AS, Huang X, Cao H, Christman-Skieller C, Bennewith K, Le QT, Koong AC. Circulating miR-210 as a Novel Hypoxia Marker in Pancreatic Cancer. Transl Oncol. 2010;3(2):109–13.
145. Wang J, Chen J, Chang P, LeBlanc A, Li D, Abbruzzesse JL, Frazier ML, Killary AM, Sen S. MicroRNAs in plasma of pancreatic ductal adenocarcinoma patients as novel blood-based biomarkers of disease. Cancer Prev Res (Phila). 2009;2(9):807–13.
146. Morimura R, Komatsu S, Ichikawa D, Takeshita H, Tsujiura M, Nagata H, Konishi H, Shiozaki A, Ikoma H, Okamoto K, et al. Novel diagnostic value of circulating miR-18a in plasma of patients with pancreatic cancer. Br J Cancer. 2011;105(11):1733–40.
147. Wong TS, Liu XB, Wong BY, Ng RW, Yuen AP, Wei WI. Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue. Clin Cancer Res. 2008;14(9):2588–92.
148. Zhang C, Wang C, Chen X, Yang C, Li K, Wang J, Dai J, Hu Z, Zhou X, Chen L, et al. Expression profile of microRNAs in serum: a fingerprint for esophageal squamous cell carcinoma. Clin Chem. 2010;56(12):1871–9.
149. Holst LM, Kaczkowski B, Glud M, Futoma-Kazmierczak E, Hansen LF, Gniadecki R. The microRNA molecular signature of atypic and common acquired melanocytic nevi: differential expression of miR-125b and let-7c. Exp Dermatol. 2011;20(3):278–80.
150. Miyachi M, Tsuchiya K, Yoshida H, Yagyu S, Kikuchi K, Misawa A, Iehara T, Hosoi H. Circulating muscle-specific microRNA, miR-206, as a potential diagnostic marker for rhabdomyosarcoma. Biochem Biophys Res Commun. 2010;400(1):89–93.
151. Xie Y, Todd NW, Liu Z, Zhan M, Fang H, Peng H, Alattar M, Deepak J, Stass SA, Jiang F. Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. Lung Cancer. 2010;67(2):170–6.
152. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, Fang H, Zhang J, Katz RL, Jiang F. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. Int J Cancer. 2010;127(12):2870–8.
153. Xing L, Todd NW, Yu L, Fang H, Jiang F. Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. Mod Pathol. 2010;23(8):1157–64.
154. Xie L, Chen X, Wang L, Qian X, Wang T, Wei J, Yu L, Ding Y, Zhang C, Liu B. Cell-free miRNAs may indicate diagnosis and docetaxel sensitivity of tumor cells in malignant effusions. BMC Cancer. 2010;10:591.
155. Hanke M, Hoefig K, Merz H, Feller AC, Kausch I, Jocham D, Warnecke JM, Sczakiel G. A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. Urol Oncol. 2010;28(6):655–61.
156. Liu CJ, Lin SC, Yang CC, Cheng HW, Chang KW. Exploiting salivary miR-31 as a clinical biomarker of oral squamous cell carcinoma. Head Neck. 2012;34(2):219–24.
157. Park NJ, Zhou H, Elashoff D, Henson BS, Kastratovic DA, Abemayor E, Wong DT. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clin Cancer Res. 2009;15(17):5473–7.

158. Shigoka M, Tsuchida A, Matsudo T, Nagakawa Y, Saito H, Suzuki Y, Aoki T, Murakami Y, Toyoda H, Kumada T, et al. Deregulation of miR-92a expression is implicated in hepatocellular carcinoma development. Pathol Int. 2010;60(5):351–7.
159. Hu Z, Chen X, Zhao Y, Tian T, Jin G, Shu Y, Chen Y, Xu L, Zen K, Zhang C, et al. Serum microRNA signatures identified in a genome-wide serum microRNA expression profiling predict survival of non-small-cell lung cancer. J Clin Oncol. 2010;28(10):1721–6.
160. Brase JC, Johannes M, Schlomm T, Falth M, Haese A, Steuber T, Beissbarth T, Kuner R, Sultmann H. Circulating miRNAs are correlated with tumor progression in prostate cancer. Int J Cancer. 2011;128(3):608–16.
161. Grinchuk OV, Jenjaroenpun P, Orlov YL, Zhou J, Kuznetsov VA. Integrative analysis of the human cis-antisense gene pairs, miRNAs and their transcription regulation patterns. Nucleic Acids Res. 2010;38(2):534–47.
162. Selcuklu SD, Yakicier MC, Erson AE. An investigation of microRNAs mapping to breast cancer related genomic gain and loss regions. Cancer Genet Cytogenet. 2009;189(1):15–23.
163. Roth C, Rack B, Muller V, Janni W, Pantel K, Schwarzenbach H. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. Breast Cancer Res. 2010;12(6):R90.
164. Gui J, Tian Y, Wen X, Zhang W, Zhang P, Gao J, Run W, Tian L, Jia X, Gao Y. Serum microRNA characterization identifies miR-885-5p as a potential marker for detecting liver pathologies. Clin Sci (Lond). 2010;120(5):183–93.
165. Bianchi F, Nicassio F, Marzi M, Belloni E, Dall'olio V, Bernard L, Pelosi G, Maisonneuve P, Veronesi G, Di Fiore PP. A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. EMBO Mol Med. 2011; 3(8):495–503.
166. Liu CJ, Kao SY, Tu HF, Tsai MM, Chang KW, Lin SC. Increase of microRNA miR-31 level in plasma could be a potential marker of oral cancer. Oral Dis. 2010;16(4):360–4.
167. Leidinger P, Keller A, Borries A, Reichrath J, Rass K, Jager SU, Lenhof HP, Meese E. High-throughput miRNA profiling of human melanoma blood samples. BMC Cancer. 2010;10:262.
168. Glud M, Rossing M, Hother C, Holst L, Hastrup N, Nielsen FC, Gniadecki R, Drzewiecki KT. Downregulation of miR-125b in metastatic cutaneous malignant melanoma. Melanoma Res. 2010;20(6):479–84.
169. Pritzker KP. Cancer biomarkers: easier said than done. Clin Chem. 2002;48(8):1147–50.
170. Polanski M, Anderson NL. Candidate cancer markers. Biomarker Insights. 2006;2:48.

# Chapter 7
# Unraveling the Complex Network of Interactions Between Noncoding RNAs and Epigenetics in Cancer

**Veronica Davalos and Manel Esteller**

**Abstract** Epigenetics is the study of heritable changes in gene expression that do not involve changes in the underlying DNA sequence. The most studied epigenetic modifications include DNA methylation and histone changes. These modifications are able to modulate the chromatin conformation and have a critical role in regulating gene expression. Over the last years, growing evidences have revealed the crucial role of epigenetic mechanisms controlling noncoding RNAs (ncRNAs) expression, in the same way as previously shown for protein-coding genes. Most interestingly, the link between ncRNAs and epigenetics is not limited to epigenetic regulation of ncRNAs, but also takes place in the opposite direction, meaning that these RNA molecules are able to control gene expression by regulating effectors of the epigenetic machinery. In this chapter both of these scenarios will be discussed, focusing in the cancer context. The complex network of reciprocal interactions between ncRNAs and epigenetics is just beginning to unravel and an exciting future in research about the role of ncRNAs in cancer epigenetics is guaranteed.

**Keywords** Epigenetics • Cancer • ncRNA • miRNA • lncRNA • DNA methylation • Histone modifications • Chromatin

V. Davalos
Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Hospital Duran i Reynals, L'Hospitalet, 3rd Floor, Av. Gran Via 199-203, 08907 Barcelona, Catalonia, Spain

M. Esteller (✉)
Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Hospital Duran i Reynals, L'Hospitalet, 3rd Floor, Av. Gran Via 199-203, 08907 Barcelona, Catalonia, Spain

Departament de Ciencies Fisiologiques II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain

Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain
e-mail: mesteller@idibell.cat

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_7,

## 7.1 Introduction

The term Epigenetics was introduced by C.H Waddington in 1939 to name "the causal interactions between genes and their products, which bring the phenotype into being" [1]. It was later redefined as those heritable changes in gene expression that do not involve changes in the underlying DNA sequence [2]. The most studied epigenetic modifications include DNA methylation and histone changes. These modifications are able to modulate the chromatin conformation and have a critical role in regulating gene expression and the accessibility of transcription factors, co-activators, and co-repressors.

In humans, DNA methylation occurs at the carbon-5 of the cytosine in CpG dinucleotides. This reaction is catalyzed by DNA methyltransferases (DNMTs): DNMT3a and DNMT3b (de novo DNMTs) transfer a methyl group from *S-adenosyl-methionine* to previously unmethylated cytosines, while DNMT1 (*maintenance* DNMT) preserves the methylation patterns throughout each cell division. CpG sites are not randomly distributed in the genome, but instead there are CpG-rich zones known as CpG islands, located mainly at the regulatory regions in 40–60 % of all genes [3, 4]. Methylation of CpG islands is a rare event in normal cells, restricted to untranscribed genes in X chromosome, imprinted genes, germ line genes, and some tissue-specific genes. However, CpG island hypermethylation is a common hallmark in cancer cells, first associated to tumor suppressor gene silencing [4, 5].

Along with DNA methylation, histone modifications are the most studied epigenetic events related to cancer progression. Regulation of gene expression can occur through posttranslational modifications of the histone tails, including covalent changes such as acetylation, methylation, phosphorylation, ubiquitination, sumoylation, proline isomerization, and ADP-ribosylation. Their presence on histones form the called "histone code" that dictates the chromatin structure in which DNA is packaged and can orchestrate the ordered recruitment of enzyme complexes to wrap the DNA [6, 7]. This histone code is "written" and "erased" by histone modifying enzymes. The "writer" of histone modification refers to an enzyme that catalyzes a chemical modification of histones in a residue-specific manner (i.e., histone methyltransferases HMTs or histone acetyltransferases HATs), and the "eraser" of histone marks refers to an enzyme that removes a chemical modification from histones (i.e., histone demethylases HDMs or histone deacetylases HDACs). This code is interpreted by "reader" or "effector" proteins that specifically bind to a certain type or a combination of histone modifications and translate the histone code into a meaningful biological outcome, whether it is transcriptional activation or silencing, or other cellular responses. In addition to this recruitment mechanism, histone marks per se can modulate the chromatin conformation based on steric or charge interactions (i.e., neutralization of the positive charges of histones by acetylation of lysines) [6, 8].

Another crucial player in epigenetic silencing is the Polycomb system. In mammals, two main Polycomb complexes have been identified: Polycomb repressive complex 1 (PRC1) and 2 (PRC2). PRC1 compacts chromatin and catalyzes the monoubiquitylation of histone H2A, and PRC2 also contributes to chromatin folding and catalyzes

the methylation of histone H3 at lysine 27 (H3K27). Polycomb complexes have been implicated in diverse biological processes such as epigenetic inheritance, differentiation, stem cell plasticity, proliferation, and senescence [9]. Altogether, this complex epigenetic network guarantees a dynamic and accurate control of gene expression.

Over the last years, growing evidences have revealed the crucial role of epigenetic mechanisms controlling noncoding RNAs (ncRNAs) expression, in the same way as previously shown for protein-coding genes. Most interestingly, the link between ncRNAs and epigenetics is not limited to epigenetic regulation of ncRNAs, but also takes place in the opposite direction, meaning that these RNA molecules are able to control gene expression by regulating effectors of the epigenetic machinery. In this chapter both of these scenarios will be discussed, focusing in the cancer context.

## 7.2 Epigenetically Regulated Noncoding RNAs in Cancer

### *7.2.1 Short and Midsize ncRNAs*

Epigenetic silencing is considered a hallmark of cancer. The first evidences showing how this mechanism is able to control ncRNA expression were found in microRNAs (miRNAs), the most widely studied class of ncRNAs. Using different strategies, independent groups were able to demonstrate that miRNA levels are epigenetically regulated [10–12]. Treating the human bladder cell line T24 with the chromatin-modifying drugs 5-*aza*-2′-deoxycytidine (a DNA demethylating agent) and 4-phenylbutyric acid (a histone deacetylase inhibitor, HDACi), Saito et al. found that about 5 % of a set of 313 miRNAs incorporated in a microarray platform were upregulated more than threefold, including **miR-127** [10]. Using an analogous approach but another HDACi, the hydroxamic acid LAQ824, Scott et al. reported that levels of 27 miRNAs were rapidly deregulated after treatment in the breast cancer cell line SKBr3; for instance expression of **miR-27** was significantly decreased [11]. Our group also contributed with the first proofs of the role of epigenetics in ncRNAs. Thus, using as a model system a colon cancer cell line genetically deficient for the DNA methyltransferases DNMT1 and DNMT3b (HCT-116 double knockout) and comparing it with the wild-type cell line (HCT-116), we identified epigenetic silencing of **miR-124a**. Furthermore, we demonstrated that miR-124a inhibition results in an increased expression of its target CDK6, and consequent phosphorylation of the downstream CDK6-regulated Rb protein [12]. CpG hypermethylation-associated silencing of this miRNA has also been detected in other tumors, such as glioblastoma multiforme [13], gastric cancer [14], hematopoietic malignancies [15, 16], hepatocellular carcinoma [17], and cervical tumors [18]. The HCT-116 DNMT1/3b double knockout cell line has been a valuable tool in the identification not only of epigenetically regulated miRNAs with tumor suppressor capacity, but also oncogenic miRNAs, as is the case for **let-7a-3**. This member of the archetypal let-7 miRNA gene family was found heavily methylated in normal human tissues but hypomethylated in some lung adenocarcinomas

[19]. In contrast, hypermethylation of this miRNA has been detected in ovarian and breast cancer [20, 21].

During the last few years, the list of miRNAs regulated by CpG methylation and histone modifications is in accelerated expansion. Some relevant examples are included in Table 7.1 and Fig. 7.1, e.g., **miR-1** in hepatocarcinoma [22]; **miR-107** in pancreatic cancer [34]; **miR-129-2** in colorectal [25], endometrial [35] and gastric cancer [33, 36]; **miR-137** in colorectal [25, 37]; gastric [39] and oral cancer [38, 41, 42]**; miR-193a** in hematological malignances [44], oral [43] and lung cancer [27]; **miR-196b** in gastric cancer [46] and hematopoietic malignancies [45]; **miR-199a** in gastric [47], ovarian [48], and testicular tumors [49]; **miR-203** in hematological malignances [32, 55, 56] and hepatocellular carcinomas [17] or **miR-9-1** in colorectal cancer. Interestingly, in this case its methylation was associated with the presence of lymph node metastasis [25]. The relevance of epigenetic control of miRNAs in cancer progression and metastasis has also been extensively documented. In 2008, our group described a miRNA DNA methylation signature for metastasis. By means of a pharmacological and genomic approach, treating lymph node metastatic cancer cells with a DNA demethylating agent followed by hybridization to an expression microarray, we detected reexpression of **miR-9**, **miR-34b/c,** and **miR-148a** upon drug treatment and confirmed cancer-specific hypermethylation-associated silencing. Most relevantly, the reintroduction of miR-148a and miR-34b/c in cancer cells with epigenetic inactivation inhibited their motility, reduced tumor growth, and inhibited metastasis formation in xenograft models, with an associated downregulation of the miRNA oncogenic target genes, such as C-MYC, E2F3, CDK6, and TGIF2. Moreover, miR-9, miR-34b/c and miR-148a hypermethylation was significantly associated with the appearance of lymph node human metastasis [24]. Epigenetic control of these miRNAs has been also confirmed by other studies [16, 23, 26, 27, 31–33]. Recently, our group and others have also shown the epigenetic silencing of a miRNA family, the **miR-200 family**, recognized as a master regulator of the epithelial phenotype, by targeting ZEB1 and ZEB2, important repressors of E-cadherin and cell polarity genes [50–54]. During tumor progression, the loss of epithelial features and acquisition of mesenchymal attributes through epithelial–mesenchymal transition (EMT) define tumor fate and metastatic dissemination. However, once cancer cells have reached a secondary site, regain of some epithelial properties (mesenchymal to epithelial transition, MET) could be advantageous for colonizing and establishing a secondary tumor. The plasticity and reversibility make epigenetic mechanisms excellent strategies to modulate miRNAs with variable expression during tumor progression. In this regard, we described a reversible mechanism to regulate expression of both clusters of miR-200 family, **miR-200ba429** and **miR-200c141**, during EMT and cancer progression through dynamic epigenetic regulation mediated by CpG island promoter methylation [53]. Interestingly, in bladder and lung cancer a coordinated epigenetic silencing of miR-200 and **miR-205** has been reported, the last one also involved in EMT [54, 57].

Most of the aforementioned examples describe miRNAs regulated by CpG methylation. Epigenetic control is nevertheless mainly achieved not only by DNA methylation but also by histone modifications, both epigenetic marks being closely

**Table 7.1** Epigenetically regulated noncoding RNAs in cancer

| ncRNA | Cancer type | Target genes | References |
|---|---|---|---|
| Let-7a-3 | Breast, lung, ovarian | IGF-II | [19–21] |
| miR-1 | Hepatocellular carcinoma | HDAC4 | [22] |
| miR-9 family | Metastatic tumors: breast, lung, melanoma; hematological malignances, renal cell carcinoma | | [16, 23–27] |
| miR-15a/miR-16-1 cluster | Chronic lymphocytic leukemia (CLL) | BCL2, MCL1 | [28] |
| miR-17-92 cluster | Colorectal | PTEN, BCL2L11, CDKN1A, miR-18a: NEDD9, and CDK19 | [29] |
| miR-27 | Breast | ZBTB10/RINZF, RYBP/DEDAF | [11] |
| miR-29a | Chronic lymphocytic leukemia (CLL) | BCL2, MCL1 | [28] |
| miR-31 | Breast | WAVE3 | [30] |
| miR-34b/c | Metastatic tumors: breast, lung, melanoma; hematological malignances; gastric and colorectal cancer | c-Myc, CDK6, E2F3 | [16, 24, 31–33] |
| miR-107 | Pancreas | CDK6 | [34] |
| miR-124a | Breast, colorectal, lung, gastric, cervical, HCC, glioblastoma, leukemia, lymphoma | CDK6, FOXA2 | [12–18] |
| miR-127 | Bladder | BCL6 | [10] |
| miR-129-2 | Endometrial, gastric, colorectal | SOX4 | [25, 33, 35, 36] |
| miR-137 | Colorectal, gastric, oral | Cdc42, LSD-1 | ([25, 37– 39]; Chen et al. [41, 42]) |
| miR-148a | Metastatic tumors: breast, lung, melanoma; breast cancer | TGIF2 | [23, 24] |
| miR-193a | Hematological malignances, lung, oral | c-kit, E2F6 | [27, 43, 44] |
| miR-196b | Gastric, hematopoietic malignancies | | [45, 46] |
| miR-199a | Gastric, ovarian, testicular | PODXL, IKKβ | [47–49] |
| miR-200 family | Colorectal, breast, lung, bladder | ZEB1, ZEB2, CRB3, LGL2 | [50–54] |
| miR-203 | Hematological malignances, hepatocellular carcinoma | ABL1 | [17, 32, 55, 56] |

(continued)

**Table 7.1** (continued)

| ncRNA | Cancer type | Target genes | References |
|---|---|---|---|
| miR-205 | Bladder, lung | PTEN, ZEB1, ZEB2 | [54, 57] |
| miR-224 | Hepatocellular carcinoma | | [58] |
| SNORD123 | Acute lymphoblastoid leukemia, acute myelogenous leukemia, and cancer cell lines | | [59] |
| U70C | Acute lymphoblastoid leukemia and cancer cell lines | | [59] |
| ACA59B | Acute lymphoblastoid leukemia, primary multiple myeloma, and cancer cell lines | | [59] |
| HBII-240 | Testicular germ cell tumors | | [60] |
| ACA8 | Testicular germ cell tumors | | [60] |
| ACA33 | Testicular germ cell tumors | | [60] |
| Uc.160+ | Colon, breast, lung | | [61] |
| Uc.283+A | Colon, breast, lung | | [61] |
| Uc.346+ | Colon, breast, lung | | [61] |
| MEG3 | Pituitary adenomas, hematological malignances | | [62–65] |
| LOC554202 | Breast | | [30] |

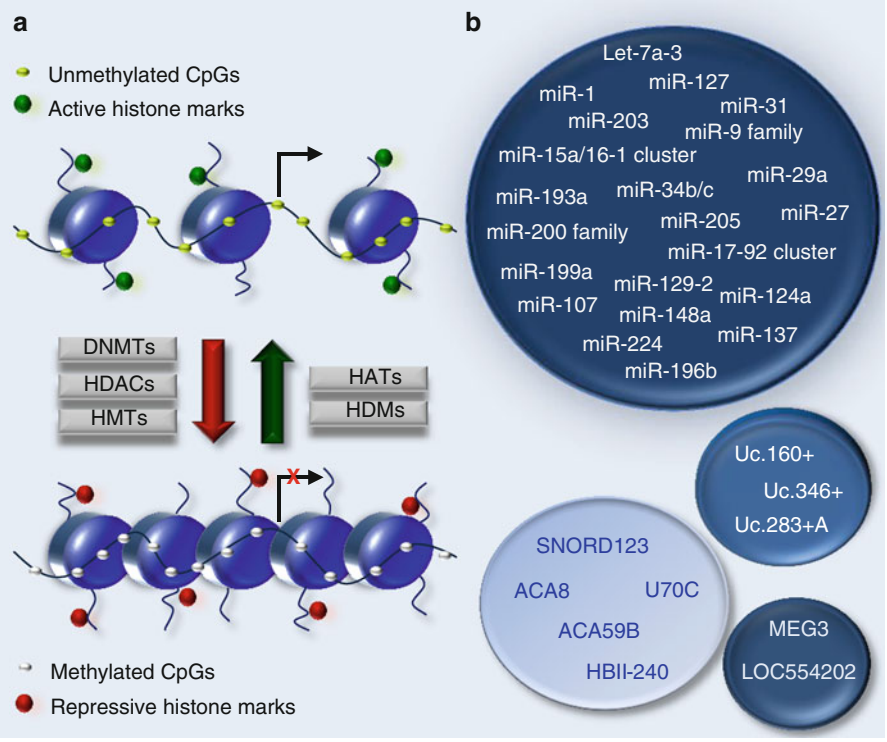


**Fig. 7.1** Epigenetically regulated noncoding RNAs in cancer. (**a**). Epigenetic mechanisms involved in regulating ncRNA expression. Unmethylated CpGs and presence of active histone marks in regulatory regions of ncRNAs are related to active transcription, while CpG hypermethylation and presence of repressive histone marks are common events associated to gene silencing in cancer. The enzymes responsible of modifying DNA and histones include DNA methyltransferases (DNMTs), histone acetyltransferases (HATs), histone deacetylases (HDACs), histone methyltransferases (HMTs), and histone demethylases (HDMs). (**b**). Epigenetically controlled ncRNAs, classified according to size (nt): miRNAs (*big circle*), snoRNAs (*medium circle*), and lncRNAs (*small circles*)

interconnected in order to guarantee a strict control of gene expression. However, many studies are focused on DNA methylation, perhaps largely due to technical constrains. But it is important to remark the crucial role of histone modifications controlling miRNA expression. Illustrative examples include **miR-15a**, **miR-16**, and **miR-29b** in chronic lymphocytic leukemia (CLL) [28]. miR-15a/miR-16 cluster was the first evidence linking miRNAs to human cancer, and common deletions in their genomic loci (13q14) were described as the main causal factor [66]. However, genomic deletions are not able to explain downregulation of these miRNAs in all the CLL cases. Recently, Sampath et al. have demonstrated that the common overexpression of HDACs in CLL mediates epigenetic silencing not only of the miR-15a/miR-16 cluster, but also of miR-29b [28]. HDACs are also important players in **miR-17-92** cluster regulation, as recently described in colorectal

cancer, where HDAC inhibitors decreased expression of miR-17-92 cluster with a corresponding increase in their target genes, including PTEN, BCL2L11 and CDKN1A [29]. In hepatocellular carcinoma, it has been reported that **miR-224** is reciprocally regulated by HDAC1, HDAC3, and the histone acetylase protein EP300 (E1A binding protein p300) [58]. Moreover, the recent development of genome-wide technologies, such as high-resolution ChIP-seq, has improved our knowledge about the role of histone modifications in miRNA regulation in cancer [67].

The new technical advances and vanguard strategies guarantee a booming understanding of the role of epigenetics not only in miRNA regulation, but also in other ncRNAs. It is the case of Small Nucleolar RNAs (snoRNAs), midsize ncRNAs of about 60 to 300nt located in the nucleolus whose main role is to guide chemical modifications of other RNAs. Thus, they are responsible for methylation and pseudouridylation of ribosomal RNA (rRNA) at about 50–100 sites per eukaryotic ribosome [68–70]. Our group has recently shown the existence of cancer-specific hypermethylation events in CpG islands associated with snoRNAs that lead to their transcriptional inactivation in transformed cells [59]. By data mining snoRNA databases and the scientific literature, we selected forty-nine snoRNAs that had a CpG island within ≤2 Kb or that were processed from a host gene with a 5′-CpG island. Interestingly, the host gene-associated 5′-CpG islands of the snoRNAs **SNORD123**, **U70C**, and **ACA59B** were hypermethylated in the cancer cell lines but not in the corresponding normal tissues. Most importantly, CpG island hypermethylation was associated with the transcriptional silencing of the respective snoRNAs not only in cancer cell lines, but also in a comprehensive cohort of primary tumors, demonstrating that the observed hypermethylation of snoRNAs was a common feature of various cancer types, particularly in leukemia [59]. Cheung et al. have also described epigenetic regulation of snoRNAs. They detected CpG hypomethylation in **HBII-240**, **ACA8**, and **ACA33** in testicular germ cell tumors, associated to snoRNAs upregulation in comparison with normal testis tissue [60]. Although the specific role of these snoRNAs in cancer remains to be elucidated, they are likely to contribute to tumorigenesis through an effect on ribosomes and protein translation, especially if we consider that translation is often perturbed in cancer cells [71].

### *7.2.2 Long ncRNAs*

Regarding long ncRNAs (lncRNAs), we have found that Transcribed Ultraconserved Regions (T-UCRs) undergo epigenetic silencing in cancer [61]. T-UCRs are ncRNAs >200 nt that are absolutely conserved between orthologous regions of the human, rat and mouse genomes. Previous studies have shown that UCR expression levels are altered in cancer and that T-UCR expression signatures define different human tumor types [72]. Treating cancer cells with a DNA-demethylating agent followed by hybridization to an expression microarray containing T-UCR sequences, we detected that **Uc.160+**, **Uc283+A** and **Uc.346+** undergo specific CpG island hypermethylation-associated silencing in cancer cells compared with

normal tissues. Most importantly, this finding was not only an in vitro phenomenon but also it was confirmed in a large set of primary human tumors [61]. Other studies describe negative regulation of T-UCRs by direct interaction with miRNAs in neuroblastoma [73].

Another example of epigenetically controlled lncRNA is the Maternally Expressed Gene 3 (**MEG3**). MEG3 was the first lncRNA proposed to function as a tumor suppressor, studying human pituitary adenomas [74]. CpG hypermethylation within its regulatory region has been involved in the loss of MEG3 expression in clinically nonfunctioning pituitary adenomas [62, 64], multiple myeloma [63], acute myeloid leukemia, and myelodysplastic syndromes [65]. Other interesting examples include **miR-31** and its host gene, the lncRNA **LOC554202**. The CpG island upstream of the miR-31 locus, which also spans the first exon of LOC554202, is hypermethylated in breast cancer resulting in silencing of both miRNA and the host lncRNA [30].

Remarkably, most lincRNAs (long noncoding RNAs located in intergenic regions) were first identified using a histone mark signature previously associated to active transcription in protein-coding genes: the "K4K36 signature" that consists of a short stretch of trimethylation of lysine 4 in the histone H3 (H3K4me3), which corresponds to promoter regions, followed by a longer stretch of trimethylation of lysine 36 in the histone H3 (H3K36me3), which covers the entire transcribed region. After excluding known genes, chromatin signatures revealed approximately 1,600 regions in the mouse genome and 3,000 regions in the human genome that were actively transcribed [75, 76]. This fact highlights the key role of epigenetic marks in ncRNA regulation. Moreover, it has been reported that numerous lncRNAs that are expressed at lower levels in embryonic stem cells exhibit higher levels of H3K27me3 at their promoters. In agreement with this fact, knockdown of the H3K27me3 methyltransferase Enhancer of zeste homolog 2 (Ezh2) results in derepression of these lncRNAs [77].

## 7.3 Noncoding RNAs as Regulators of the Cancer Epigenome

### *7.3.1 Short and Midsize ncRNAs*

Noncoding RNAs are not only regulated by epigenetic mechanisms but also modulate DNA methylation, histone modifications and chromatin remodeling by interfering with the effectors of the epigenetic machinery (Table 7.2). Thus, ncRNAs are able to regulate gene expression by two different ways: direct interaction with their target mRNAs and/or acting as master regulators of the epigenetic processes. Recent studies in this field have shown that DNMTs, HDACs, and HMTs are direct targets of miRNAs (Fig. 7.2). This subgroup of miRNAs that modulate the epigenetic machinery has been called "epi-miRNAs." The first evidence was reported by Fabbri et al.,

**Table 7.2** Noncoding RNAs as regulators of the cancer epigenome

| ncRNA | Cancer type | Target genes | References |
|---|---|---|---|
| miR-29 family | Acute myeloid leukemia, lung | DNMT1, DNMT3a, DNMT3b | [78, 79] |
| miR-101 | Bladder, prostate | EZH2 | [80, 81] |
| miR-140 | Colon and osteosarcoma | HDAC4 | [82] |
| miR-143 | Colorectal | DNMT3a | [83] |
| miR-148 | Cervical, cholangiocarcinoma | DNMT1, DNMT3b | [84, 85] |
| miR-152 | Cholangiocarcinoma | DNMT1 | [85] |
| miR-185 | Glioma | DNMT1 | [86] |
| miR-212 | Gastric | MeCP2 | [87] |
| miR-342 | Colorectal | DNMT1 | [88] |
| miR-373 | Cholangiocarcinoma | MBD2 | (Chen et al. [40]) |
| miR-449a | Prostate | HDAC1 | [89] |
| HOTAIR | Breast, colorectal, gastrointestinal stromal tumors, hepatocellular carcinoma | PRC2, LSD1 | [90–95] |
| ANRIL | Leukemia, Prostate | PRC1 (CBX7), PRC2 (SUZ12) | [96–98] |
| p21 antisense | EST sequenced from neuroblastoma. Potentially oncogenic | Ago-1 | [99] |
| lincRNA-p21 | Potentially oncogenic | hnRNP-K | [100] |
| lncRNA-HEIH | Hepatocellular carcinoma | EZH2 | [101] |
| ncRNAs-Cyclin D1 | Potential tumor suppressor | p300/CBP | [102] |

who demonstrated that **miR-29 family** induces global DNA hypomethylation by decreasing DNMTs expression, through direct targeting in the case of DNMT3a and DNMT3b or indirectly in DNMT1. Restoration of miR-29 expression in cancer cell lines resulted in demethylation and consequent reexpression of p15/INK4b and ESR1 comparable to that of using DNMTs inhibitors [78, 79]. DNMT3a is also target of **miR-143**, a frequent downregulated miRNA in CRC. More significantly, enforced expression of miR-143 in colon cancer cells decreases tumor cell growth and soft-agar colony formation [83]. In addition, **miR-148** is able to repress DNMT3b expression targeting a region in its coding sequence instead of the 3′UTR as is usual in miRNA–mRNA interactions [84]. **miR-148** and **miR-152** also target DNMT1, increasing Rassf1a and p16/INK4a expression, which in turn reduce cell proliferation in cholangiocarcinoma cells [85]. **miR-342** and **miR-185** also directly target the 3′ untranslated region of DNMT1, as has been described in colorectal cancer [88] and gliomas [86]. **miR-449a** is another miRNA targeting enzymes involved in epigenetic regulation, specifically HDAC1. Thus, tumor-related downregulation of this miRNA could be one of the mechanisms responsible for the frequent HDAC1 overexpression found in several cancer types. Introduction of miR-449a into prostate cancer cells results in cell-cycle arrest, a senescence-like phenotype and apoptosis [89]. Another HDAC, HDAC4, is targeted by **miR-140**. Remarkably, blocking endogenous miR-140 partially sensitized resistant colon

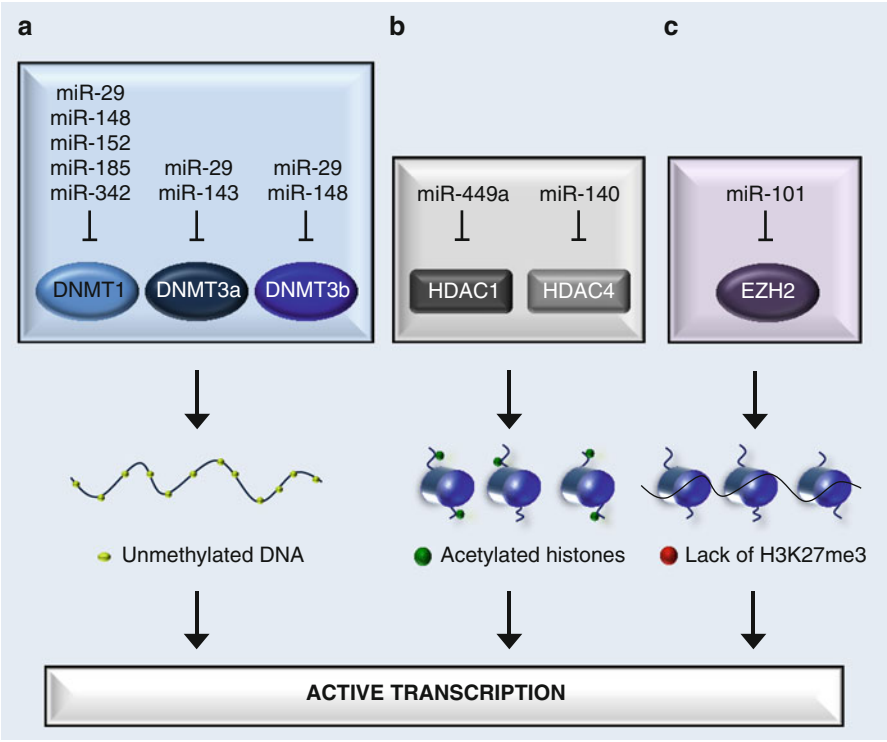


**Fig. 7.2** Noncoding RNAs as regulators of the cancer epigenome. ncRNAs are able to regulate the effectors of the epigenetic machinery. (**a**) DNA methyltransferases (DNMTs), (**b**) histone deacetylases (HDACs), and (**c**) members of the Polycomb repressive complexes, such as Enhancer of zeste homolog 2 (EZH2) are direct targets of miRNAs. Posttranscriptional inhibition of these molecules by miRNAs leads to a global deregulation of gene expression in cancer

cancer cells to 5-fluorouracil chemotherapeutic treatment [82]. In addition, HMTs are also targets of miRNAs. Studies in bladder transitional cell carcinomas and prostate tumors have revealed that EZH2, the catalytic subunit of the Polycomb repressive complex 2 (PRC2), is target of **miR-101**. Hence, the miR-101 downregulation observed in cancer could be involved in the common overexpression of EZH2 in aggressive solid tumors. Significantly, restoration of miR-101 expression attenuates cancer invasion, whilst miR-101 inhibition induces an invasive phenotype [80, 81]. Another interesting example is **miR-373,** which has been recognized as a negative regulator of the Methyl-CpG-binding Domain Protein 2 (MBD2) in hilar cholangiocarcinoma. MBD proteins bind selectively to methylated DNA and serve as a molecular link to recruit histone deacetylases (HDAC) and other transcription repression factors to the chromatin. miR-373 inhibition leads to increase of MBD2, which in turn inhibits methylation-silenced genes such as RASSF1A [40]. Another MBD controlled through miRNAs is MeCP2, targeted by **miR-212**. This miRNA has been found downregulated in gastric cancer, resulting in higher MeCP2 protein levels

that could favor the epigenetic deregulation observed in these tumors [87]. These miRNA-controlled adjustments in the epigenetic machinery could in turn affect the regulation not only of protein-coding genes, but also of ncRNAs epigenetically regulated.

In addition to the role of miRNAs as regulators of the epigenome, other short ncRNAs have been proposed as components of the epigenetic regulatory networks. **piRNAs** (PIWI-interacting RNAs) are germ line-specific ncRNAs of 24–30 nt in length that bind to Piwi proteins and guide them to their targets. These ncRNAs have been implicated in DNA methylation [103]. Previously, it had been described that loss of Piwi proteins MIWI2 (mouse piwi 2) and MILI (miwi-like) in male germ cells results in defective DNA methylation of regulatory regions of retrotransposons in a similar way to that in DNMT3L-deficient mice, indicating that the piRNA pathway plays essential roles in establishing de novo DNA methylation of retrotransposons in fetal male germ cells in mice [104, 105]. Along the same lines, Watanabe et al. have recently reported that the components of the piRNA pathway are required for de novo methylation of the differentially methylated region (DMR) of the imprinted mouse Rasgrf1 locus [103]. Although their specific functions in tumorigenesis are not clearly defined, piRNAs and PIWI proteins have been implicated in cancer [106–109]. Strikingly, piRNAs have been involved not only in testicular tumors but also are aberrantly expressed in human somatic tumors, implying a role outside germ line cells that make it worth an exhaustive research. For instance, expression of pi-651 was found upregulated in several cancer types [110], and tumor suppressive properties in human gastric cancer cells have been described for pi-823 [111]. In addition, these piRNAs have been proposed as useful biomarkers for detecting circulating gastric cancer cells [111].

A role for another type of ncRNA, **pRNAs** (promoter-associated RNAs), in the regulatory networks controlling the epigenetic state of chromatin, have also been reported. Schmitz et al. have shown that pRNA interacts with the target site of the transcription factor TTF-I, forming a DNA:RNA triplex that is specifically recognized by the DNA methyltransferase DNMT3b. More relevantly, they found that elevated levels of synthetic pRNA trigger de novo DNA methylation, heterochromatin formation and transcriptional silencing. Remarkably, as commented by the authors, recruitment of DNMT3b by DNA–RNA triplexes may be a common and generally used pathway in epigenetic regulation that could act on virtually any gene that is silenced by DNA methylation [112].

### *7.3.2 Long ncRNAs*

The abovementioned example of a triplex structure consisting of ncRNA:DNA: DNMT3b illustrates the versatility of ncRNAs, feature that makes them ideal orchestrators of epigenetic networks. In particular, long noncoding RNAs are considered flexible modular scaffolds of proteins that bring specific regulatory components into proximity with each other, which results in the formation of functional

complexes. However, as described for several ncRNAs, the model of lncRNAs as modular scaffolds should not be limited to protein interactions. Base-pairing alignment with DNA could be used to guide complexes to specific DNA sequences or bridge together sets of DNA-binding proteins. Additionally, lncRNAs could also interact with other RNAs in order to assemble complexes that can interact with proteins [113]. This attribute of lncRNAs is critical in their function as effectors of epigenetic control and gene expression regulation. The ability of lncRNAs to bind to multiple histone modifier enzymes, recruit chromatin remodeling complexes, and recruit RNA binding proteins to gene promoters is broadly recognized.

Many of the identified lncRNAs have a spatial and temporal expression patterns, indicating that lncRNA expression is strongly regulated. This fact suggests that lncRNAs have specific biological functions. In agreement with this idea, increasing evidences are demonstrating their role in a wide repertoire of biological processes, such as cell cycle control, cell differentiation, translation, splicing, etc. Considering that disruption of some of these functions has been implicated in tumorigenesis, it is expected that lncRNAs deregulation might be associated with cancer. In fact, there is growing information showing differential lncRNA expression in tumors and supporting the repercussion of lncRNA disruption in this disease. Below some examples of lncRNAs acting as epigenetic regulators with a proposed role in cancer will be discussed.

The most well-known lncRNA that acts as mediator of epigenetic modifications by recruiting chromatin remodeling complexes to specific loci is **XIST** (X-inactivation specific transcript). During early development in females, XIST transcript is expressed from the inactive X and coats the X chromosome from which it is transcribed [114]. Another lncRNA, RepA, which is transcribed from the repeat A element of the XIST locus, recruits the Polycomb repressive complex 2 (PRC2) to the XIST promoter [115]. PRC2 in turns trimethylates H3K27 creating a heterochromatic patch that leads to XIST transactivation resulting in widespread X silencing. Interestingly TSIX, another lncRNA, is transcribed from the XIST promoter in the antisense direction and regulates XIST levels during X-chromosome inactivation, at least in mice [116, 117]. Although a role for XIST in cancer has been suggested [118, 119], this issue remains controversial. While only one X chromosome is active in human cells, many tumors from both sexes have supernumerary X chromosomes and dysregulation of XIST expression has been detected. However, it remains to be determined whether the X chromosome duplications/reactivations, XIST dysregulation, and overexpression of X-linked genes are involved in cancer development or are merely a consequence of overall epigenetic instability in these cancers [120]. It is remarkable that in a mouse thymic lymphoma model, the induction of Xist results in the initiation of X inactivation and the inhibition of tumor growth, suggesting that Xist deregulation might play a significant role in tumorigenesis [121]. Moreover SATB1, a cofactor required for Xist-mediated silencing [121], is able to reprogramme chromatin organization upregulating metastasis-associated genes while downregulating tumor suppressor genes to promote tumor growth and metastasis in breast [122].

A mechanism of lncRNA-mediated epigenetic silencing similar to the aforementioned for XIST has been described for AIR and KCNQ1OT1. Both lncRNAs epigenetically silence large domains of the genome through their interaction with chromatin. In mice, **Air** is imprinted and expressed only from the paternal allele, whereas a near cluster including Igf2r, Slc22a2 and Slc22a3 protein-coding genes is expressed only from the maternal allele. Air is transcribed in antisense direction to Igf2r and Slc22a3 and is responsible for *cis*-silencing of the three paternally inherited genes, through binding to the Slc22a3 promoter and recruiting the histone lysine methyltransferase Kmt1c, which methylates H3K9 and drives the epigenetic silencing. Conversely, imprinted expression of Air itself is caused by a DNA methylation imprint acquired during oocyte development that silences the maternal Air promoter [123]. In humans, the IGF2R gene contains an intronic CpG island promoter that expresses the AIR ncRNA. The human AIR ncRNA is expressed in 16–40 % of Wilms' tumors, however the AIR-mediated silencing of IGF2R is not clear. Only in a 50 % of tumor samples, high expression of the AIR ncRNA correlated with reduced IGF2R expression [124].

A similar example is **KCNQ1OT1**, expressed from the paternal allele and responsible for the *cis*-silencing of a neighbor cluster of protein-coding genes. This lncRNA immunoprecipitates with members of the PRC2 complex (EZH2 and SUZ12) and the HMT G9, suggesting that KCNQ1OT1 is able to recruit these histone methyltransferases in *cis* to establish the repressive H3K27me3 and H3K9me3 marks, respectively [125]. A role in loss of imprinting for KCNQ1OT1 (LIT1) via epigenetic disruption has been suggested in colorectal carcinogenesis [126].

It has been proposed that as many as 20 % of lncRNAs expressed in various cell types are bound by PRC2, and that additional lincRNAs are bound by other chromatin-modifying complexes [76]. In this context, it is noteworthy that members of PRC2 complex, such as EZH2, have been associated with cancer, although contrasting evidences point to both oncogenic and tumor suppressor roles [127]. For example, Simon et al. have recently demonstrated that disruption of Ezh2 is sufficient to cause T-acute lymphoblastic leukemia (T-ALL) in mice [128]. This tumor suppressor behavior of EZH2 has been also observed in human T-ALL [129]. In contrast, EZH2 seems to have an oncogenic role in nasopharyngeal carcinoma, where it is able to form a co-repressor complex with HDAC1/HDAC2/Snail to repress E-cadherin, regulating cell invasion and metastasis [130].

One of the first lncRNAs described to have a direct implication in cancer progression by remodeling the chromatin landscape was **HOTAIR** (Hox antisense intergenic RNA) (Fig. 7.3). Among other 231 transcripts with low-coding potential, this ncRNA was identified in a comprehensive study evaluating the transcriptional activity of human HOX loci [131]. HOTAIR is a 2.2-kb spliced and polyadenylated transcript that is transcribed from the HOXC locus. This ncRNA negatively transregulates the distant HOXD locus by acting as a molecular scaffold, binding and recruiting at least two distinct histone modification complexes. The 5′ domain of the HOTAIR binds the PRC2 complex responsible for H3K27 methylation, while the 3′ region binds LSD1/CoREST/REST complex, which mediates enzymatic demethylation of H3K4me2 [90]. This lncRNA has been found highly upregulated in both primary and metastatic breast tumors. More significantly, it has been demonstrated that

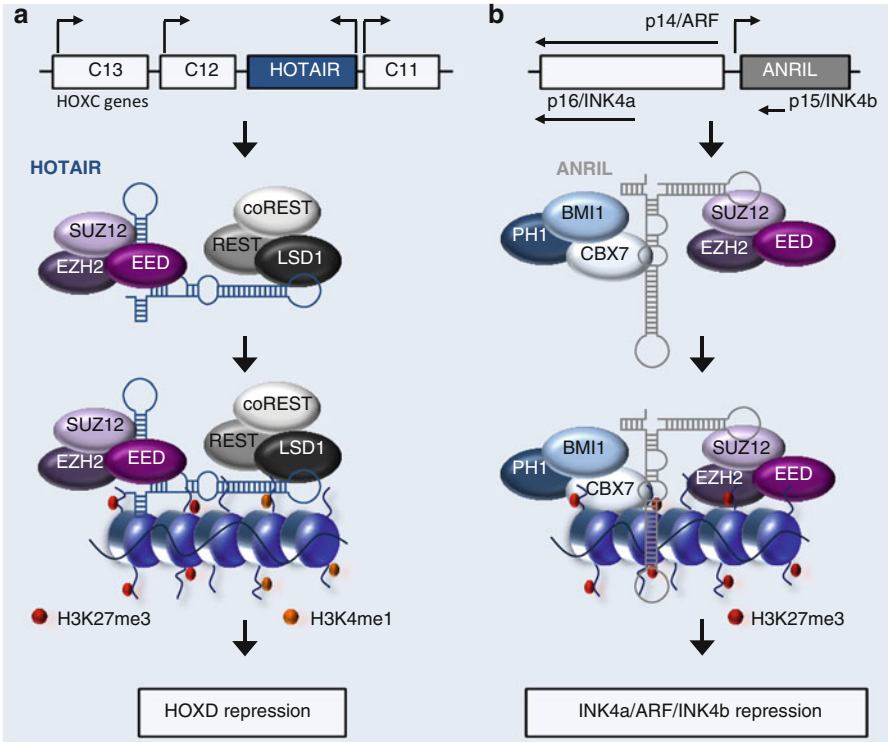


**Fig. 7.3** Role of long noncoding RNAs in chromatin remodeling in cancer. lncRNAs, like HOTAIR and ANRIL, can recruit chromatin remodeling complexes to specific genomic loci to mediate tumor suppressor gene silencing. (**a**). HOTAIR is transcribed antisense to HOXC genes and its expression is correlated with the repression of genes in the HOXD locus. The 5′ domain of the HOTAIR binds the Polycomb repressive complex 2 (PRC2) responsible for H3K27 methylation, while the 3′ region binds LSD1/CoREST/REST complex, which mediates enzymatic demethylation of H3K4me2. Recruitment of PRC2 complex to specific target genes genome-wide leads to epigenetic silencing of metastasis suppressor genes. (**b**). Long ncRNA ANRIL is transcribed antisense of the INK4a/ARF/INK4b tumor suppressor genes. ANRIL mediates gene silencing of this locus by interaction and recruitment of members of the Polycomb repressive complexes PRC1 and PRC2. CBX7, a H3K27me3-recognizing component of PRC1, can bind directly to both ANRIL and H3K27me3 via its chromodomain, and both interactions are required for CBX7 to repress the INK4a and INK4b loci

HOTAIR acts as an epigenomic reprogrammer regulating metastatic progression in breast tumors. Thus, this lncRNA recruits PRC2 complex to specific target genes genome-wide, leading to H3K27 trimethylation and epigenetic silencing of metastasis suppressor genes [91]. Remarkably, their function in cancer progression and metastasis is not restricted to breast tumors, since a significant correlation between HOTAIR overexpression and metastasis has also been detected in colorectal cancer [92], gastrointestinal stromal tumors [93], and hepatocellular carcinoma (HCC) [94]. In addition, in HCC patients who have undergone liver transplant therapy, HOTAIR has been proposed as candidate biomarker for predicting tumor

recurrence. Also, its potential as therapeutic target has been suggested in view of experiments in which siRNA-mediated suppression of this lncRNA in a liver cancer cell line increased chemotherapeutic sensitivity to cisplatin and doxorubicin [95].

**ANRIL** (Antisense noncoding RNA in the INK4 locus) (Fig. 7.3) is another example of cancer-related lncRNA. It is transcribed from the INK4a/ARF locus that encodes the p15, p16, and ARF proteins, which regulate cell cycle progression and senescence. ANRIL is transcribed antisense to the INK4b/ARF/INK4a promoter and overlaps with two exons of p15/CDKN2B. It has been reported that ANRIL overexpression results in silencing of INK4b/ARF/INK4a and p15/CDKN2B in leukemia [96] and prostate cancer [97]. ANRIL-mediated epigenetic silencing is achieved by their interaction with members of the Polycomb repressive complexes PRC1 and PRC2. Yap and colleagues have reported that CBX7, a H3K27me3-recognizing component of PRC1, can bind directly to both ANRIL and H3K27me3 via its chromodomain, and both interactions are required for CBX7 to repress the INK4a and INK4b loci [97]. Along the same lines, studies from Kotake et al. have revealed that ANRIL binds and is required for the recruitment of SUZ12, a component of PRC2, to p15/INK4B locus [98]. p15/INK4B silencing has been correlated with an increase in the repressive mark H3K9me2 and a decrease of the active mark H3K4me2 at the promoter region, but no increase in DNA methylation has been observed [96]. These studies revealed the critical role of ANRIL-*cis*-mediated epigenetic silencing of a well-recognized cancer-related locus. It is noteworthy that specific SNPs in ANRIL have been correlated with an increased susceptibility to cancer and other diseases. For instance, SNPs in ANRIL influence the number of plexiform neurofibromas in Neurofibromatosis type 1 (NF1), a tumor predisposition syndrome [132].

LncRNA-mediated silencing of tumor suppressor genes has also been reported as a mechanism controlling the cell cycle regulator p21/CDKN1A. Morris et al. described the bidirectional transcription at the p21 genomic locus where p21 antisense (**p21AS**) acts as effector molecule driving transcriptional gene silencing of p21. p21AS maintains low-level epigenetic silencing by direct recruitment of Argonaute 1 (Ago-1) and the repressive mark H3K27me3 to the p21 sense promoter [99]. Although p21AS has not been associated to cancer thus far, the direct consequence of an imbalance in endogenous bidirectional transcription could potentially lead to the deregulation of p21 expression and contribute to tumorigenesis, specially taking into account the crucial role of p21 in cell cycle control. Additionally, another ncRNA located approximately 15 Kb upstream of the p21/CDKN1A gene has been identified. The so-called **lincRNA-p21**, (due to its proximity to p21 locus), is transcribed from an independent promoter in the opposite orientation to the p21/CDKN1A gene. This transcript was identified as a p53-activated lincRNA whose binding to hnRNP-K (heterogeneous nuclear ribonucleoprotein K) is required for the proper localization of the ribonucleoprotein [100]. hnRNP-K had been previously described as a key component of a repressor complex that acts in the p53 pathway [133] and lincRNA-p21 was found to be a global repressor of genes in the p53 pathway, playing an important role in the p53-dependent induction of apoptosis [100].

Collectively, these examples highlight the lncRNA-mediated epigenetic silencing as an important mechanism driving critical cancer-related pathways.

Recently, a specific lncRNA High Expressed In Hepatocellular Carcinoma (HCC), called **lncRNA-HEIH**, has been identified. Interestingly, the expression level of this lncRNA has been significantly associated with recurrence and has been proposed as an independent prognostic factor for survival in hepatitis B virus-related HCC. This lncRNA is able to associate with EZH2 and experimental evidences have demonstrated that this association is required for the repression of EZH2 target genes in HCC progression [101].

In addition to binding to and recruiting chromatin remodeling complexes and histone modifier enzymes, the ability of lncRNAs to regulate gene expression is also achieved through their capacity to recruit RNA binding proteins and/or regulate its accessibility to gene promoters. ncRNAs transcribed from 5′ end of the Cyclin D1 gene are significant examples. Cyclin D1/CCND1 is a cell cycle regulator frequently disrupted in several cancer types [134]. In response to DNA damage, **ncRNAs-Cyclin D1** are expressed and interact with the TLS (Translocated in Liposarcoma) protein, inducing an allosteric modification that allows its association with the CCND1 promoter. Recruitment of TLS to the promoter causes Cyclin D1 repression by inhibition of enzymatic activities of the histone acetyltransferases CREB-binding protein (CBP) and p300 [102].

Another example is an lncRNA transcribed from the DHFR (Dihydrofolate Reductase) minor promoter. This **lncRNA-DHFR** has a key function in an epigenetic mechanism of promoter-specific transcriptional repression of the gene encoding DHFR. The lncRNA forms a triplex structure with the major promoter while interacting directly with the transcription factor TFIIB, which results in the disruption of the preinitiation complex at the major promoter [135]. Considering the role of DHFR in the synthesis of DNA precursors, competitive inhibitors of DHFR (i.e., methotrexate) are used in anticancer therapy in order to limit the growth and proliferation of tumor cells [136]. Therefore, although so far a link between lncRNA-DHFR and cancer has not been identified, this represents a topic worthy of future research.

Finally, ncRNAs can also function as molecular “decoys” by preventing correct regulation through competitive binding. For instance, **GAS5** (Growth arrest-specific 5) is an ncRNA whose conformation mimics that of the glucocorticoid responsive element (GRE) DNA. Consequently, GAS5 is able to bind to the DNA-binding domain of the glucocorticoid receptor (GR) by acting as a decoy glucocorticoid response element (GRE). The competition with DNA GREs for binding to the receptor blocks the transcriptional induction by GR of the target genes. GAS5 is highly expressed in cells whose growth has been arrested because of lack of nutrients or growth factors and sensitizes cells to apoptosis by suppressing glucocorticoid-mediated induction of several responsive genes [137]. Interestingly, GAS5 has been found downregulated in breast cancer and some introns of GAS5 encode snoRNAs that sensitize mammalian cells to apoptosis, supporting its tumor suppressor role [138]. Overall, multiple examples have been discussed emphasizing the function of ncRNAs as regulators of the cancer epigenome.

## 7.4 Perspectives

The complex network of reciprocal interactions between ncRNAs and epigenetics is just beginning to unravel. The exponential increase in the number of studies and publications in the ncRNA field highlights the fast advance of our knowledge in this matter, directly promoted by technological and bioinformatics advances at genome-wide level. In general, it has been demonstrated that the more complex an organism, the greater the number of its ncRNAs. Moreover, gene expression regulation by ncRNAs has significant advantages, including rapid response speed (attributable to the proximity between ncRNA production and the target gene), or a lower energetic cost to the cell (due to the lack of protein synthesis), among others. Furthermore, their molecular structure consisting of primary nucleotide sequences leads to more plastic structure–function constraints than proteins. Collectively, these facts support the proposition that expansion of regulatory RNAs was fundamental in cellular reprogramming and ultimately in the evolution of eukaryotes. Although until quite recently argued to be spurious transcriptional noise, the biological significance and critical roles of ncRNA in cellular function are now supported by increasingly strong evidences. Versatility of these molecules makes them not only ideal orchestrators of biological networks, but also key components in tumorigenesis. An exciting future in research about the role of ncRNAs in cancer epigenetics is guaranteed.

**Acknowledgements** V.D. is supported by Instituto de Salud Carlos III, Sara Borrell postdoctoral contract. M.E. is an Institucio Catalana de Recerca i Estudis Avançats (ICREA) Research Professor.

## References

1. Waddington CH. Preliminary notes on the development of the wings in normal and mutant strains of Drosophila. Proc Natl Acad Sci U S A. 1939;25(7):299–307.
2. Holliday R. The inheritance of epigenetic defects. Science. 1987;238(4824):163–70.
3. Antequera F. Structure, function and evolution of CpG island promoters. Cell Mol Life Sci. 2003;60(8):1647–58.
4. Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358(11):1148–59.
5. Esteller M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene. 2002;21(35):5427–40.
6. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000; 403(6765):41–5.
7. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.
8. Chi P, Allis CD, Wang GG. Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer. 2010;10(7):457–69.
9. Margueron R, Reinberg D. The polycomb complex PRC2 and its mark in life. Nature. 2011;469(7330):343–9.
10. Saito Y, Liang G, Egger G, et al. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell. 2006;9(6):435–43.
11. Scott GK, Mattie MD, Berger CE, et al. Rapid alteration of microRNA levels by histone deacetylase inhibition. Cancer Res. 2006;66(3):1277–81.

12. Lujambio A, Ropero S, Ballestar E, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res. 2007;67(4):1424–9.
13. Silber J, Lim DA, Petritsch C, et al. miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. BMC Med 2008;6:14.
14. Ando T, Yoshida T, Enomoto S, et al. DNA methylation of microRNA genes in gastric mucosae of gastric cancer patients: its possible involvement in the formation of epigenetic field defect. Int J Cancer. 2009;124(10):2367–74.
15. Agirre X, Vilas-Zornoza A, Jimenez-Velasco A, et al. Epigenetic silencing of the tumor suppressor microRNA Hsa-miR-124a regulates CDK6 expression and confers a poor prognosis in acute lymphoblastic leukemia. Cancer Res. 2009;69(10):4443–53.
16. Roman-Gomez J, Agirre X, Jimenez-Velasco A, et al. Epigenetic regulation of microRNAs in acute lymphoblastic leukemia. J Clin Oncol. 2009;27(8):1316–22.
17. Furuta M, Kozaki KI, Tanaka S, et al. miR-124 and miR-203 are epigenetically silenced tumor-suppressive microRNAs in hepatocellular carcinoma. Carcinogenesis. 2010;31(5):766–76.
18. Wilting SM, van Boerdonk RA, Henken FE, et al. Methylation-mediated silencing and tumour suppressive function of hsa-miR-124 in cervical cancer. Mol Cancer. 2010;9167.
19. Brueckner B, Stresemann C, Kuner R, et al. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. Cancer Res. 2007;67(4):1419–23.
20. Lu L, Katsaros D, de la Longrais IA, et al. Hypermethylation of let-7a-3 in epithelial ovarian cancer is associated with low insulin-like growth factor-II expression and favorable prognosis. Cancer Res. 2007;67(21):10117–22.
21. Lu L, Katsaros D, Zhu Y, et al. Let-7a regulation of insulin-like growth factors in breast cancer. Breast Cancer Res Treat. 2011;126(3):687–94.
22. Datta J, Kutay H, Nasser MW, et al. Methylation mediated silencing of MicroRNA-1 gene and its role in hepatocellular carcinogenesis. Cancer Res. 2008;68(13):5049–58.
23. Lehmann U, Hasemeier B, Christgen M, et al. Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. J Pathol. 2008;214(1):17–24.
24. Lujambio A, Calin GA, Villanueva A, et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A. 2008;105(36):13556–61.
25. Bandres E, Agirre X, Bitarte N, et al. Epigenetic regulation of microRNA expression in colorectal cancer. Int J Cancer. 2009;125(11):2737–43.
26. Hildebrandt MA, Gu J, Lin J, et al. Hsa-miR-9 methylation status is associated with cancer development and metastatic recurrence in patients with clear cell renal cell carcinoma. Oncogene. 2010;29(42):5724–8.
27. Heller G, Weinzierl M, Noll C, et al. Genome-wide miRNA expression profiling identifies miR-9-3 and miR-193a as targets for DNA methylation in Non-small cell lung cancers. Clin Cancer Res. 2012;18(6):1619–29.
28. Sampath D, Liu C, Vasan K, et al. Histone deacetylases mediate the silencing of miR-15a, miR-16, and miR-29b in chronic lymphocytic leukemia. Blood. 2012;119(5):1162–72.
29. Humphreys KJ, Cobiac L, Le Leu RK, et al. Histone deacetylase inhibition in colorectal cancer cells reveals competing roles for members of the oncogenic miR-17-92 cluster. Mol Carcinog. 2013;52(6):459–74. doi:10.1002/mc.21879.
30. Augoff K, Mccue B, Plow EF, et al. miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. Mol Cancer 2012;11:5.
31. Toyota M, Suzuki H, Sasaki Y, et al. Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. Cancer Res. 2008;68(11):4123–32.
32. Chim CS, Wan TS, Wong KY, et al. Methylation of miR-34a, miR-34b/c, miR-124-1 and miR-203 in Ph-negative myeloproliferative neoplasms. J Transl Med. 2011;9:197.
33. Tsai KW, Wu CW, Hu LY, et al. Epigenetic regulation of miR-34b and miR-129 expression in gastric cancer. Int J Cancer. 2011;129(11):2600–10.

34. Lee KH, Lotterman C, Karikari C, et al. Epigenetic silencing of MicroRNA miR-107 regulates cyclin-dependent kinase 6 expression in pancreatic cancer. Pancreatology. 2009;9(3):293–301.
35. Huang YW, Liu JC, Deatherage DE, et al. Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 oncogene in endometrial cancer. Cancer Res. 2009;69(23):9038–46.
36. Shen R, Pan S, Qi S, et al. Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 in gastric cancer. Biochem Biophys Res Commun. 2010;394(4):1047–52.
37. Balaguer F, Link A, Lozano JJ, et al. Epigenetic silencing of miR-137 is an early event in colorectal carcinogenesis. Cancer Res. 2010;70(16):6609–18.
38. Langevin SM, Stone RA, Bunker CH, et al. MicroRNA-137 promoter methylation in oral rinses from patients with squamous cell carcinoma of the head and neck is associated with gender and body mass index. Carcinogenesis. 2010;31(5):864–70.
39. Chen Q, Chen X, Zhang M, et al. miR-137 is frequently down-regulated in gastric cancer and is a negative regulator of Cdc42. Dig Dis Sci. 2011;56(7):2009–16.
40. Chen Y, Luo J, Tian R, et al. miR-373 negatively regulates methyl-CpG-binding domain protein 2 (MBD2) in hilar cholangiocarcinoma. Dig Dis Sci. 2011;56(6):1693–701.
41. Langevin SM, Stone RA, Bunker CH, et al. MicroRNA-137 promoter methylation is associated with poorer overall survival in patients with squamous cell carcinoma of the head and neck. Cancer. 2011;117(7):1454–62.
42. Wiklund ED, Gao S, Hulf T, et al. MicroRNA alterations and associated aberrant DNA methylation patterns across multiple sample types in oral squamous cell carcinoma. PLoS One. 2011;6(11):e27840.
43. Kozaki K, Imoto I, Mogi S, et al. Exploration of tumor-suppressive microRNAs silenced by DNA hypermethylation in oral cancer. Cancer Res. 2008;68(7):2094–105.
44. Gao XN, Lin J, Li YH, et al. MicroRNA-193a represses c-kit expression and functions as a methylation-silenced tumor suppressor in acute myeloid leukemia. Oncogene. 2011;30(31): 3416–28.
45. Schotte D, Lange-Turenhout EA, Stumpel DJ, et al. Expression of miR-196b is not exclusively MLL-driven but is especially linked to activation of HOXA genes in pediatric acute lymphoblastic leukemia. Haematologica. 2010;95(10):1675–82.
46. Tsai KW, Hu LY, Wu CW, et al. Epigenetic regulation of miR-196b expression in gastric cancer. Genes Chromosomes Cancer. 2010;49(11):969–80.
47. Ueda T, Volinia S, Okumura H, et al. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncol. 2010; 11(2):136–46.
48. Yin G, Chen R, Alvero AB, et al. TWISTing stemness, inflammation and proliferation of epithelial ovarian cancer cells through MIR199A2/214. Oncogene. 2010;29(24):3545–53.
49. Cheung HH, Davis AJ, Lee TL, et al. Methylation of an intronic region regulates miR-199a in testicular tumor malignancy. Oncogene. 2011;30(31):3404–15.
50. Ceppi P, Mudduluru G, Kumarswamy R, et al. Loss of miR-200c expression induces an aggressive, invasive, and chemoresistant phenotype in non-small cell lung cancer. Mol Cancer Res. 2010;8(9):1207–16.
51. Neves R, Scheel C, Weinhold S, et al. Role of DNA methylation in miR-200c/141 cluster silencing in invasive breast cancer cells. BMC Res Notes. 2010;3:219.
52. Vrba L, Jensen TJ, Garbe JC, et al. Role for DNA methylation in the regulation of miR-200c and miR-141 expression in normal and cancer cells. PLoS One. 2010;5(1):e8697.
53. Davalos V, Moutinho C, Villanueva A, et al. Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. Oncogene. 2011;31(16):2062–74.
54. Wiklund ED, Bramsen JB, Hulf T, et al. Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. Int J Cancer. 2011;128(6):1327–34.
55. Craig VJ, Cogliatti SB, Rehrauer H, et al. Epigenetic silencing of microRNA-203 dysregulates ABL1 expression and drives Helicobacter-associated gastric lymphomagenesis. Cancer Res. 2011;71(10):3616–24.

56. Chim CS, Wong KY, Leung CY, et al. Epigenetic inactivation of the hsa-miR-203 in haematological malignancies. J Cell Mol Med. 2011;15(12):2760–7.
57. Tellez CS, Juri DE, Do K, et al. EMT and stem cell-like properties associated with miR-205 and miR-200 epigenetic silencing are early manifestations during carcinogen-induced transformation of human lung epithelial cells. Cancer Res. 2011;71(8):3087–97.
58. Wang Y, Toh HC, Chow P, et al. MicroRNA-224 is up-regulated in hepatocellular carcinoma through epigenetic mechanisms. FASEB J. 2012;26(7):3032–41. doi:10.1096/fj.11-201855.
59. Ferreira HJ, Heyn H, Moutinho C, et al. CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer. RNA Biol. 2012;9(6):881–90.
60. Cheung HH, Lee TL, Davis AJ, et al. Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. Br J Cancer. 2010;102(2):419–27.
61. Lujambio A, Portela A, Liz J, et al. CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. Oncogene. 2010;29(48):6390–401.
62. Zhao J, Dahle D, Zhou Y, et al. Hypermethylation of the promoter region is associated with the loss of MEG3 gene expression in human pituitary tumors. J Clin Endocrinol Metab. 2005; 90(4):2179–86.
63. Benetatos L, Dasoula A, Hatzimichael E, et al. Promoter hypermethylation of the MEG3 (DLK1/MEG3) imprinted gene in multiple myeloma. Clin Lymphoma Myeloma. 2008; 8(3):171–5.
64. Gejman R, Batista DL, Zhong Y, et al. Selective loss of MEG3 expression and intergenic differentially methylated region hypermethylation in the MEG3/DLK1 locus in human clinically nonfunctioning pituitary adenomas. J Clin Endocrinol Metab. 2008;93(10): 4119–25.
65. Benetatos L, Hatzimichael E, Dasoula A, et al. CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. Leuk Res. 2010;34(2):148–53.
66. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of microRNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524–9.
67. Suzuki H, Takatsuka S, Akashi H, et al. Genome-wide profiling of chromatin signatures reveals epigenetic regulation of MicroRNA genes in colorectal cancer. Cancer Res. 2011; 71(17):5646–58.
68. Gerbi SA. Small nucleolar RNA. Biochem Cell Biol. 1995;73(11–12):845–58.
69. Kiss-Laszlo Z, Henry Y, Bachellerie JP, et al. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell. 1996;85(7):1077–88.
70. Ganot P, Bortolin ML, Kiss T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell. 1997;89(5):799–809.
71. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74.
72. Calin GA, Liu CG, Ferracin M, et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell. 2007;12(3):215–29.
73. Scaruffi P, Stigliani S, Moretti S, et al. Transcribed-ultra conserved region expression is associated with outcome in high-risk neuroblastoma. BMC Cancer. 2009;9:441.
74. Zhang X, Zhou Y, Mehta KR, et al. A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. J Clin Endocrinol Metab. 2003;88(11):5119–26.
75. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–7.
76. Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009;106(28):11667–72.
77. Wu SC, Kallin EM, Zhang Y. Role of H3K27 methylation in the regulation of lncRNA expression. Cell Res. 2010;20(10):1109–16.

78. Fabbri M, Garzon R, Cimmino A, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci U S A. 2007;104(40):15805–10.
79. Garzon R, Liu S, Fabbri M, et al. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. Blood. 2009;113(25):6411–8.
80. Varambally S, Cao Q, Mani RS, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science. 2008;322(5908):1695–9.
81. Friedman JM, Liang G, Liu CC, et al. The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. Cancer Res. 2009;69(6):2623–9.
82. Song B, Wang Y, Xi Y, et al. Mechanism of chemoresistance mediated by miR-140 in human osteosarcoma and colon cancer cells. Oncogene. 2009;28(46):4065–74.
83. Ng EK, Tsang WP, Ng SS, et al. MicroRNA-143 targets DNA methyltransferases 3A in colorectal cancer. Br J Cancer. 2009;101(4):699–706.
84. Duursma AM, Kedde M, Schrier M, et al. miR-148 targets human DNMT3b protein coding region. RNA. 2008;14(5):872–7.
85. Braconi C, Huang N, Patel T. MicroRNA-dependent regulation of DNA methyltransferase-1 and tumor suppressor gene expression by interleukin-6 in human malignant cholangiocytes. Hepatology. 2010;51(3):881–90.
86. Zhang Z, Tang H, Wang Z, et al. MiR-185 targets the DNA methyltransferases 1 and regulates global DNA methylation in human glioma. Mol Cancer. 2011;10:124.
87. Wada R, Akiyama Y, Hashimoto Y, et al. miR-212 is downregulated and suppresses methyl-CpG-binding protein MeCP2 in human gastric cancer. Int J Cancer. 2010;127(5):1106–14.
88. Wang H, Wu J, Meng X, et al. MicroRNA-342 inhibits colorectal cancer cell proliferation and invasion by directly targeting DNA methyltransferase 1. Carcinogenesis. 2011;32(7): 1033–42.
89. Noonan EJ, Place RF, Pookot D, et al. miR-449a targets HDAC-1 and induces growth arrest in prostate cancer. Oncogene. 2009;28(14):1714–24.
90. Tsai MC, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689–93.
91. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6.
92. Kogo R, Shimamura T, Mimori K, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011;71(20):6320–6.
93. Niinuma T, Suzuki H, Nojima M, et al. Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. Cancer Res. 2012;72(5):1126–36.
94. Geng YJ, Xie SL, Li Q, et al. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res. 2011;39(6):2119–28.
95. Yang Z, Zhou L, Wu LM, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011;18(5):1243–50.
96. Yu W, Gius D, Onyango P, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature. 2008;451(7175):202–6.
97. Yap KL, Li S, Munoz-Cabello AM, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell. 2010;38(5):662–74.
98. Kotake Y, Nakagawa T, Kitagawa K, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 2011;30(16):1956–62.
99. Morris KV, Santoso S, Turner AM, et al. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet. 2008;4(11):e1000258.

100. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010;142(3):409–19.
101. Yang F, Zhang L, Huo XS, et al. Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. Hepatology. 2011;54(5):1679–89.
102. Wang X, Arai S, Song X, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. 2008;454(7200):126–30.
103. Watanabe T, Tomizawa S, Mitsuya K, et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science. 2011;332(6031): 848–52.
104. Aravin AA, Sachidanandam R, Bourc'his D, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell. 2008;31(6):785–99.
105. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 2008;22(7):908–17.
106. Lee JH, Schutte D, Wulf G, et al. Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/Bcl-XL pathway. Hum Mol Genet. 2006; 15(2):201–11.
107. Liu X, Sun Y, Guo J, et al. Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. Int J Cancer. 2006;118(8):1922–9.
108. Taubert H, Greither T, Kaushal D, et al. Expression of the stem cell self-renewal gene Hiwi and risk of tumour-related death in patients with soft-tissue sarcoma. Oncogene. 2007;26(7):1098–100.
109. Liu JJ, Shen R, Chen L, et al. Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. Int J Clin Exp Pathol. 2010;3(4):328–37.
110. Cheng J, Guo JM, Xiao BX, et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. Clin Chim Acta. 2011;412(17–18):1621–5.
111. Cheng J, Deng H, Xiao B, et al. piR-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells. Cancer Lett. 2012;315(1):12–7.
112. Schmitz KM, Mayer C, Postepska A, et al. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. 2010;24(20):2264–9.
113. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482(7385):339–46.
114. Penny GD, Kay GF, Sheardown SA, et al. Requirement for Xist in X chromosome inactivation. Nature. 1996;379(6561):131–7.
115. Zhao J, Sun BK, Erwin JA, et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322(5902):750–6.
116. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet. 1999;21(4):400–4.
117. Ohhata T, Hoki Y, Sasaki H, et al. Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development. 2008;135(2):227–35.
118. Huang KC, Rao PH, Lau CC, et al. Relationship of XIST expression and responses of ovarian cancer to chemotherapy. Mol Cancer Ther. 2002;1(10):769–76.
119. Sirchia SM, Tabano S, Monti L, et al. Misbehaviour of XIST RNA in breast cancer cells. PLoS One. 2009;4(5):e5559.
120. Weakley SM, Wang H, Yao Q, et al. Expression and function of a large non-coding RNA gene XIST in human cancer. World J Surg. 2012;35(8):1751–6.
121. Agrelo R, Souabni A, Novatchkova M, et al. SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. Dev Cell. 2009;16(4):507–16.
122. Han HJ, Russo J, Kohwi Y, et al. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. Nature. 2008;452(7184):187–93.

123. Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature. 2002;415(6873):810–3.
124. Yotova IY, Vlatkovic IM, Pauler FM, et al. Identification of the human homolog of the imprinted mouse Air non-coding RNA. Genomics. 2008;92(6):464–73.
125. Pandey RR, Mondal T, Mohammad F, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell. 2008; 32(2):232–46.
126. Nakano S, Murakami K, Meguro M, et al. Expression profile of LIT1/KCNQ1OT1 and epigenetic status at the KvDMR1 in colorectal cancers. Cancer Sci. 2006;97(11):1147–54.
127. Hock H. A complex Polycomb issue: the two faces of EZH2 in cancer. Genes Dev. 2012; 26(8):751–5.
128. Simon C, Chagraoui J, Krosl J, et al. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. Genes Dev. 2012;26(7):651–6.
129. Ntziachristos P, Tsirigos A, Van Vlierberghe P, et al. Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med. 2012;18(2):298–301.
130. Tong ZT, Cai MY, Wang XG, et al. EZH2 supports nasopharyngeal carcinoma cell aggressiveness by forming a co-repressor complex with HDAC1/HDAC2 and Snail to inhibit E-cadherin. Oncogene. 2012;31(5):583–94.
131. Rinn JL, Kertesz M, Wang JK, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129(7):1311–23.
132. Pasmant E, Sabbagh A, Masliah-Planchon J, et al. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. J Natl Cancer Inst. 2011;103(22): 1713–22.
133. Kim K, Choi J, Heo K, et al. Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. J Biol Chem. 2008;283(14):9113–26.
134. Diehl JA. Cycling to cancer with cyclin D1. Cancer Biol Ther. 2002;1(3):226–31.
135. Martianov I, Ramadass A, Serra Barros A, et al. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 2007;445(7128):666–70.
136. McGuire JJ. Anticancer antifolates: current status and future directions. Curr Pharm Des. 2003;9(31):2593–613.
137. Kino T, Hurt DE, Ichijo T, et al. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal. 2010;3(107):ra8.
138. Mourtada-Maarabouni M, Pickard MR, Hedge VL, et al. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene. 2009;28(2):195–208.

# Chapter 8
# MicroRNA SNPs in Cancer

**Yujing Li and Peng Jin**

**Abstract** MiRNAs could serve as the regulators of cell events, including such as differentiation, propagation, and apoptosis. MiRNAs could act as natural oncogenes or tumor suppressor genes. Whether a particular miRNA serves as either could almost be moot when the additional problems of SNPs enter the fray. A miRNA involved with SNPs (miR-SNPs) on any regulatory level, whether naturally cancer-inducing or not, could easily undergo an oncogenic transformation. This chapter reviews targets of miRNAs and the miRNAs themselves frequently containing SNPs reflecting different risks and markers of cancer with emphasis on familial groups and populations of shared heredity.

**Keywords** MicroRNAs • SNPs • Cancer

## 8.1 Introduction

MiRNAs are short (19–25 nucleotides long in humans [1]), evolutionarily conserved RNA structures that can bind to the mRNA of protein coding genes. Most often miRNAs bind within the 3′ untranslated region (UTR) [2, 3] though there are cases in which miRNAs can bind the 5′ UTR [4] and even within coding regions of some mRNAs [4, 5]. MiRNAs, like mRNAs, possess 5′ phosphate groups and 3′ hydroxyl termini. They are often found within introns of coding genes, but some miRNAs have their expression driven by a separate promoter. Nucleotides seven to eight bases in length, often referred to as seed sites, are located at the 5′ end of the mature form of miRNA [6]. The seed sites are usually

Y. Li • P. Jin (✉)
Department of Human Genetics, Emory University School of Medicine,
Atlanta, GA 30322, USA
e-mail: yli29@emory.edu; peng.jin@emory.edu

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_8,

defined by evolutionary comparison; they are generally conserved among distantly related species, though there is a weaker evolutionary conservation with the 3′ end of miRNAs as the tail is marginally involved in binding to the target site [7]. It has been computationally estimated that up to 30–60 % of genes can have their expression altered by the presence or absence of miRNAs that bind to the seed site target located on the mRNA. Single nucleotide polymorphisms (SNPs) are base pair changes with DNA that occur with a frequency of about 1 in 12,500 base pair or at approximately 99 % of the sites in which the same residue is present on both homologues of chromosomes [8]. SNPs are by far the most common form of mutation in the human genome. SNPs serve as guides to delineate possible markers for disease causing loci, or the loci themselves in databases such as HapMap. SNPs have typically been used for cancer-association studies in different ways. One involves direct examination of genes known to be involved in the cancer pathway; these studies are not always fruitful as they lack statistical power and are limited to a few genes known to interact with a specific oncogene or tumor-suppressor gene [9]. The other uses genome-wide association studies (GWAS) in order to examine cancer association within a large population or SNPs can work with and within miRNAs to influence translational control of mRNAs [9–11]. SNPs are able in some cases to generate or abolish miRNA binding sites [12, 13]. SNPs have also been credited as activating miRNAs to become oncogenes or tumor-suppressors [14, 15].

These SNP base pair changes, whether within the target site of the miRNA or the miRNA itself have been associated with many cancers, both in vitro and in vivo [16–18]. It is not likely coincidental that about half of all miRNAs are located at fragile sites as well as sites known to be involved in cancer [19]. This chapter largely covers the interaction of miRNAs with their target sites, but it should be noted that miRNA containing polymorphic SNPs can affect transcription of the primary transcript, and additionally, how the precursor-miRNA interacts with downstream miRNA processing proteins. After screening more than a 100 tumor tissues representative of 20 cancers, the expression of one miRNA, *let-7e*, was significantly downregulated in vivo when a SNP transforming an A to a G (A>G) 17 bp downstream of the miRNA was examine [20]. Though this was not a bioinformatics study it demonstrated that SNPs within the pri- or pre-regions of miRNA could affect miRNA processing. Exact knowledge on the manufacture of the atypical expression remains elusive [20].

Currently, the in silico prediction of miRNA interaction with a purported target site does not always agree with in vivo studies, though these predictions do lead to further avenues of exploration via in vivo studies and effective case–control studies [21]. The association of population-based SNPs with cancers, however, is a somewhat contested issue. It has been suggested that many of the population sizes used to measure the connotation with SNPs and cancer are not large enough to make some of the claims of association and as such more careful case–control studies are needed [22]. Also imperative is the need to link bioinformatics, in vitro examination, in vivo research and large case–control studies [23].

## 8.2 MiRNAs and Their Relationship to Cancer

Many miRNAs have been associated with certain cancer phenotypes. The first known reporting of miRNAs and their association with some cancers was shown in Calin et al. [24]. This study showed a deletion of *miR-15a* and *miR-17-92* in chronic lymphocytic leukemia (CLL). This group further demonstrated that a germ-line mutation in the *pri-miR-16-1* results in down-regulation of the miRNA both in vivo and in vitro [24]. Other studies have also shown linkage between specific and non-specific cancers [17]. For instance, the miR-19-92 cluster is frequently found rearranged within lymphomas [25] and the *miR-17-92* cluster is found to be highly expressed in a variety of tumors [26, 27] and is associated with the binding of c-myc to E-boxes for activation of transcription [28]. In vivo and in vitro studies confirm *miR-130a* targets transcription factor V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (*MAFB*) and that depletion of *miR-10a* upregulates *HOXA1* expression. It was also shown that *miR-10a* directly targets the 3′ UTR of *HOXA1* RNA [28].

Conversely, leukemic megakaryocytes show upregulation of *miR-101*, *miR-126*, *miR-99a*, *miR-135*, *miR-20* [28]. Additional works have pointed to miRNA differential expression leading to context dependent effects in some cancers. Expression signatures of cancer gene targets within solid tumors are also beginning to be explored [29] and recently solid tumors were used for deep sequencing and discovery of new miRNA SNP regions [30]. However it is unknown whether these novel sequences will shed light on SNP regions that are differentially expressed across cancers within the same familial clades.

### *8.2.1 Breast Cancer*

A SNP in the precursor form of *miR-146a* could be a target for predicting age of onset for both ovarian and breast cancer [31] though there is some doubt about the case–control methodologies [32]. A SNP in the gene antecedent (rs2910164) changing a G>U pair to a C>U pair in the stem region was recently associated with age of onset of breast cancer (BC) and ovarian cancer (OC) in unrelated groups [31]. In vitro analysis demonstrated that the rare SNP variant binds the 3′ UTR of *BRCA1* more commonly than the more common allele. This study suggests that the *miR-146a* mutant precursor may be concomitant with ovarian cancer and breast cancer [31]. A later study showed that the *miR-146a* pre-miRNA rs2910164 C>G allele was in Hardy–Weinberg equilibrium with the rest of the comparative population with in a case–control study among Chinese women [32]. Other studies point toward bioinformatics methodologies that could shed light on both miRNAs and their target sites with a role in cancer [33–35]. Recently, in a case–control study involving unrelated Chinese women of Han ethnicity, two out of four pre-miRNAs studied have shown to have significance with increased risk of BC. The Hu et al. study

indicated that *hsa-mir-196a2* rs11614913: T>C and *hsa-mir-499* rs3746444: A>G were distributed more heavily in women of like descent [32]. The research also points out two genes, *LSP1* and *TOX3*, according to GWAS studies, are associated with *hsa-mir-196a2-3p* and *hsa-mir-196a2-5p* as newly identified BC susceptibility markers [32].

In a smaller case study, the *estrogen receptor 1* (*ESR1*) protein product has been shown to affect BC risk in women; based on a study predicting polymorphic SNPs effect on gene expression [36] an *ESR1* miRNA binding site was examined for association with BC onset [37]. The populations amassed for the study included familial BC cases and isolated cases of early onset BC. A minor allele of *ESR1* (*ESR1* rs2747648T>C) within a predicted *miR-453* binding site was negatively correlated with premenopausal women and the onset of BC [37]. The allele (T>C) has a protective effect against BC, even more so in cases of familial BC and when the C allele was present in the homozygous condition.

*BRCA1* and *BRCA2* gene mutations are involved in a majority of BCs and OCs [38, 39], however, both their penetrance and expressivity are questioned as neither gene (or both genes together) can truly predict an accurate outcome of patient disease onset [40, 41]. Kontorovich et al. examined a population of Jewish women at risk for BC and OC in a case–control study [42]. Specifically they address both miRNA binding site SNPs as well as SNPs with the miRNAs themselves. This is an interesting revelation in that three *BRCA2* SNPs within miRNA binding sites were found to have different modalities in their effect on BC and OC onset [42]. Two miRNA precursor SNPs, rs6505162 and rs895819 are associated weakly with cancer risk. The *hsa-mir-423* SNP rs6505162 is unusual in that it is located outside of the mature product, but various RNA folding programs are unable to predict any other genes with which this miRNA with its mutant SNP interact [42].

Another SNP within the *activating transcription factor 1* (*ATF1*) gene miRNA-binding site, rs11169571 is strongly associated with onset of cancer, but its mechanism is also unknown. Rs11169571 has an affinity for binding the *hsa-mir-320* family and the heterozygote SNPs have an approximately twofold risk for developing BC and OC [42]. One line of reasoning suggests that the *miR-320* family binds a particular SNP blocking access to many other miRNAs that could in theory seek presentation to ATF1. The rs895819 SNP in its heterozygous form has a much lower rate of cancer and is located within the *has-mir-27a* pre-miRNA. Once again, RNA fold programs cannot predict the relevance of this particular SNP and its attachment to BC and OC [42]. Nicoloso et al. looked at SNPs that interrupt miRNA target sites [16]. The study found that BC associated SNPs within different populations at risk for developing BC (*BCRA1* rs799917 and *TGFR1* rs334348) present in differing amounts within somatic DNA. *TGFB1* SNP rs982073, associating with *miR-187* and *XRCC1* rs1799782, associating with *miR-138* possess the ability to alter expression by changing the target sites of each miRNA [16]. An A>G SNP is also found related BC, though its exact mechanism of action is unknown [43]. Another SNP, rs89519, was assessed with a reduction in BC within related individuals, though again the achievement of this SNP to aid in circumventing BC is a quandary [44].

### 8.2.2 *Lung Cancer*

Lung cancer is the third largest cause of cancer-related deaths among men and women in the United States [45]. *Let-7* has been implicated as an oncogene in many human cancers. *Let-7* is a direct downstream target of the *RAS* gene family and a recent report by Chin et al. examines the connection between *Let-7* and *KRAS*, a gene within the *RAS* superfamily [46]. LCS6 is a newly found SNP in the 3′ UTR of *KRAS*, a target of *Let-7*. Expression of *KRAS*, and a consequent lowering of expression of *Let-7*, was found to be significantly associated with non-small-cell lung cancer [46]. Shortly after LCS6 was shown to be tied to lung cancer involving low-dose smokers, a reexamination of the data cast doubt on the association [12]. Though involvement of KRAS as a lung cancer oncogene and lowered *Let-7* expression is not in doubt, the LCS6 SNP is not found to be involved with greater risk of lung cancer [12]. The second study, involving the same populations as the Chin et al. paper uses a slightly different analysis of the data and the authors discuss that the use of a smaller subset of cases coupled with a very low association of LCS6 with lung tumors may play a role in the most current study negating the former study [12].

Other extremely important genes that require deeper investigation are those directly responsible for of miRNA processing. It cannot be ignored that any polymorphisms including SNPs within miRNA biosynthesis genes can have a direct effect on an individual's cancer susceptibility. In a two stage study using a Sequenome mass spectrometry-based genotyping assay for stage 1, 11 miRNA were examined for lung cancer association [47]. The intronic *AGO1* rs636832A>G was found to be a good candidate for further study using a larger population for analysis in a case–control study investigating a Korean population with a little over both 500 cancer patients and control healthy patients. Interestingly, individuals with at least one AA allele at rs636832 have a higher risk of lung cancer while those with an AG or GG alleles have a protective effect against lung cancer [47]. A possible link between cancer risk of smokers and nonsmokers was examined with regards to *AGO1* rs636832. The AA allele was found to be initially correlated to lung cancer prediction in heavy smokers, but in a multivariate logistic regression this correlation was not found [47]. Further study on this *AGO1* SNP in larger case–control studies and in different ethnic groups is needed to elucidate its relationship with lung cancer.

### 8.2.3 *Hepatocellular Carcinoma*

Hepatocellular carcinoma (HCC) is responsible for the majority of liver cancers [48] with the prevalence occurring in China [49]. Worldwide, HCC is the fifth most widespread cancer and is responsible for a third of cancer deaths [50]. *Mir-146a* rs2910164 SNP GG genotype was coupled to hepatocellular carcinoma (HCC) in

males [48]. This is in contrast to another SNP change in the *miR-146a*, a G>U pair to a C>U pair [31]. This case–control study used male and female unrelated Han Chinese participants, 479 HCC patients and 504 controls without occurrence of HCC [48]. Hepatitis B Virus (HBV) was found in a large cohort of the HCC patients (88.9 %), suggesting that HBV plays a role in cancer onset of HCC and may have some interaction with the *miR-146a* variant allele. Production of mature miR-146a was also studies by transiently transfecting 293T cells with either the GG or CC allele. The cells transfected with the GG allele produced more mature miR-146a. Also discovered was the ability of the GG allele to promote colony formation and proliferation in transfected NIH/3T3 cells [48].

Two recent studies investigated origins of HCC and found that *miR-196a2*, coupled with cirrhosis of the liver, has prognostic implications for HCC [51, 52]. Using a Han Chinese population, 310 HCC patients with cirrhosis and 222 individuals with cirrhosis but without HCC, were examined for an rs11614913 *miR-196a2* polymorphism in the first study [51]. Patients with HCC and cirrhosis had a higher level of the rs11614913 CC genotype. Various stages of tumor tissue were collected from 59 HCC patients and the expression levels of *miR-196a* were examined. No significant differences between rs11614913 phenotypes were seen among the different grades and stages of tumors, though a slight association with the T allele was shown with tumor progression [51]. Patients with a CC or CT genotype overall had a higher preponderance of HCC. Because *miR-196a2* has been previously shown to effect expression of mature *miR-196a* it was thought that levels of *miR-196a2* might increase levels of *miR-196a*. Indeed *miR-196a* expression levels were increased in patients with a CC or CT genotype in *miR-196a2* suggesting these miRNA polymorphisms play a role in HCC onset in patients also displaying cirrhosis [51]. In the second study 560 patients examined had HCC and 391 individuals without HCC were used as the control population [52]. As in the previous study males with *miR-196a2* rs11614913 CC genotype had higher levels of HCC diagnosis. This case–control study also examined miR-196a2 expression in different tumor stages and concluded that the miR-196a2 rs11614913 C allele was indicative of patients with certain types of tumors but not in patients with large tumor, advanced-stage tumor or lymphatic metastasis, thus suggesting that differential gene regulation is playing a role in HCC stages [52].

### *8.2.4 Other Cancers*

Other miR-SNPs, through case–control studies, have been found to amplify or diminish risk of other cancers. For instance, cervical cancer is the second most globally reported cancer for which Human Papilloma Virus (HPV) is responsible [53]. Han Chinese women were used in a case–control study recently and a miR-SNP was examined in the *LAMB3* pathway [54]. HPV through genes *E6* and *E7* blocks the expression of *miR-218*. *LAMB3* expresses laminin-5 and this protein is greatly reduced in the absence of *miR-218*. The lack of laminin-5 then further stimulates the

HPV. The study yielded an interesting result; SNP rs11134527 within *pri-miR-218* has a variant linked to increased cervical cancer [54]. It is even postulated that the variant may play a role in increased risk of HPV infection. *LAMB3* was shown to be a direct target of *miR-218* through this work. The study could, by the author's admission, be expanded as the numbers involved in the case study were rather low and the controls were women who has self-reported to have no cancers, but may have been harboring other unknown cancers [54].

A non-synonymous substitution in *GEMIN3* has been coupled with increased bladder cancer risk [55]. This mutation is somewhat similar to the *AGO1* SNP as there is an increased risk of lung cancer, which examines a miRNA biosynthesis gene's relationship with cancer [47]. *GEMIN3* codes for a core protein of a larger complex that plays a role in pre-miRNA splicing; [56] the protein is also in a 15S ribonucleoprotein complex containing eIF2C, another protein that is of great consequence with regard to miRNA processing [57]. The population was a large and homogenous, composed of Caucasian patients diagnosed with bladder cancer and a control group [55]. This analysis once again points to the importance of further study of miRNA processing genes that may alter expression of a myriad of miRNAs potentially involved in tumorigenesis.

Also targeting MiR-SNPs for targeted exploration of linkage in an effort to aid in identification, early-stage head and neck cancer patients with high and low risk secondary primary tumor (STP) and high- and low-risk cancer reoccurrence [58]. The population contained only 150 patients and 300 controls matched by age, gender and ethnicity [58]. Though 18 miR-SNPs were found to be associated with STP and/or reoccurrence, one miR-SNP in particular, rs3747238, is located in a miRNA binding site within *SMC1B* [58]. The 18 miR-SNPs were examined and found to be tied to STP/reoccurrence in a dose dependent fashion [58]. Almost half of the SNPs were located in *RNASEN* (*DROSHA*) [58]. Though mutations in the *RNASEN* are likely to equally affect the processing of all pri-mRNA equally, it is postulated that since miRNAs are expressed differentially in tissue the *RNASEN* would then differentially affect tumorigenesis [58]. *SMC1B* is suggested to play a role in chromosome structure during meiosis and mitosis [59]. Polymorphism in microRNA Target Site (PolymiRTS) is a database of DNA changes in presumed microRNA target sites [60]. PolymiRTS found that a SNP within *SMCB1* is likely to create miRNA binding sites for *miR-609* and *miR-124a* [59]. This SNP is thought to lower expression of *SMC1B* leading to increased carcinogenic potential via further genome volatility [59].

Responsible for a third of deaths among genitourinary malignant cancers, renal cell carcinoma (RCC) causes 40 % mortality among these patients [61]. Though surgery is still the best therapy for RCC, reoccurrence will arise in 20–40 % of patients [62]. Because this cancer has such a high rate of mortality it is imperative that biological markers predictive of clinical outcome, including miR-SNPs be recorded. Seven miR-SNPs were found linked with cancer survival; however, five miR-SNPs were associated with an additional RCC episode [63]. GEMIN4, a protein functionally coupled with GEMIN3, has two SNPs, rs7813T>C and rs91025G>C associated with almost 1.75 % risk of mortality [63, 64]. SNP rs3744741C>T was found concomitant with decreased risk of mortality [63]. *MiR-146a*, *miR-196a-2*,

*miR-423*, *miR-608*, and *miR-601* are also concomitant with RCC recurrence [63]. All of the SNPs found in the miRNAs are located in the pre-miRNA form [63].

There is a scenario in which the heterozygosity of the SNP within a miRNA area can lead to greater likelihood of disease. SNP rs2910164 located in the 3′ strand of *miR-146a* can lead to an increased risk of papillary thyroid carcinoma especially when present as a heterozygote [65]. It is suggested that the heterozygosity somehow leads to a gestalt phenomena wherein the sum of the parts is less than equal to the whole via an epistatic effect between the two alleles [65]. The same group then showed that the SNP produces three miRNAs unlike the normal two produced in homozygous affected individuals. Unlike the phenotype normally seen with heterozygous individuals, two of the mature miRNAs are produced from the 3′ end of *miR-146a* and a third is produced from the leading strand [66]. These three mature miRNAs have the ability to bind various mRNAs thus interrupting the normal *miR-146a* interaction with a predicted variety of mRNAs. It is thought that DNA-damage response pathways acting on cell death signals are invoked within the SNP heterozygote [66]. This group demonstrates the genetic complexity of miRNA interactions with target site, and the importance of somatic mutation with regards to an oncogenic phenotype [66].

As related with BC risk and HCC, *miR-146a* rs2910164 is found once again to be associated with a distinct cancer, this time esophageal squamous cell carcinoma (ESCC) within a case–control study among the Chinese Han [67]. In this case the GG genotype was attached to the ESCC state [67] rather than a change from the G>U pair to a C>U [31].

Like HCC patients, a North Indian population was shown to have a significant risk of developing prostate cancer with a polymorphism in *miR-196a2* (rs11614913) [68]. In this candidate gene study, *miR-499* (rs3746444) also demonstrated significant association with prostate cancer in this case-controlled study, which examined 159 prostate cancer patients and 230 controls [68]. In this case the heterozygote allele CT in *miR-196a* was associated with disease outcome and interestingly a heterozygote CT genotype in *miR-499* also showed linkage to prostate cancer. *MiR-196a* and *miR-499* may work in concert to produce influence the onset of prostate cancer and my serve as prognostic and diagnostic indicators [68].

## 8.3 MiRNAs: Clinical Outcome Predictors?

Discussed above are some of the miRNA binding sites and miRNA SNPs that putatively affect the outcome of BC. In silico scenarios suggest that *miR-453* binds more strongly to an *ESR1* SNP and may thus effectively lower the amount of estrogen produced [37]. Interestingly the majority of BCs do express estrogen receptors, though the cancers that do not are more difficult to treat [69, 70]. The facts that these estrogen receptor positive cancers, though more easily treated with certain drugs, do become easily resistant. It is suggested that hormone replacement therapy would be

a valuable variable to study in addition to familial cancer patients and sudden onset patients [37]. It would be invaluable to know if those patients with the ESR1 SNP allowing more dynamic binding of miR-453 would positively respond to endocrine treatment, thus leading to an individualized plan of treatment [37]. The Kontorovich et al. study addresses their shortcomings [42]. More affected and unaffected individuals need to be included in the study, especially in light of the preliminary *BRCA2* data suggesting that this particular population contains a mutant SNP possibly affecting regulatory actions leading to BC [42]. Again, with more case–control studies focusing on larger and more diverse populations (this initial study only focused on an Ashkenazi population displaying little heterozygosity at the SNP mutation sites), the possibility arises that personal treatment plans could be designed for patients [42]. Yet another study involves integrins which dictate cell adhesion to the extracellular matrix [71]. In a large case–control study 746 Swedish patients with current or former instances of BC were examined along with 1,493 individuals without BC as controls. Probable target sites of miRNA SNPs were examined in integrin genes and a strong association between the ITGB4 rs743554 A allele and aggressive tumor formation was discovered [71]. This allele could be a strong predictive indicator of BC risk.

It has also been shown that several cancer cell lines will alter protein expression based on differential regulation of miR-638 and miR-628-5p and was concluded that small differences in protein expression caused by the interaction of certain regulatory genes SNPs and miRNA will influence the onset of certain cancers [16]. In one case–control study miR-21 was discovered to be involved with lung cancer and proved a chemotherapy response marker [46]. MiRNA related SNPs have also been associated with colon cancer in patients treated with 5-flurouracil and irinotecan [72]. These SNPs were associated with various genes, including rs1834306, within pri-miR-100 and rs7372209 also located in a pri-miRNA, pri-miR-26a [72]. Investigation of SNPs directly involved with deleterious effects of cancer drugs would greatly facilitate basic research studies for cancer. Further studies involving GWAS and large case–control studies could certainly go a long way to advancing these research studies into something that could be used to tailor-make a treatment plan for cancer.

A significant movement advocates personalized medicine [73, 74]. This group feels that much can be done to assist individuals with a particular SNP (or greater than one SNP) that may leave them more likely for cancer onset. Foremost amongst this concern is the large number of adverse drug reactions among cancer patients. Dihydrofolate reductase (DHFR), when overexpressed, leads to methotrexate resistance, a drug primarily used to treat cancer but is also used to treat such conditions as psoriatic arthritis [75]. SNP 892C>T mutation near the 3′ UTR of DHFR hinders *miR-24* from binding its target site located within the 3′ UTR. This causes upregulation of DHFR along with its consequent drug resistance leads to recurrence or even inability to fight cancer. Acknowledging that much more can be done to identify and validate particular miR-SNPs and their association to a diseased state a recent paper points to the new laws enacted by the United States to both encourage parents of

individuals or adult individuals to have their genome sequenced in its entirety [76, 77]. By having a trove of genetic information some sense may be made of miR-polymorphisms and how to effectively diagnose and treat individuals possessing known SNPs showing incomplete penetrance within a population or limited expressivity within an individual.

Excitingly, a new study shows that miR-SNPs have been associated with prostate cancer in men can be used to effectively predict how effective of androgen-deprivation therapy (ADT) [78]. Fifteen total SNPs spread between three prognoses (disease progression, prostate cancer-specific mortality and all-cause mortality) were found [78]. These SNPs were found within miRNAs and miRNA binding sites. Combinatorial analysis between SNPs show that during ADT, patients having a larger number of adverse genotypes have a more rapid time to progression and poorer prostate cancer-specific survival rates [78]. In effect miR-SNPs could act as prognostic markers in patients.

## 8.4 Limitations of miR-SNP Studies

MiR-SNP studies with regards to disease relationship have at least two major caveats, the population size used in the study and the diversity of that population. There are also conflicting reports about miR-SNPs and their ability to affect cancer outcome [79, 80]. A recent report on colorectal cancer (CRC) in Han Chinese using 126 CRC patients and 407 healthy individuals showed that a *miR-196a2* polymorphism (rs11614913 T>C) is not associated with CRC. This is in direct conflict with non-epidemiological studies demonstrating that the polymorphism could be involved with CRC onset [81–83]. Therefore mir-196a2 would not be indicative of a CRC condition. The gene *bcl-2*, its overexpression regulated by *miR-16 +7*, has been indicated as tumor suppressor with regards to chronic lymphocytic leukemia (CLL) [84]. A recent study using 39 CLL patient demonstrated that *miR-16 +7* would be a poor diagnostic marker of CLL [22].

It should be noted however that these case–control studies did not incorporate GWA studies. GWAS is a more useful tool than traditional methods involving linkage studies and candidate gene analysis [85]. Because many cancers have polygenic origins it is important that many loci are examined. GWAS has the statistical power to perform this task in an unbiased way [11, 85]. GWAS has its own limitations however in that the data set produced by many studies is large and complex, leading to confusion about which SNPs may be relevant to disease [86, 87]. GWAS SNP analysis also requires validation in unrelated populaces. To assist in data analysis tools have been constructed that will closely examine the amount of linkage disequilibrium between SNPs. These tools are useful in analyzing miR-SNP data as some studies do not incorporate GWAS, but instead concentrate solely on linkage data and candidate gene analysis [86].

## 8.5 Conclusions

Mir-SNPs have already been shown on a molecular level to be associated with a plethora of cancers. It is imperative not only to have insight into which cancers in general have differentiating levels of miR-SNPs, but also to have an understanding of which miR-SNPs would best predict cancer onset and outcome. MiR-SNPs used within GWA studies with a large case–control group reflective of true population heterogeneity may ultimately prove successful in predicting cancer risks [10, 11, 88].

Interestingly, *miR-146a* has been shown to be involved in a large range of cancers including BC, OC, papillary thyroid carcinoma, renal cell carcinoma and ESCC. Much work has been done on BC. MiRNA mutation as well as mRNA with modified miRNA binding sites is likely to be useful in disease prognosis. A critical point is that heterozygosity of miRNA-SNPs can have an epistatic effect on gene expression. Also crucial, not only do cancers originate from SNPs located on miRNA binding sites of genes directly associated with cancer and SNPs within the miRNAs themselves, cancers are also associated with SNPs within the miRNA machinery genes themselves. Some cancers show a relationship between miRNA binding sites in genes such as *GEMIN3* and *RNASEN* in bladder cancer and early-stage head and neck cancer respectively.

Genetic differentiation among cancers across populations is only starting to be extensively documented. As just one example, renal cell carcinoma would greatly benefit from miR-SNP markers to trace patients likely to need advanced forms of therapy to avoid remission after initial treatment [63]. One cautionary note to building large databases with miR-SNP data from many populations would be maintenance of privacy to the donors. MiR-SNPs will prove a notable trove of data and most likely be very effective as a clinical outcome predictor.

## References

1. http://www.mirbase.org/. 2011 [updated 2010 September; cited 2011 Jan 27]; Available from: http://www.mirbase.org/.
2. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005;433(7027):769–73.
3. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010;24(10):992–1009.
4. Moretti F, Thermann R, Hentze MW. Mechanism of translational regulation by miR-2 from sites in the 5′ untranslated region or the open reading frame. RNA. 2010;16(12):2493–502.
5. Lee EK, Gorospe M. Coding region: the neglected post-transcriptional code. RNA Biol. 2011; 8(1):44–8.
6. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA—target recognition. PLoS Biol. 2005;3(3):e85.
7. Bartel DP. MicroRNAs: target recognition and regulatory function. Cell. 2009;136:215–33.
8. Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet. 2001;27(3):234–6.

9. Milne RL, Antoniou AC. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. Ann Oncol. 2011;22 Suppl 1:i11–7.
10. Easton DF, Eeles RA. Genome-wide association studies in cancer. Hum Mol Genet. 2008;17(R2):R109–15.
11. Varghese JS, Easton DF. Genome-wide association studies in common cancers—what have we learnt? Curr Opin Genet Dev. 2010;20(3):201–9.
12. Nelson HH, Christensen BC, Plaza SL, Wiencke JK, Marsit CJ, Kelsey KT. KRAS mutation, KRAS-LCS6 polymorphism, and non-small cell lung cancer. Lung Cancer. 2010;69(1):51–3.
13. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res. 2004;64(11):3753–6.
14. Wynendaele J, Böhnke A, Leucci E, Nielsen SJ, Lambertz I, Hammer S, et al. An illegitimate microRNA target site within the 3′ UTR of MDM4 affects ovarian cancer progression and chemosensitivity. Cancer Res. 2010;70(23):9641–9.
15. Tsang WP, Kwok TT. The miR-18a* microRNA functions as a potential tumor suppressor by targeting on K-Ras. Carcinogenesis. 2009;30(6):953–9.
16. Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA. MicroRNAs [mdash] the micro steering wheel of tumour metastases. Nat Rev Cancer. 2009;9(4):293–302.
17. Spizzo R, Nicoloso MS, Croce CM, Calin GA. SnapShot: MicroRNAs in cancer. Cell. 2009;137(3):586–e1.
18. Xiang J, Wu J. Feud or friend? The role of the miR-17-92 cluster in tumorigenesis. Curr Genomics. 2010;11:129–35.
19. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101(9):2999–3004.
20. Wu M, Jolicoeur N, Li Z, Zhang L, Fortin Y, L'Abbe D, et al. Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. Carcinogenesis. 2008; 29(9):1710–6.
21. Landi D, Gemignani F, Barale R, Landi S. A catalog of polymorphisms falling in microRNA-binding regions of cancer genes. DNA Cell Biol. 2008;27(1):35–43.
22. Yazici H, Zipprich J, Peng T, Akisik EZ, Tigli H, Isin M, et al. Investigation of the miR16-1 (C>T)+7 substitution in seven different types of cancer from three ethnic groups. J Oncol. 2009;2009:827532.
23. Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W. Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. Carcinogenesis. 2008;29(7):1306–11.
24. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med. 2005;353(17):1793–801.
25. Davis-Dusenbery BN, Hata A. Mechanisms of control of microRNA biogenesis. J Biochem. 2010;148(4):381–92.
26. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043):828–33.
27. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839–43.
28. Garzon R, Pichiorri F, Palumbo T, Iuliano R, Cimmino A, Aqeilan R, et al. MicroRNA fingerprints during human megakaryocytopoiesis. Proc Natl Acad Sci U S A. 2006;103(13): 5078–83.
29. Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006;103(7):2257–61.
30. Meiri E, Levy A, Benjamin H, Ben-David M, Cohen L, Dov A, et al. Discovery of microRNAs and other small RNAs in solid tumors. Nucleic Acids Res. 2010;38(18):6234–46.
31. Shen J, Ambrosone CB, DiCioccio RA, Odunsi K, Lele SB, Zhao H. A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. Carcinogenesis. 2008;29(10):1963–6.

32. Hu Z, Liang J, Wang Z, Tian T, Zhou X, Chen J, et al. Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. Hum Mutat. 2009;30(1):79–84.
33. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer. 2010;10(6):389–402.
34. Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A. 2007;104(9):3300–5.
35. Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Hum Mol Genet. 2007;16(9):1124–31.
36. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet. 2006;38(12):1452–6.
37. Tchatchou S, Jung A, Hemminki K, Sutter C, Wappenschmidt B, Bugert P, et al. A variant affecting a putative miRNA target site in estrogen receptor (ESR) 1 is associated with breast cancer risk in premenopausal women. Carcinogenesis. 2009;30(1):59–64.
38. Wacholder S, Struewing JP, Hartge P, Greene MH, Tucker MA. Breast cancer risks for BRCA1/2 carriers. Science. 2004;306(5705):2187–91. author reply 2187–91.
39. Robles-Diaz L, Goldfrank DJ, Kauff ND, Robson M, Offit K. Hereditary ovarian cancer in Ashkenazi Jews. Fam Cancer. 2004;3(3–4):259–64.
40. Narod SA. Modifiers of risk of hereditary breast cancer. Oncogene. 2006;25(43):5832–6.
41. Levy-Lahad E, Friedman E. Cancer risks among BRCA1 and BRCA2 mutation carriers. Br J Cancer. 2007;96(1):11–5.
42. Kontorovich T, Levy A, Korostishevsky M, Nir U, Friedman E. Single nucleotide polymorphisms in miRNA binding sites and miRNA genes as breast/ovarian cancer risk modifiers in Jewish high-risk women. Int J Cancer. 2010;127(3):589–97.
43. Sætrom P, Biesinger J, Li SM, Smith D, Thomas LF, Majzoub K, et al. A risk variant in an miR-125b binding site in BMPR1B is associated with breast cancer pathogenesis. Cancer Res. 2009;69(18):7459–65.
44. Yang R, Schlehe B, Hemminki K, Sutter C, Bugert P, Wappenschmidt B, et al. A genetic variant in the pre-miR-27a oncogene is associated with a reduced familial breast cancer risk. Breast Cancer Res Treat. 2010;121(3):693–702.
45. Prevention CfDCa. United States Cancer Statistics. Department of Health and Human Services; 2007 [updated 2007; cited 2 Nov 2011]; Available from: http://apps.nccd.cdc.gov/uscs/toptencancers.aspx.
46. Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3′ untranslated region increases non-small cell lung cancer risk. Cancer Res. 2008;68(20):8535–40.
47. Kim JS, Choi YY, Jin G, Kang HG, Choi JE, Jeon HS, et al. Association of a common AGO1 variant with lung cancer risk: a two-stage case-control study. Mol Carcinog. 2010;49(10):913–21.
48. Xu T, Zhu Y, Wei QK, Yuan Y, Zhou F, Ge YY, et al. A functional polymorphism in the miR-146a gene is associated with the risk for hepatocellular carcinoma. Carcinogenesis. 2008;29(11):2126–31.
49. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer. 2006;118(12):3030–44.
50. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007;132(7):2557–76.
51. Li X-D, Li Z-G, Song X-X, Liu C-F. A variant in microRNA-196a2 is associated with susceptibility to hepatocellular carcinoma in Chinese patients with cirrhosis. Pathology. 2010;42(7):669–73.
52. Qi P, Dou TH, Geng L, Zhou FG, Gu X, Wang H, et al. Association of a variant in MIR 196A2 with susceptibility to hepatocellular carcinoma in male Chinese patients with chronic hepatitis B virus infection. Hum Immunol. 2010;71(6):621–6.
53. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005;55(2):74–108.

54. Zhou X, Chen X, Hu L, Han S, Qiang F, Wu Y, et al. Polymorphisms involved in the miR-218-LAMB3 pathway and susceptibility of cervical cancer, a case-control study in Chinese women. Gynecol Oncol. 2010;117(2):287–90.
55. Yang H, Dinney CP, Ye Y, Zhu Y, Grossman HB, Wu X. Evaluation of genetic variants in MicroRNA-related genes and risk of bladder cancer. Cancer Res. 2008;68(7):2530–7.
56. Shpargel KB, Matera AG. Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. Proc Natl Acad Sci U S A. 2005;102(48):17372–7.
57. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, et al. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. 2002;16(6):720–8.
58. Zhang X, Yang H, Lee JJ, Kim E, Lippman SM, Khuri FR, et al. MicroRNA-related genetic variations as predictors for risk of second primary tumor and/or recurrence in patients with early-stage head and neck cancer. Carcinogenesis. 2010;31(12):2118–23.
59. Revenkova E, Eijpe M, Heyting C, Hodges CA, Hunt PA, Liebe B, et al. Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. Nat Cell Biol. 2004;6(6):555–62.
60. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, et al. PolymiRTS database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res. 2007;35 Suppl 1:D51–4.
61. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin. 2009;59(4):225–49.
62. Janzen NK, Kim HL, Figlin RA, Belldegrun AS. Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease. Urol Clin North Am. 2003;30(4):843–52.
63. Lin J, Horikawa Y, Tamboli P, Clague J, Wood CG, Wu X. Genetic variations in microRNA-related genes are associated with survival and recurrence in patients with renal cell carcinoma. Carcinogenesis. 2010;31(10):1805–12.
64. Horikawa Y, Wood CG, Yang H, Zhao H, Ye Y, Gu J, et al. Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. Clin Cancer Res. 2008;14(23):7956–62.
65. Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. Proc Natl Acad Sci U S A. 2008;105(20):7269–74.
66. Jazdzewski K, Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, Jarzab B, et al. Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. Proc Natl Acad Sci U S A. 2009;106(5):1502–5.
67. Guo H, Wang K, Xiong G, Hu H, Wang D, Xu X, et al. A functional variant in microRNA-146a is associated with risk of esophageal squamous cell carcinoma in Chinese Han. Fam Cancer. 2010;9(4):599–603.
68. George GP, Gangwar R, Mandal RK, Sankhwar SN, Mittal RD. Genetic variation in microRNA genes and prostate cancer risk in North Indian population. Mol Biol Rep. 2010;38(3): 1609–15.
69. Fishman J, Osborne MP, Telang NT. The role of estrogen in mammary carcinogenesis. Ann N Y Acad Sci. 1995;768:91–100.
70. Martin G, Davio C, Rivera E, Melito G, Cricco G, Andrade N, et al. Hormone dependence of mammary tumors induced in rats by intraperitoneal NMU injection. Cancer Invest. 1997;15(1): 8–17.
71. Brendle A, Lei H, Brandt A, Johansson R, Enquist K, Henriksson R, et al. Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker. Carcinogenesis. 2008;29(7):1394–9.
72. Boni V, Zarate R, Villa JC, Bandres E, Gomez MA, Maiello E, et al. Role of primary miRNA polymorphic variants in metastatic colon cancer patients treated with 5-fluorouracil and irinotecan. Pharmacogenomics J. 2010;11(6):429–36.
73. Mishra PJ. MicroRNA polymorphisms: a giant leap towards personalized medicine. Per Med. 2009;6(2):119–25.

74. Rossbach M. Small non-coding RNAs as novel therapeutics. Curr Mol Med. 2010;10(4): 361–8.
75. Mishra PJ, Humeniuk R, Longo-Sorbello GS, Banerjee D, Bertino JR. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. Proc Natl Acad Sci U S A. 2007;104(33):13513–8.
76. Genetic Information Nondiscrimination Act of 2008. 2010; Available from: http://thomas.loc.gov/cgi-bin/bdquery/z?d110:h.r.00493.
77. National Human Genome Research Institute. 2010; Available from: http://www.genome.gov/10002077.
78. Bao B-Y, Pao J-B, Huang C-N, Pu Y-S, Chang T-Y, Lan Y-H, et al. Polymorphisms inside MicroRNAs and MicroRNA target sites predict clinical outcomes in prostate cancer patients receiving androgen-deprivation therapy. Clin Cancer Res. 2011;17(4):928–36.
79. Chen H, Sun L-Y, Chen L-L, Zheng H-Q, Zhang Q-F. A variant in microRNA-196a2 is not associated with susceptibility to and progression of colorectal cancer in Chinese. Intern Med J. 2012;42(6):e115–9.
80. Yazici H, Zipprich J, Peng T. Investigation of the miR16-1 (C>T)+7 substitution in seven different types of cancer from three ethnic groups. J Oncol. 2009;2009:827532
81. Motoyama K, Inoue H, Takatsuno Y, Tanaka F, Mimori K, Uetake H, et al. Over- and under-expressed microRNAs in human colorectal cancer. Int J Oncol. 2009;34(4):1069–75.
82. Schepeler T, Reinert JT, Ostenfeld MS, Christensen LL, Silahtaroglu AN, Dyrskjot L, et al. Diagnostic and prognostic microRNAs in stage II colon cancer. Cancer Res. 2008;68(15): 6416–24.
83. Bandres E, Agirre X, Bitarte N, Ramirez N, Zarate R, Roman-Gomez J, et al. Epigenetic regulation of microRNA expression in colorectal cancer. Int J Cancer. 2009;125(11):2737–43.
84. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005;102(39):13944–9.
85. Sethupathy P, Collins FS. MicroRNA target site polymorphisms and human disease. Trends Genet. 2008;24(10):489–97.
86. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. 2009;37 Suppl 2:W600–5.
87. Yoshida T, Ono H, Kuchiba A, Saeki N, Sakamoto H. Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: gastric cancer as an example. Cancer Sci. 2010; 101(7):1582–9.
88. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661–78.

# Chapter 9
# Bioinformatics Approaches to the Study of MicroRNAs

**Ravi Gupta and Ramana V. Davuluri**

**Abstract** The introduction of Next-Generation Sequencing (NGS) technologies has opened new avenues, including determination of the relationship of genomic and epigenomic variation and phenotypes to disease. These technologies are particularly well suited for discovery of both small RNAs, around 30–75 base pairs long, and long RNAs, longer than 75 base pairs, because the NGS sequencers can produce millions of short reads in a relatively rapid period of time. However, the large amounts (terabases) of data generated require proper computational resources and analytical methods to translate the rich source of genomic data into meaningful information for biomedical applications. In this chapter, we describe various bioinformatics methods for performing integrative analysis for the identification of miRNAs and their target mRNAs from the NGS sequencers. We describe the most commonly used databases and prediction programs that are available on the World Wide Web and demonstrate the use of some of these programs by an example. We provide a list of these programs along with their Web URLs and suggest guidelines for successful application.

**Keywords** Next-generation sequencing • miRNA • ncRNAs • Target sites • Computational prediction • Databases

R. Gupta
Center for Systems and Computational Biology, The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA

Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, PA, USA

SciGenom Labs. Pvt. Ltd., Cochin, Kerala, India

R.V. Davuluri (✉)
Center for Systems and Computational Biology, The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA

Molecular and Cellular Oncogenesis Program, The Wistar Institute, Philadelphia, PA, USA
e-mail: rdavuluri@wistar.org

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_9,

## 9.1 Introduction

RNA interference (RNAi) is an evolutionary conserved gene silencing pathway by which double-stranded RNA (dsRNA) induces highly specific gene silencing [1]. RNAi is mechanistically related to a number of other conserved RNA silencing pathways, which are involved in the cellular control of gene expression and in protection of the genome against mobile repetitive DNA sequences, retroelements and transposons. These RNA silencing pathways are all associated with small non-coding RNAs (ncRNAs) that guides protein-complex for mostly repressing homologous sequence [2, 3]. Several classes of small ncRNAs have been discovered, such as small interfering RNAs (siRNA), micro RNAs (miRNAs), and Piwi-interacting RNAs (piRNAs). Small ncRNAs are of size 20–31 nucleotides and are divided into different classes based on their biogenesis mechanism and their association with different types of Argonaute (Ago) proteins [4–7]. The discovery of small ncRNAs has revolutionized the field of gene regulation and particularly the miRNAs have emerged as key regulators at both transcriptional and posttranscriptional levels. Although in many cases the target of small RNAs and their mechanism is still not clear, it is certain that many of these small RNAs play important role in several biological processes including development timing, cell differentiation, cell proliferation, cell death, metabolic control, transposon silencing, and antiviral defense [8, 9].

The best understood among the three small RNA classes is miRNA. MiRNAs are single-stranded which are generated from local hairpin structures by the action of two Rnase III-type proteins, Drosha and Dicer [10–14]. Mature miRNAs are 20–24-nucleotide (nt) long and highly conserved during evolution [15, 16]. Mature miRNAs and Argonaute (Ago) proteins form the RNA-induced silencing complex (RISC), a ribonucleoprotein complex mediating posttranscriptional gene silencing. The miRNA was first discovered in *Caenorhabditis elegans* by Victor Ambros and his colleagues in 1993 [17]. Till date several miRNAs have been discovered in both animals and plants. They play important roles in several biological processes including development, differentiation and proliferation of cells, angiogenesis, apoptosis, stress adaptation, and hormone signaling [14, 16, 18–20]. Many studies have shown that miRNAs are also involved in wide range of diseases, including multiple cancers [21–27]. Utilizing the miRNAs as cancer biomarkers is recognized as a valuable diagnostic strategy, because aberrant expression levels of miRNAs are found in tissues and in sera from patients with different types/stages of cancers [27–29].

Discovery of miRNA sequences and their genomic coordinates using in silico-based approaches is a very challenging and complicated task. Various computational and experimental approaches have been developed to predict miRNA in several organisms. The introduction of recent massive parallel sequencing or next-generation sequencing (NGS) technologies, such as Applied Biosystems' SOLiD3, Genome Analyzer or HiSeq from Illumina, has revolutionized the field of small

ncRNAs [30]. These technologies produces several millions of DNA short reads (36–250 bp, depending on platform) in a single run and are perfect for genome-wide study of small ncRNAs. Several programs and pipelines have been developed to analyze NGS datasets extensively [31–36]. In addition to identification of miRNA sequences, prediction of miRNA targets has also been an active area of research [37–44]. Bioinformatics analysis has indicted that approximately one-third of all human protein-coding genes are targeted by miRNAs [45]. The small RNA-Seq technology is also allowing researchers to identify RNA editing sites in miRNAs [46–48]. The major focus of this chapter is to introduce different bioinformatics techniques and programs that identify miRNAs, miRNA–mRNA interaction sites, and miRNA editing position.

## 9.2 MiRNA Prediction

The first miRNA, lin-4, was discovered by forward genetic experiments in worm in 1993 [17]. After 7 years another miRNA, let-7, was discovered in worms using forward genetic experiments [49]. This approach for miRNA discovery is very time consuming and expensive. Later on cloning based methods was introduced [50] that led to discovery of several miRNAs in different organisms [51–56]. However, this approach has limitations as it cannot identify low expressed or condition specific or tissue specific miRNAs. In the last one decade, several computational techniques have been developed to predict miRNA. But the computational approaches when used alone can produce too many false positive preditions. The recent advent of next-generation sequencing (NGS) technologies has transformed the field of miRNA discovery. Using these massive parallel techniques, one can not only identify miRNAs but can also measure miRNA expression at unprecedented sensitivity [57–60]. The NGS approach is a hybrid technique because it requires both experimental and computational expertise to study miRNAs. Millions of sequence reads generated by NGS sequencer need a systematic and robust approach for analysis that can help researchers to study the existing miRNAs and also discover novel miRNAs. Here, we review the computational and NGS analysis programs. In addition, we describe general steps followed to analyze NGS small RNA data (small RNA-Seq).

### *9.2.1 Computational Appoarches*

#### 9.2.1.1 MiRscan

MiRscan first scans the conserved region of the genome to identify hairpin structures [61]. The hairpin structure is predicted using RNAfold program [62].

The predicted structures with at least 25 base-paired stem-loop structures and folding energy ≤−25 kcal/mol are retained for further processing. MiRscan identified 36,000 hairpins in *C. elegans* that satisfied hairpin and conservation criterias. The hairpins were then compared with a training set that was generated from known miRNAs. MiRscan used several features including strict conservation in the 5′ half of pre-miRNA, sequence bases in the first five bases, base pairing in fold-back and non fold-back regions, preference of symmetric internal loops and buldges in the pre-miRNA region and others to predict pre-miRNAs. MiRscan predicted 50 of the 53 miRNAs previously known in *C. elegans* [61].

#### 9.2.1.2 miRseeker

Lai et al. [63] proposed miRseeker program to predict pre-miRNA from conserved intronic and intergenic regions of the genome. Each conserved region was divided into 100 base sequences that were folded using mfold algorithm in both forward and reverse complement direction. Each candidate pre-miRNA structure with length of longest helical arm (≥23 bases) and free energy of isolated arm ≤−23.0 kcal/mol was further evaluated. The continuous helical pairing was positively scored whereas the internal loops, asymmetric loops and bulged were negatively scored. miRseeker identified 18 out of 24 previously known miRNAs in *Drosophila* genome. In total, 48 novel miRNAs were identified and expression of 20 miRNAs was validated using northern analysis.

#### 9.2.1.3 RNAmicro

RNAmicro [64] first computes consensus structure from consensus sequence of multiple-sequence alignment using RNAfold algorithm [62]. In the second step it calculates descriptors which includes stem and hairpin loop regions length, G+C content, average folding energy, consensus structure energy, sequence conservation. In the last step a support vector machine classification model is constructed using libsvm package.

#### 9.2.1.4 miRRim

Goroi et al. [65] proposed hidden markov model (HMM) based approach to predict pre-miRNA sequence. A five dimension vector is generated using evolutionary and secondary structural features of miRNA and surrounding regions. The first dimension is the conservation score calculated using phylo-HMM [66], a multiple sequence alignment tool. The second dimension represents statistical significance of the potential minimum free energy (MFE) with respect to both sequence length and base composition. The remaining vectors are based on base-pairing probability.

#### 9.2.1.5 Phylogenetic Shadowing

Berezikov et al. [67] studied 122 miRNAs in ten primate species. They found that while the stem region of the folder miRNA is extremly conserved, the loop region has high variations. The classical phylogenetic footprinting approach has several disadvatages expecially when looking at short sequences and increasing phylogenetic distance. In [67], authors proposed phylogenetic shadowing approach to accurately identify conserved regions at single nucleotide resolution level [68]. Phylogenetic shadowing approach is based on the alignment of phylogenetically closely related species.

### *9.2.2 Next-Generation Sequencing Approach*

The initial step in NGS is to isolate RNAs of specific size from the sample that can be used to create library by attaching specific adapters on the ends of the molecules. With the availability of improved kits and protocols the sequencing and library creation has become a routine task. But the final analysis of millions of short sequenced RNAs which are obtained after sequencing has become the major bottleneck. The large amount of data generated by next-generation sequencer requires new set of tools for translating the huge data into useful and meaningful information. The main steps in small RNA-Seq analysis are as follows:

1. Preprocessing
2. Alignment
3. Annotation
4. Expression estimation
5. Novel miRNA prediction

1. *Preprocessing*—This step is an essential step to improve the quality of the data for final analysis. First the low quality of reads is removed from the sample. Typically reads with error rate >0.01 are removed from further analysis. Then the adapter sequences, which are added prior to actual sequencing step, should be properly removed from the reads before aligning them to the reference genome. If the adapter sequences are not removed correctly then it can either result in drop of alignment percentage and/or could increase incorrect alignment. The adapter sequences are provided by the user and depend on the protocol used for sequencing. The average error rate and frequency of individual nucleotides for each cycle need to be checked. The distribution of average error rate at each position shows if the quality of each nucleotide degrades towards end of the sequencing cycle. If so, before removing the adapter sequences from raw reads, the user (choosing a cutoff by user) should neglect the last few nucleotides in order to minimize the error due to last low quality nucleotides. Some of the programs remove trimmed reads matching

rRNAs, tRNA, and other noncoding RNAs from further analysis. These RNAs are assumed to be present in the sample due of degradation.

2. *Alignment*—Alignment of reads to the reference miRNA and genome is a very important step for accurate analysis of the data. Due to short reads and presence of error in sequencing step alignment of read to correct place is a challenging task. Recently several programs have been developed to align short reads from NGS sequencers along with quality of nucleotides assigned by the sequencers [69]. For example, Bowtie (http://bowtie.cbcb.umd.edu), an open-source program ultrafast, memory-efficient alignment program can be used for aligning short DNA sequence reads to large genomes [70]. Bowtie extends previous Burrows–Wheeler techniques with a novel quality-aware backtracking algorithm that permits mismatches. Bowtie provides an option to use multiple processor cores simultaneously to achieve amazingly greater alignment speeds. This program also provides several useful options such as maximum read length, number of mismatches for alignment, number of multiple-mapping, and many others for alignment.

   For the alignment of reads an index file of the reference miRNA and genome is necessary for majority of the short read aligners. The reference miRNA and genome sequences are downloaded miRBase and NCBI respectively. miRBase maintains a set of high quality pre-miRNA and mature miRNA sequences which includes some the latest discovered miRNAs. The trimmed and filtered reads are first aligned to known reference pre-miRNA sequences. The unaligned reads are then mapped to the reference genome. For alignment total multiple mapping and maximum mistmatches are need to be allowed need to be decided by the user.

   Quality assessment is a necessary step after alignment of reads. This step is very important to keep a further check on the quality of sequence reads that would be used in the downstream analysis. Two different correlation tests need to be performed: (a) alignment mismatches and position of mismatch in the sequence read (b) alignment mismatch and quality of nucleotide provided by the sequencers. Performing this quality assessment is necessary because there have been reports that quality of bases decreases towards end and there exits some biasing while making base calls in sequencing step [71].

   After removing adapter sequences and quality assessment we can evaluate the lengths distribution of mapped sequence reads. If the length distribution pattern of the mapped sequences follows tri-modal distribution then it implies that there exist three classes of small RNAs, 18–20 nucleotides (siRNAs), 20–24 nucleotides (miRNA), and $\geq$25 nucleotides (piRNAs).

3. *Annotation*—Characterization of aligned reads into different categories is an important step for analysis of small RNA sequences. Annotation categories are assigned to the reads depending on the annotation of corresponding genomic co-ordinates present in different databases (Table 9.1). The reads are annotated with known genes, noncoding RNAs, and repeats.

**Table 9.1** List of annotation types and their sources

| Annotation type | Resource |
|---|---|
| miRNAs | miRBase—ftp://mirbase.org/pub/mirbase/CURRENT/ |
| Other noncoding RNAs (rRNA, tRNA, snoRNA, srpRNA, scRNA) | NCBI—www.ncbi.nlm.nih.gov<br>GtRNAdb—http://gtrnadb.ucsc.edu/<br>snoRNABase—http://www-snorna.biotoul.fr/<br>UCSC—http://genome.ucsc.edu/<br>Ensembl—www.ensembl.org/<br>Rfam—http://rfam.sanger.ac.uk/ |
| Repeats (LINE, SINE, LTR, DNA, simple repeat, satellite, and others) | Repeat masker track from UCSC |
| Known genes | UCSC—http://genome.ucsc.edu/<br>Ensembl—www.ensembl.org/<br>Mouse Genome Informatics (http://www.informatics.jax.org/) |
| Pseudo genes | Vega genome browser (http://vega.sanger.ac.uk)<br>Pseudofam (http://pseudofam.pseudogene.org) |
| piRNAs | piRNA Bank (http://pirnabank.ibab.ac.in/) and literature |
| siRNAs | Published literature |

The annotation is performed in the following order: miRNAs, noncoding RNAs, repeats, known genes. If the reads do not belong to any of these three categories then it can be assigned to "no annotation" category. The reads which are mapped to multiple locations of the reference genome are put into one of the above discussed categories (i.e., noncoding RNAs, repeat, known gene, and no annotation) if all the genomic locations have same annotation and all mapping position are taken as possible read location. However, if at least one of the locations is annotated as different from others then the reads are kept in separate category named as "multi-annotation category." The reason for doing this is to avoid any error that could arise if it was randomly assigned to one of the mapped locations. Further, separate analysis can be performed on the multi-annotation category class. The annotations of different categories of ncRNAs can be downloaded from public databases (Table 9.1).

4. *Expression estimation*—One of the objectives of small RNA-Seq experiment is to measure the expression of miRNAs in the sample and perform differential expression analysis. The miRNA expression is quantfied in terms of RPM (Reads Per Million) value, which is normalized against total reads in the sample. RPM value is defined as follow

$$RPM = \frac{\#\,of\ reads\ mapping\ precursor\ or\ mature\ miRNA}{Total\ reads\ in\ the\ sample} \times 10^6$$

It is a direct measure for the amount of mature miRNA in the sample. However, the miRNA expression for next-generation sequencing data is highly dependent on RNA extraction and library preparation [72]. Issues such as reads with errors, multiple mapped reads, low read count and read length should be taken care for expression calculation. The substitution errors was a big concern initially but with advancement in quality of reads by next-generation sequencer, the number of error occurred during sequencing is very small now. Multiple mapping of reads is a big concern because a large fraction of total reads are mapped to more than one precursor miRNA. This happens because of high similarity between miRNA species. So filtering out multiple mapped reads will affect the gene expression anlysis of miRNAs. Another challenge faced by the researcher is filtering out miRNAs with low read count. Usually, the threshold is selected arbitrarily for filtering. However, some of the latest studies have used statistical approach to remove low expressed miRNAs. For identifying differentially expressed miRNAs various programs have proposed statistical approaches including fold-change, chi-square test, *t*-test, ANOVA, negative bionomial. The available tools that can perform differential expression analysis are mirExpress, SeqBuster, mirTools, miRNAKey.

5. *Novel miRNA prediction*—One of the main adavantages of performing small RNA-Seq experiments is the discovery of novel miRNAs. The next-generation sequencing approach has led to discovery to many new miRNAs in various species. In the last few years many programs have been developed that can predict novel miRNAs from small RNA-Seq data. These programs are based on the current knowledge on biogenesis of miRNAs in the cell, secondary structure and folding energy and sequence conservation. Some of the programs that can discover novel miRNAs are mirDeep, mirTools, mirAnalyzer, UEA sRNA Tookit.

Some of the tools developed in past few years for analysis of small RNA-Seq datasets are briefly described below.

#### 9.2.2.1 miRDeep

The miRDeep program can discover both known and novel miRNAs present in small RNA-Seq sample [31, 73]. miRDeep uses Bayesian framework of miRNA biogenesis to score the aligned reads for miRNA prediction. It can identify both canonical and noncanonical miRNAs especially generated from transposable elements and detected in many samples. miRDeep includes several scripts that can preprocess the input data, perform alignment and a core algorithm to predict miRNAs.

The miRDeep analysis pipeline is divided into three different modules: mapper module, qualifier module, and miRDeep2 module. The mapping module process the fastq input file and align it to the reference genome. The qualifier module counts total reads falling in the known miRNA precursor. The miRDeep2 module is the core of the pipeline, which predicts known and novel miRNAs in the input sample.

First, it formats the input files to the module. Second, it identifies the potential miRNA precursor region of the genome based on the read mapping and stacking. Third, using Bowtie program the reads are aligned to the candidate miRNA precursor sequences. Fourth, the RNAfold program is applied to candidate miRNA precursor regions to check if it forms a miRNA hairpin like structure. The miRDeep2 core algorithm scores the precursor if it forms hairpin structure and if the reads falling on the hairpin structure satisfy the Dicer processing. The output of miRDeep typically consists of a list of mature miRNAs, known and novel miRNA precursors along with score and the estimates for the number of false positives.

#### 9.2.2.2 SeqBuster

SeqBuster is a Web-based application for analysis of small RNA-Seq datasets [32]. It first trims the adapter sequences from high-throughput data. Second, it aligns the trimmed data to the reference precursor and mature miRNA reference sequences. Third, the sequences are aligned to the reference mRNA and genome sequences. A unique feature of SeqBuster is the characterization of miRNA variants, known as IsomiRs. SeqBuster identified several type of IsomiRs in human stem cells datasets [32]. Recently, several studies have shown functional role for IsomiRs [74–78]. Differential expression analysis module of SeqBuster use several statistical tests to identifiy differentially expressed genes. The functional analysis module of SeqBuster provides a list of statistically significant pathways targeted by the expressed miRNAs and IsomiRs. The pathway enrichement analysis is performed using ingenuity pathway analysis.

#### 9.2.2.3 mirTools

mirTools performs extensive analysis of small RNA-Seq dataset. It performs low-quality read filtering, adapter trimming, alignment to reference genome, annotation, miRNA expression estimation, predicts novel miRNAs and identifies differentially expressed miRNAs [34]. The reads are aligned to the reference genome using SOAP program [79]. The reads are also aligned to miRBase [80], Rfam [81], repeat database generated by RepeatMasker [82] and protein-coding genes of the reference organism. A hierarchy based approach is then followed to assign read to a specific class. The differential expressed miRNAs are identified using Bayesian approach. The reads that do not fall in the known RNA catergory are used for identification of novel miRNAs. mirTools uses the miRDeep program to identify novel miRNAs.

#### 9.2.2.4 miRanalyzer

miRanalyzer is a Web-based tool for the identification of known and novel miRNAs from small RNA-Seq samples. The reads are first aligned to the reference precursor

miRNA sequences to identify known expressed miRNAs. The remaining reads are aligned to mRNAs, noncoding RNA (Rfam), and retrotransposons [81]. miRanalyzer implements random forest based machine learning approach to predict novel miRNAs. Prediction model for novel miRNA identification is built using human, rat and *C. elegans* known miRNAs. Annotation-Modules program [83] is used to identify significant pathway targeted by the miRNA present in the sample.

#### 9.2.2.5 miRNAKey

miRNAKey is a program that can be installed locally for analysis of small RNA-seq data [36]. The program performs the following analysis: adapter removal, alignment to known precursor miRNA, miRNA expression quantification, and differential expression. It implements SEQ-RM algorithm [84] for optimizing the multiple-mapped reads and is a unique feature of miRNAKey. Differential miRNA expression is calculated using chi-squared statistics. The *p*-values obtained are corrected using the Bonferroni correction for multiple hypothesis testing.

#### 9.2.2.6 UEA sRNA Tookit

This toolkit consists a collection of programs for analyzing plants and animals small RNA-Seq data [85]. The programs included in the toolkit are preprocessing, filtering, miRProf, miRCAT, FiRePat, SiLoCo, SiLoMa, RNAfold annotation. The preprocessing module removes the adapter sequences from the sequences. A filtering process is implemented to select trimmed sequences within user defined length. For example, the mature miRNA sequences are of length $\geq$18 and $\leq$25 bases. The miRCAT program is used to identify known miRNAs in the sample. It first aligns the read to the reference genome. Then it looks at alignment distribution patterns and secondary structure of genomic regions corresponding to small RNA hits to predict miRNAs. The miRProf program generates expression profiles of the miRNAs in the sample. The FiRePat looks for patterns in the small RNA-Seq and gene expression level from microarray. SiLoMa program maps the reads in the sample to the reference genome and generates a graphical map of alignment.

## 9.3 MiRNA Target Prediction

Several computational approaches have been developed for miRNA target prediction to come up with a manageable number of predicted target sites for experimental follow-up (see review [86] for experimental strategies for microRNA target identification). Different computational approaches to model the miRNA–mRNA target sites have led to several miRNA target prediction methods (Table 9.2). Although

**Table 9.2** List of miRNA resource database

| Database name | Brief description | Weblink |
|---|---|---|
| miRBase [80] | Contains 21,264 entries representing precursor miRNAs, 25,141 mature miRNAs of 193 species. Reads from small RNA-Seq can be viewed | http://www.mirbase.org/ |
| miRDB [120] | Provides predicted miRNA targets and functional annotation for human, mouse, rat, dog, and chicken | http://mirdb.org/miRDB/ |
| microRNA.org [91] | Provides miRNA targets and expression values for human, mouse, rat, fruitfly, and nematode | http://www.microrna.org/microrna/home.do |
| microPIR [122] | Provides miRNA targets present within human promoter regions. It also integrates various other information including EST expression, sequence conservation across vertebrate species, transposable element and repeat position, transcription factor binding sites, CpG islands, and SNPs | http://www4a.biotec.or.th/micropir/ |
| miRTarBase [124] | Contains experimentally validated miRNA–target interactions (MTIs) sites for 14 different species | http://mirtarbase.mbc.nctu.edu.tw/ |
| TarBase [125] | Contains manually curated 65,814 experimentally validated miRNA–gene interactions sites | http://www.microrna.gr/tarbase/ |
| miRWalk [135] | Contains predicted and validated information on miRNA–target interaction | http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/ |
| starBase [136] | Database of miRNA–target regulatory relationships, identified from Argonaute CLIP-Seq and Degradome-Seq data | http://starbase.sysu.edu.cn/ |
| mESAdb [128] | Provides function and gene expression values of miRNAs in human, mouse, and zebra fish | http://konulab.fen.bilkent.edu.tr/mirna/ |
| Patrocles [137] | A database of SNP that are predicted to perturb miRNA-mediated gene regulation | http://www.patrocles.org/Patrocles.htm |
| RNAimmuno [138] | Provides information regarding the nonspecific effects generated in cells by RNA interference triggers and microRNA regulators | http://rnaimmuno.ibch.poznan.pl/ |
| miR2Disease [127] | Contains manually curated database of microRNAs that have been linked to different human diseases | http://www.mir2disease.org/ |
| HOCTAR [139] | Database of prediction target lists for 290 human intragenic miRNAs and tentative assignments of miRNA function based on Gene Ontology analyses of their predicted targets | http://hoctar.tigem.it/ |

(continued)

**Table 9.2** (continued)

| Database name | Brief description | Weblink |
|---|---|---|
| Oncomir | Contains miRNA expression profiles derived from various human tumors and select normal tissues | http://www.oncomir.umn.edu/ |
| miRdSNP [133] | A comprehensive database of dSNPs that affect gene expression and are related to various diseases | http://mirdsnp.ccr.buffalo.edu |
| miRecords [126] | Contains manually curated experimentally validated miRNA–target interaction sites | http://mirecords.biolead.org/ |
| miRGator [140] | Provides miRNA-associated gene expression, predicted targets, disease association, and genomic annotation | http://mirgator.kobic.re.kr:8080/MEXWebApp/ |
| DIANA-mirGen [37] | Provides miRNA annotation and their regulation by transcription factors. Contains expression profiles of miRNAs in several tissues and cell lines, SNP locations, predicted miRNA targets, and co-regulated miRNA targets biological pathways | http://diana.cslab.ece.ntua.gr/mirgen/ |
| miRSystem [141] | Provides predicted miRNA target gene and function/pathway analysis | http://mirsystem.cgm.ntu.edu.tw/ |
| miRvestigator [142] | Provides list of miRNAs targeting a list of co-expressed genes | http://mirvestigator.systemsbiology.net |
| PhenomiR [143] | Provides manually curated differentially expressed miRNAs in various diseases and biological processed | http://mips.helmholtz-muenchen.de/phenomir |
| PolymiRTS [144] | A database of SNPs affected miRNA targets | http://compbio.uthsc.edu/miRSNP/ |
| miRNEST [145] | A database of animal, plant, and virus miRNAs. Integrates literature data and 13 databases that includes miRNA sequences, small RNA-Seq, expression, SNPs, miRNA targets | http://lemur.amu.edu.pl/share/php/mirnest/home.php |

there are several programs for target prediction, the following features are in general used for predicting miRNA targets.

1. *Sequence complementarity*—The presence of complementary miRNA seed sequence in the target sequence is the most important feature for target prediction. The nucleotides 2–7 from 5′ of the miRNA are defined as the "seed region" of miRNA. Most of the programs prefer only Watson–Crick pairing in the seed region [38, 39]. Mismatches and wobble pairing in general are not allowed in the seed region.
2. *Thermodynamics*—The thermodynamics for pairing between the miRNA and the target sequence (also called duplex or "hybridization energy") is calculated, and only sites with the hybridization energy below a certain cutoff (pairs with the strongest interaction) are selected as the predicted targets [39, 40, 87]. Some algorithms use such energies to rank the final set of predictions. However, the assumption of the stronger the physical interaction the more likely the predicted miRNA–mRNA pair will be functional has been questioned. It was recently shown that the ranking according to hybridization energy does not improve the prediction accuracy once a perfect complementarity to the seed (positions 2–8) is present [88].
3. *Conservation*—The cross-species conservation is the best filter frequently used in the prediction of both miRNA genes and their target sites. Although conservation reduces the false positive predictions, it may sometimes miss species specific miRNAs that have evolved to act on non-conserved 3′ UTRs.
4. *Accessibility*—Consideration of embedded miRNA target sites accessibility in the folded mRNA can help in reducing the false positive cases. This feature basically looks for interaction of miRNA with target sites in the folded mRNA [41–44, 89].

Some of the well-known programs used for target prediction are briefly described below.

### *9.3.1 miRanda*

The miRanda algorithm [90, 91] identifies the miRNA targets in three different steps: (a) At first it identifies the best complementary pattern between miRNA and mRNA; (b) calculate the free energy change for the miRNA–mRNA duplex; (c) filter the predicted target by considering the evolutionary conservation. To find the complementary pattern it applies the Smith–Waterman alignment. During alignment it looks for Watson–Crick complementary pairs (A:U or G:C) or the wobble base pair (G:U). The Watson–Crick pairs are scored higher than the wobble pair. For all other base pairs (mismatches), a negative penalty value is applied. There is higher weight for 2–8 positions than other positions and the terminal miRNA

nucleotides. Fist nucleotide and last two nucleotides do not contribute to the alignment score, regardless of base pairing. The free energy change is calculated using folding routine present in the Vienna package [62]. miRanda uses PhastCons conservation scores [66] for identifying evolutionary conserved targets.

### *9.3.2 TargetScan*

TargetScan first searches for perfect complementary sequence of the seed region in the target site [38]. It then extends the complementary matches to the remaining region of the miRNA, allowing wobble base pairing. The identified targets are classified into three different categories depending on the length of the exact matching and on the occurrence of adenine at the first base of target sequence.

### *9.3.3 PicTar*

The PicTar program starts the miRNA target search by generating a set of orthologous 3′ UTR regions. The orthologous regions are identified using multiple sequence alignment from eight vertebrates. Looking into the orthologous sites reduces false positive cases. It scans the orthologous regions looking for perfect complementary match to miRNA seed region. The target sites are filtered based on optimal free energy of the predicted miRNA–mRNA duplex (defined as anchors). The target UTRs are further filtered based on total numbers of anchors found in the UTR. The filtered identified targets are then scored using a Hidden Markov Model (HMM) maximum-likelihood approach. The PicTar scoring system takes into consideration the fact that an mRNA can be targeted by multiple miRNAs.

### *9.3.4 DIANA-microT*

This algorithm is based on several parameters known as miRNA recognition elements (MREs) calculated individually for each miRNA [37, 92]. The parameters include complementary sequence pattern, conservation, binding site structural accessibility, nucleotide composition around the flanking sites of binding region, and closeness of one binding site to another in the same target sequence. The final predicted score of the target gene is the weighted sum of MREs. The final score is compared with the score obtained from shuffled sequence with the same dinucleotide composition as that of 3′ UTR sequence.

### 9.3.5 *RNAHybrid*

RNAHybrid [40, 93] is an improvement of the classical RNA secondary structure prediction programs: MFold and Vienna package [62, 94, 95]. RNAHybrid instead of folding a single sequence it scans the target sequence to identify potential regions that can form a thermodynamically duplex with a specific miRNA in an energetically optimal way. RNAHybrid does not allow intramolecular base pairing and branching structures.

### 9.3.6 *Rna22*

Rna22 is a pattern-based approach for the identification of miRNA target sites and miRNAs [96]. The prediction algorithm first looks for statistically significant complementary pattern in the mRNAs based on known miRNA pattern. It then predicts the miRNAs that targets a given mRNA. The candidate miRNAs are identified based on miRNA–mRNA duplex free energy calculated using Vienna package. Since knowledge of miRNA is not needed in prior, Rna22 program can identify mRNAs targeted by novel miRNAs. It does not use conservation filter and hence can identify miRNA targets, which are not conserved across species.

### 9.3.7 *PITA*

The secondary structure of the mRNA plays a very crucial role in miRNA targeting. This algorithm predicts targets based on the miRNA interaction with binding site embedded in the folded mRNA [41]. PITA first scans for complementary sequence in the seed region, allowing one mismatch or one wobble pair. It then calculates the energy-based score for miRNA–target interaction. The calculated score is the difference between the energy gained by miRNA binding to the target site and the loss of free energy by unpairing the target site to make it accessible to the miRNA [97].

Almost all miRNA target prediction programs use key features, such as complementarity to 5′ seed of miRNAs and evolutionary conservation. However, not all miRNA target sites are conserved and adhere to canonical seed complementarity, leading to missing a large fraction of the targeted mRNAs and additionally predicting a large number of false positives. In order to overcome this limitation, recent algorithms combine multiple novel target site features through machine learning tools. For example, DIANA-microT-ANN program combines multiple novel target site features through an artificial neural network (ANN) model, which was trained using published high-throughput data measuring the change of protein levels after

miRNA overexpression [98]. Similarly, in PACCMIT algorithm [89], the concept of partial accessibility of the complementary sites in the 3′ UTR was incorporated into a successful statistical framework to predict highly precise miRNA target sites. This method does not require conservation information and considers multiple binding sites in calculating a single score.

## 9.4 MiRNA Editing

Adenosine-to-inosine (A-to-I) RNA editing is a co-transcriptional or posttranscriptional event that changes adenosine to inosine in a double-stranded RNA [47, 99, 100]. RNA sequence editing can result in a protein which is different from the one encoded by the genome and thereby increase protein diversity [100–102]. The A-to-I transition is catalyzed by enzymes known as adenosine deaminases acting on RNA (ADARs) [47, 103, 104]. Long primary miRNA (pri-miRNA) transcript is processed into ~60–70 nucleotide precursor miRNA (pre-miRNA) that is further processed to generate ~22 nucleotide mature miRNA(s) [105–108]. Pri-miRNAs contain a short double-stranded hairpin like structure and can generate up to six miRNAs. Due to double stranded structure of pri-miRNA it can act as a substrate for ADAR [109–113]. Although the extent of A-to-I editing was low, targeted adenosines can influence the expression and function of miRNAs. If the editing of pri-miRNA is in the seed region then the target gene will change. For example, A-to-I editing in the seed region of mouse miR-376 alters the gene target and has severe effects on cellular processes [111].

In the last few years several studies have reported editing sites in miRNA including novel several editing sites using small RNA-Seq data [46, 48, 114, 115]. The major challenge in detecting A-to-I editing site is to discriminate sequencing and alignment errors from actual editing sites. Although there does not exist a gold standard program for predicting A-to-I editing from small RNA-Seq dataset, a systematic approach needs to be followed to identify actual A-to-I editing sites. The following bioinformatics steps are followed to detect true A-to-I editing site from small RNA-Seq samples.

1. *Preprocessing*—The 3′ adapter sequence should be properly trimmed before alignment. No mismatches are allowed while removing adapter sequences. Some studies also suggest trimming of last two bases from the mature miRNA because they undergo large-scale RNA modifications in form of adenylation and uridylation [46, 114, 116]. In addition, many studies also put a length filter because most of the mature miRNAs are of length $\geq$20 bp and $\leq$25 bp.
2. *Alignment*—The processed reads are aligned to either known pri-miRNA or reference genome. If the reads are aligned to the reference genome then the reads that fall in the known pri-miRNAs are taken further for analysis. Reads can be cross-mapped to multiple location in the genome due to presence of similar sequence pattern in the genome, sequencing errors or posttranscriptional modifications in the read. For example let-7b and let-7c mature miRNA sequence

differs by one base. An error or editing in let-7b can lead to cross mapping to let-7c. The read counts for multiple mapping reads are commonly divided equally between the aligned positions, randomly aligned to one of the positions, or removed completely. In [117], an expectation-maximization based approach was proposed to avoid error due to cross mapping of miRNAs. Some of the studies allow only up to one mismatch during alignment.

3. *Editing site prediction*—For identification of true editing sites from read alignment file, first different filters are applied and then a statistical test is performed. The filtering options include the following: position of mismatch in the read, quality score of the mismatched nucleotide, read depth at the mismatch location, minimum editing percentage. After filtering some of the potential editing sites statistical tests such as Benjamini–Hochberg test, Chi-Square test, Bonferroni's correction [46, 115, 117] are applied to identify A-to-I editing sites in the miRNAs.

## 9.5 Databases of MiRNAs, Expression Profiles and MiRNA Target Sites

Increasingly it has been shown that microRNAs play a significant role in regulating gene expression within the animal and plant kingdom. There has been a rapid accumulation of miRNA knowledge in the past few years. Rapid retrieval of miRNA information is very important for researchers. Several miRNA databases have been established to systematically organize miRNA related data (Table 9.2). The databases catalogue known miRNAs, experimentally validated and/or computationally predicted miRNA targets, miRNA expression information in different tissues, miRNA–disease association. Below we briefly discuss some of the widely used and important miRNA databases.

### *9.5.1 miRBase*

miRBase is one of the earliest and most widely used databases that provides integrated interfaces to comprehensive miRNA sequence data, annotation and predicted mRNA targets [118]. The latest release of miRBase (Release 19, August 2012) contains 21,264 entries representing precursor miRNAs, 25,141 mature miRNAs of 193 species [80]. Data are collected from publications reporting novel miRNAs and also from datasets submitted by user to database such as NCBI GEO [119]. It provides a consistent nomenclature for naming newly discovered miRNAs. miRBase also assigns a stable accession number to each stem-loop and mature sequence. The latest version of miRBase provides an interface to view reads from RNA deep-sequencing data mapped to microRNA loci.

### 9.5.2 miRDB

miRDB is a database of miRNA target prediction and functional annotation [120]. It follows wiki model to keep track of all the latest development in the miRNA field. All miRNA researchers are invited to continuously contribute miRNA functional annotations and actively interact with each other. The miRNA targets maintained at miRDB was predicted using MirTarget2 program [121]. MirTarget2 study genes impacted by miRNAs and train prediction model using support vector machine (SVM). miRDB maintains predicted miRNA targets in five species: human, mouse, rat, dog, and chicken.

### 9.5.3 microRNA.org

MicroRNA.org is a database of predicted miRNA targets and expression values [91]. MiRNA targets are predicted using miRanda algorithm. It provides an interface to explore miRNA target genes and their expression in different tissues. The expression value provided by the database is in the normalized format. The August 2010 release of the database provides information for human, mouse, rat, fruitfly, and nematode.

### 9.5.4 microPIR

microRNA-Promoter Interaction Resource (microPIR) is a database that provides miRNA targets present within human promoter regions [122]. The RNAHybrid program was used to predict the target sites in the promoter region. More than 15 million target sites were identified by scanning 5,000 bp upstream of all human genes, on both sense and antisense strands. To increase the sensitivity of target site selection it also provides validated AGO protein binding sites information that has been reported to be involved in promoter-targeting miRNAs [123]. microPIR also integrates various other information including EST expression, sequence conservation across vertebrate species, transposable element and repeat position, transcription factor binding sites, CpG islands, and SNPs. Integrating other genomic features with target sites will help the researcher in taking intelligent decision. The complete information is presented on a genome browser.

### 9.5.5 mirTarBase

mirTarBase is a curated database of miRNA–target interactions (MTIs) sites that have been experimentally validated [124]. The articles related to the miRNAs are first

identified using a systematic text-mining process. Then each article is carefully reviewed by at least two developers at miTarBase to extract MTIs reported in the articles. The latest release (Release 2.5, Oct. 15, 2011) consists of 4,270 MTI sites and 669 miRNAs from 14 different species. mirTarBase also provides comparison summary with other databases including TarBase [125], miRecords [126], and miR2Disease [127].

### *9.5.6 TarBase*

The latest version of TarBase implements advance text mining-assisted human pipeline for paper curation. The pipeline included miRNA target association identification, name entity recognition and scoring techniques to select publications from PubMed. After an article is selected for curation, the curator extracts several relevant information including miRNAs, the target genes and experimental validation information. TarBase 6.0 version of database contains 65,814 experimentally validated miRNA–gene interactions [125].

### *9.5.7 mESAdb*

microRNA expression and sequence analysis database (mESAdb) is a database for multivariate analysis of miRNA sequence and expression from multiple taxa [128]. It allows user to extract miRNA expression for a subset of miRNA selected manually or by motif. It also performs meta-analysis for comparative analysis of function and expression for microRNA subset across multiple taxa, including human, mouse, and zebra fish. mESAdb also perform gene enrichment analysis using KEGG, GO, HUGE [129–131] annotation databases. Users can also upload their data to perform analysis.

### *9.5.8 miR2Disease*

miR2Disease is a manually curated database of microRNAs that have been linked to different human diseases [127]. It contains several information on a miRNA–disease relationship including miRNA id, disease name, brief description of the relationship, relative expression pattern of miRNA in the disease state, expression detection method, and experimental validated miRNA targets. The disease terms are organized based upon a controlled medical vocabulary (http://diseaseontology.sourceforge.net/). The database can be queried using miRNA name or disease or target gene. Additionally, it provides external link to several other miRNA databases. The current version of miR2Disease covers 163 diseases, 349 miRNAs.

### *9.5.9 miRdSNP*

Recent findings indicate that Single Nucleotide Polymorphisms (SNPs) could create, destroy, or modify the efficiency of miRNA binding to the 3′ UTR of a gene, resulting in altered target gene expression. For example, genome-wide bioinformatics analysis predicted approximately 64 % of transcribed SNPs as target SNPs that can modify (increase/decrease) the binding energy of putative miRNA–mRNA duplexes by greater than 90 % [132]. miRdSNP is a comprehensive database of dSNPs that affect gene expression that are related to various diseases [133]. miRdSNP manually curates dSNPs from PubMed articles linked in Enterz and overlap the 3′ UTRs of human. The miRNA targets are predicted using TargetScan and PicTar algorithms. It also helps in studying the genomic distance between miRNA target sites and dSNPs. The latest release of miRdSNP contains information about 786 disease–SNP pairs, where 180 of these are predicted to disrupt the miRNA targets.

## 9.6 Worked Example

We downloaded small RNA-Seq data from NCBI SRA database (Accession = SRP006574) that was deposited by a recently published work on breast cancer. We processed 14 samples (Breast_Normal = 7; Breast_IDC = 7) from the published study. Following steps are followed for analysis of small RNA-Seq datasets.

1. Preprocessing: First the read quality and base composition of the samples are studied. The average base composition and quality distribution obtained for one of the samples is shown in Fig. 9.1. Figure 9.1a clearly indicates random base composition in the initial portion of the reads, whereas the later portion reads show some bias. The bias is due to multiplexing index and 3′ adapter sequences present in each read. Figure 9.1b shows the quality of bases across reads in the sample. More than 90 % of the bases have read quality ≥Q20 (Phred score). The reads with average base quality ≥Q20 were taken further for analysis. The adapter sequences and multiplex sequence are later on trimmed from each read of the sample. The read length distribution of trimmed reads for all samples is shown in Fig. 9.2. Majority of the trimmed reads are of length 21–23 bp that is the range for mature miRNAs (Table S1).
2. Alignment and annotation: The trimmed reads are aligned to primary miRNA sequences, noncoding RNAs (rRNA, tRNA, snoRNA, snRNA), and genome. The alignment is performed in following order: rRNA → tRNA → miRNA → sno RNA → snRNA → chrM → genome. The alignment to different RNA class and genome is shown in Fig. 9.3 (Tables 9.3 and S2). The majority of the reads map to the primary miRNA sequences. The read length distribution for the RNA class was also studied. It can be clearly seen that miRNA class has a specific pattern,

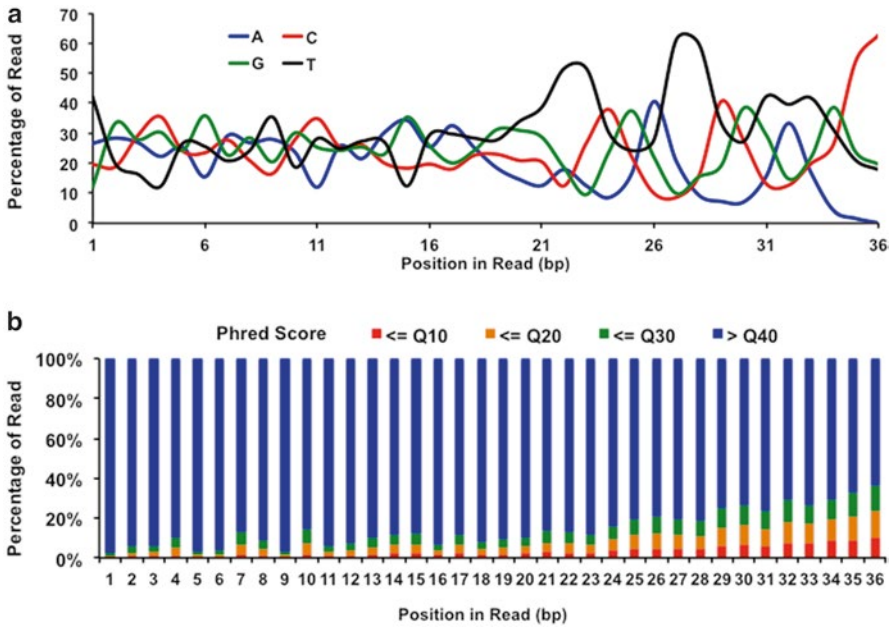


**Fig. 9.1** (**a**) Base composition distribution, (**b**) Base quality distribution for one of the small RNA sample

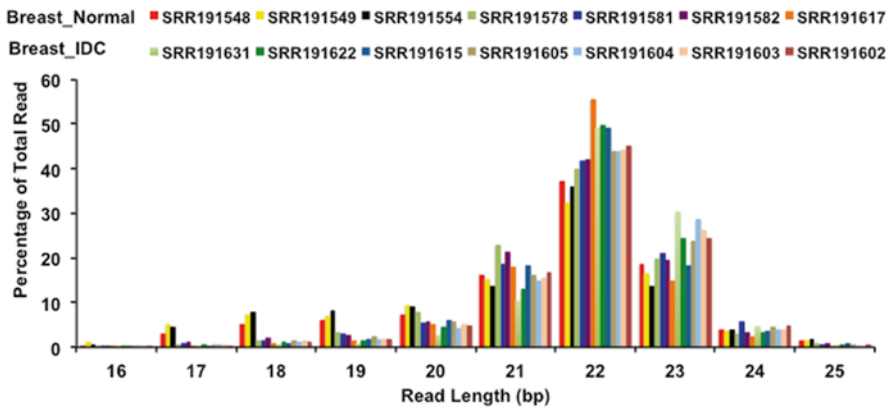


**Fig. 9.2** Read length distribution after trimming of adapter sequences from small RNA samples

peaking at 22 bases whereas other small RNA classes does not show any specific pattern (Fig. 9.4, Table S3).

3. MiRNA expression analysis: The total reads overlapping pre-miRNA regions are counted for miRNA expression quantification (Table S4). The multiple mapped read count is commonly divided equally between the aligned positions.

**Table S1** Read length summary after trimming and pre-processing

| | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191631 | SRR 191622 | SRR 191615 | SRR 191605 | SRR 191604 | SRR 191603 | SRR 191602 |
| 16 | 3516 | 2939 | 2635 | 446 | 570 | 647 | 174 | 491 | 654 | 291 | 431 | 341 | 370 | 222 |
| 17 | 23803 | 12310 | 17639 | 2493 | 2875 | 3147 | 1710 | 4602 | 6765 | 2880 | 7697 | 6153 | 6968 | 5907 |
| 18 | 41254 | 17685 | 29419 | 6187 | 4824 | 5106 | 3466 | 8241 | 11668 | 8096 | 23825 | 15727 | 18437 | 18394 |
| 19 | 46978 | 17066 | 31444 | 14305 | 8987 | 6457 | 5894 | 10877 | 14717 | 16441 | 33999 | 21956 | 23519 | 28110 |
| 20 | 57293 | 22536 | 34458 | 33886 | 16319 | 13384 | 19612 | 36322 | 39302 | 50146 | 84913 | 50313 | 61143 | 77718 |
| 21 | 125927 | 35844 | 51377 | 98457 | 54477 | 49345 | 66582 | 135161 | 112077 | 152569 | 235030 | 176137 | 184083 | 269660 |
| 22 | 291628 | 76928 | 135093 | 172389 | 122231 | 97861 | 204314 | 636819 | 423792 | 407053 | 637872 | 521916 | 517337 | 724009 |
| 23 | 145941 | 39094 | 51343 | 85053 | 61843 | 45451 | 55552 | 393919 | 209862 | 151173 | 346773 | 339011 | 306350 | 393338 |
| 24 | 32287 | 8615 | 14858 | 12949 | 16916 | 8126 | 8941 | 62179 | 28339 | 31264 | 68716 | 48656 | 46835 | 76909 |
| 25 | 12109 | 3374 | 7349 | 3952 | 2293 | 1974 | 1401 | 6442 | 5747 | 6838 | 8375 | 5021 | 5095 | 8201 |
| Total | 780736 | 236391 | 375615 | 430117 | 291335 | 231498 | 367646 | 1295053 | 852923 | 826751 | 1447631 | 1185231 | 1170137 | 1602468 |

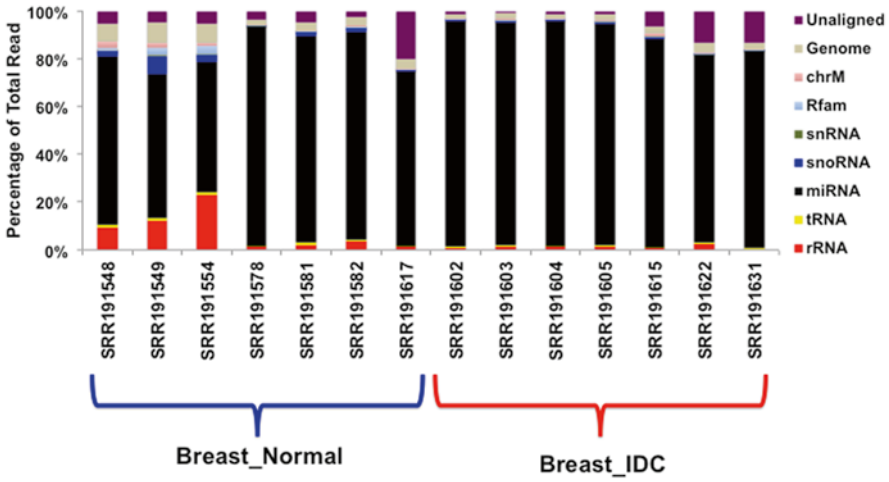


**Fig. 9.3** Alignment summary of trimmed reads to different RNA class across all samples

**Table 9.3** Small RNA-Seq read summary

| Tissue type | NCBI SRA ID | Total raw reads | Total reads with quality ≥Q20 | Total reads after adapter trimming | Total reads aligned |
|---|---|---|---|---|---|
| Breast normal | SRR191548 | 880,564 | 863,369 | 780,736 | 740,595 |
| | SRR191549 | 288,539 | 276,906 | 236,391 | 225,186 |
| | SRR191554 | 436,890 | 420,729 | 375,615 | 355,350 |
| | SRR191578 | 500,705 | 439,419 | 430,117 | 414,695 |
| | SRR191581 | 342,815 | 303,855 | 291,355 | 276,606 |
| | SRR191582 | 283,125 | 241,959 | 231,498 | 226,208 |
| | SRR191617 | 388,816 | 376,488 | 367,646 | 294,371 |
| Breast IDC | SRR191602 | 1,631,075 | 1,615,298 | 1,602,468 | 1,584,151 |
| | SRR191603 | 1,191,991 | 1,180,635 | 1,170,137 | 1,158,625 |
| | SRR191604 | 1,203,352 | 1,195,729 | 1,185,231 | 1,168,644 |
| | SRR191605 | 1,472,658 | 1,458,642 | 1,447,631 | 1,430,470 |
| | SRR191615 | 1,021,848 | 831,239 | 826,751 | 773,983 |
| | SRR191622 | 890,845 | 863,530 | 852,923 | 738,638 |
| | SRR191631 | 1,340,601 | 1,303,399 | 1,295,053 | 1,123,003 |

Differential miRNA expression analysis was performed using DESeq program. MiRNAs with an adjusted $p$-value $\leq 0.01$ and at least 10 reads in four samples of the seven sample of either normal breast or IDC are selected as candidates. Our analysis identified 8 up-regulated and 23 down-regulated miRNAs in breast IDC samples as compared to breast normal (Table S5).

**Table S2** Read count summary into various RNA class and genome

| | Breast_Normal | | | | | | | Breast_IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Annotation class | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| rRNA | 72291 | 28851 | 86726 | 5800 | 6513 | 8585 | 4261 | 17074 | 18326 | 14637 | 22960 | 5231 | 22398 | 6025 |
| tRNA | 11290 | 3046 | 3203 | 1348 | 1847 | 1862 | 1621 | 7835 | 7579 | 4691 | 7478 | 3160 | 6450 | 2834 |
| miRNA | 551457 | 142272 | 205844 | 394773 | 252351 | 200926 | 268824 | 1510152 | 1089360 | 1116548 | 1342425 | 723523 | 665503 | 1068212 |
| snoRNA | 15783 | 17524 | 10707 | 1298 | 4565 | 3235 | 1526 | 6442 | 5114 | 3701 | 7506 | 4906 | 3352 | 2135 |
| snRNA | 1934 | 1269 | 778 | 144 | 175 | 195 | 107 | 541 | 520 | 430 | 545 | 263 | 555 | 184 |
| Rfam | 9926 | 6944 | 13221 | 983 | 1156 | 1503 | 1036 | 4725 | 5250 | 3826 | 5918 | 1696 | 3245 | 4041 |
| chrM | 19966 | 4534 | 4225 | 1372 | 1381 | 1675 | 750 | 4803 | 4406 | 2979 | 6429 | 9735 | 3257 | 5189 |
| Genome | 57948 | 20746 | 30646 | 8977 | 8618 | 8227 | 16246 | 32579 | 28070 | 21832 | 37209 | 25469 | 33878 | 34383 |
| Unaligned | 43441 | 11205 | 20265 | 15422 | 14729 | 5290 | 73275 | 18317 | 11512 | 16587 | 17161 | 52768 | 114285 | 172050 |

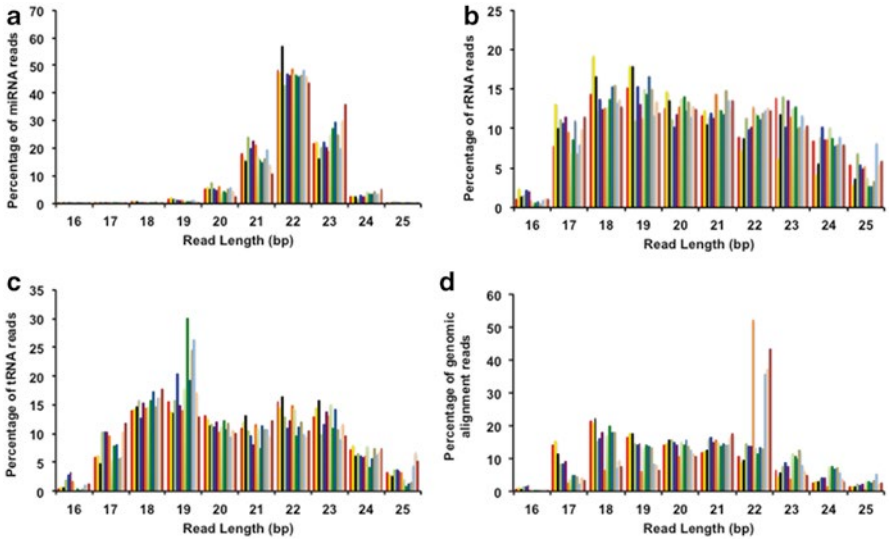


**Fig. 9.4** Read length distribution for the sample aligned to different RNA class, (**a**) miRNA, (**b**) rRNA, (**c**) tRNA, (**d**) genome

## 9.7 Notes

1. Despite great progress, miRNA target prediction by computational approaches alone is still far from perfect. Almost all miRNA target prediction programs use key features, such as complementarity to 5′ seed of miRNAs and evolutionary conservation. While these features reduce the false positive prediction rate, not all miRNA target sites are conserved and adhere to canonical seed complementarity. Recent results indicate significant promise for energy-based miRNA target prediction that includes a broader range of targets without having to use conservation or impose stringent seed match rules [134]. Inclusion of such features is not considered in the approaches suggested in this chapter, although we expect development of prediction programs in near future.
2. Recent evidence from NGS data, such as RNA-seq, suggest that a gene can yield transcript variants that differ either in their regulatory UTRs or/and protein coding regions, thereby expanding the complexity of mammalian transcriptomes and proteomes. The transcript variants thus generated should be considered separately in miRNA target prediction programs, because one transcript variant of gene could be a target while the others may not depending on their 3′ UTRs.
3. An automated pipeline would soon be made available at http://bioinformatics.wistar.upenn.edu. The users may contact the author for R code.

**Table S3** Read length summary for various annotation class

| rRNA | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 53 | 24 | 26 | 27 | 53 | 64 | 28 | 11 | 32 | 6 | 31 | 38 | 77 | 41 |
| 17 | 675 | 190 | 154 | 141 | 193 | 195 | 158 | 440 | 605 | 386 | 427 | 187 | 671 | 339 |
| 18 | 1599 | 434 | 476 | 215 | 234 | 285 | 234 | 1154 | 1196 | 812 | 1097 | 514 | 1052 | 505 |
| 19 | 1757 | 424 | 436 | 215 | 378 | 279 | 227 | 1403 | 2281 | 908 | 1838 | 837 | 1111 | 371 |
| 20 | 1496 | 385 | 368 | 158 | 207 | 227 | 169 | 881 | 932 | 508 | 888 | 302 | 689 | 288 |
| 21 | 1250 | 370 | 423 | 143 | 179 | 151 | 190 | 856 | 564 | 540 | 808 | 342 | 616 | 352 |
| 22 | 1767 | 443 | 529 | 174 | 204 | 229 | 244 | 1122 | 737 | 531 | 899 | 316 | 610 | 298 |
| 23 | 1478 | 443 | 508 | 135 | 214 | 257 | 215 | 1185 | 834 | 671 | 809 | 284 | 747 | 274 |
| 24 | 831 | 243 | 200 | 88 | 114 | 110 | 101 | 615 | 325 | 266 | 556 | 200 | 441 | 214 |
| 25 | 384 | 90 | 83 | 52 | 71 | 65 | 55 | 168 | 73 | 63 | 125 | 140 | 436 | 152 |
| Total | 11290 | 3046 | 3203 | 1348 | 1847 | 1862 | 1621 | 7835 | 7579 | 4691 | 7478 | 3160 | 6450 | 2834 |

| miRNA | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 151 | 105 | 58 | 70 | 38 | 35 | 13 | 53 | 47 | 50 | 62 | 45 | 35 | 48 |
| 17 | 1535 | 640 | 549 | 348 | 288 | 254 | 193 | 1203 | 1021 | 1384 | 1366 | 372 | 509 | 451 |
| 18 | 5062 | 1395 | 1694 | 3162 | 1267 | 1101 | 1022 | 6117 | 5481 | 6074 | 8203 | 2975 | 2449 | 2360 |
| 19 | 8688 | 3404 | 3850 | 7268 | 3045 | 2331 | 3282 | 14337 | 11257 | 10791 | 17172 | 10141 | 5869 | 5622 |
| 20 | 30426 | 8248 | 10774 | 31056 | 13166 | 10001 | 16443 | 66936 | 51345 | 42668 | 72621 | 42908 | 29864 | 29035 |
| 21 | 99697 | 24259 | 32286 | 95224 | 50876 | 45810 | 57427 | 258976 | 174744 | 168534 | 221536 | 143079 | 93966 | 115973 |
| 22 | 266392 | 67827 | 117729 | 168457 | 118207 | 93968 | 131235 | 712132 | 507745 | 513856 | 625226 | 348973 | 306457 | 469278 |
| 23 | 121680 | 31950 | 33337 | 80306 | 56794 | 41551 | 51869 | 380303 | 296274 | 330692 | 333931 | 145440 | 201160 | 385208 |
| 24 | 15804 | 4010 | 5018 | 6470 | 7901 | 5316 | 6839 | 66071 | 39314 | 39785 | 58179 | 26979 | 22938 | 56591 |
| 25 | 2022 | 434 | 549 | 2412 | 769 | 559 | 501 | 4024 | 2132 | 2714 | 4129 | 2611 | 2256 | 3646 |
| Total | 551457 | 142272 | 205844 | 394773 | 252351 | 200926 | 268824 | 1510152 | 1089360 | 1116548 | 1342425 | 723523 | 665503 | 1068212 |

| tRNA | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 53 | 24 | 26 | 27 | 53 | 64 | 28 | 11 | 32 | 6 | 31 | 38 | 77 | 41 |
| 17 | 675 | 190 | 154 | 141 | 193 | 195 | 158 | 440 | 605 | 386 | 427 | 187 | 671 | 339 |
| 18 | 1599 | 434 | 476 | 215 | 234 | 285 | 234 | 1154 | 1196 | 812 | 1097 | 514 | 1052 | 505 |
| 19 | 1757 | 424 | 436 | 215 | 378 | 279 | 227 | 1403 | 2281 | 908 | 1838 | 837 | 1111 | 371 |
| 20 | 1496 | 385 | 368 | 158 | 207 | 227 | 169 | 881 | 932 | 508 | 888 | 302 | 689 | 288 |
| 21 | 1250 | 370 | 423 | 143 | 179 | 151 | 190 | 856 | 564 | 540 | 808 | 342 | 616 | 352 |
| 22 | 1767 | 443 | 529 | 174 | 204 | 229 | 244 | 1122 | 737 | 531 | 899 | 316 | 610 | 298 |
| 23 | 1478 | 443 | 508 | 135 | 214 | 257 | 215 | 1185 | 834 | 671 | 809 | 284 | 747 | 274 |
| 24 | 831 | 243 | 200 | 88 | 114 | 110 | 101 | 615 | 325 | 266 | 556 | 200 | 441 | 214 |
| 25 | 384 | 90 | 83 | 52 | 71 | 65 | 55 | 168 | 73 | 63 | 125 | 140 | 436 | 152 |
| Total | 11290 | 3046 | 3203 | 1348 | 1847 | 1862 | 1621 | 7835 | 7579 | 4691 | 7478 | 3160 | 6450 | 2834 |
| snoRNA | Breast Normal | | | | | | | Breast IDC | | | | | | |
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 199 | 316 | 147 | 30 | 106 | 82 | 15 | 12 | 11 | 23 | 23 | 26 | 29 | 43 |
| 17 | 1625 | 2082 | 1308 | 113 | 383 | 275 | 146 | 301 | 375 | 339 | 470 | 241 | 303 | 217 |
| 18 | 2449 | 2941 | 1793 | 145 | 454 | 295 | 123 | 651 | 635 | 468 | 673 | 364 | 308 | 187 |
| 19 | 2346 | 2420 | 1489 | 140 | 559 | 419 | 132 | 787 | 607 | 527 | 1344 | 429 | 327 | 200 |
| 20 | 3531 | 4392 | 2414 | 141 | 506 | 412 | 132 | 785 | 714 | 694 | 846 | 481 | 348 | 230 |
| 21 | 2021 | 2071 | 1370 | 146 | 573 | 409 | 198 | 809 | 626 | 451 | 1517 | 629 | 457 | 268 |
| 22 | 1524 | 1436 | 1042 | 199 | 625 | 445 | 258 | 1154 | 860 | 481 | 1098 | 1018 | 658 | 385 |
| 23 | 1272 | 1244 | 793 | 180 | 651 | 453 | 206 | 1016 | 709 | 394 | 807 | 774 | 480 | 306 |
| 24 | 454 | 366 | 222 | 107 | 394 | 229 | 185 | 577 | 356 | 221 | 473 | 449 | 232 | 150 |
| 25 | 362 | 256 | 129 | 97 | 314 | 216 | 131 | 350 | 221 | 103 | 255 | 495 | 210 | 149 |
| Total | 15783 | 17524 | 10707 | 1298 | 4565 | 3235 | 1526 | 6442 | 5114 | 3701 | 7506 | 4906 | 3352 | 2135 |

(continued)

**Table S3** (continued)

| snRNA | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 1303 | 977 | 560 | 22 | 24 | 41 | 2 | 16 | 56 | 70 | 61 | 5 | 9 | 3 |
| 17 | 61 | 39 | 13 | 16 | 15 | 17 | 12 | 42 | 55 | 27 | 31 | 20 | 53 | 20 |
| 18 | 99 | 68 | 39 | 12 | 23 | 20 | 17 | 86 | 95 | 54 | 76 | 33 | 88 | 29 |
| 19 | 108 | 45 | 33 | 8 | 19 | 18 | 7 | 72 | 65 | 56 | 60 | 30 | 72 | 23 |
| 20 | 99 | 38 | 40 | 21 | 28 | 23 | 11 | 109 | 71 | 68 | 72 | 37 | 97 | 25 |
| 21 | 93 | 36 | 29 | 12 | 16 | 21 | 12 | 63 | 54 | 58 | 79 | 40 | 65 | 34 |
| 22 | 70 | 26 | 26 | 12 | 12 | 17 | 11 | 59 | 58 | 32 | 81 | 33 | 60 | 24 |
| 23 | 49 | 20 | 20 | 15 | 17 | 17 | 16 | 39 | 36 | 35 | 45 | 23 | 46 | 11 |
| 24 | 36 | 14 | 10 | 13 | 9 | 12 | 10 | 45 | 19 | 23 | 32 | 25 | 38 | 5 |
| 25 | 16 | 6 | 8 | 13 | 12 | 9 | 9 | 10 | 11 | 7 | 8 | 17 | 27 | 10 |
| Total | 1934 | 1269 | 778 | 144 | 175 | 195 | 107 | 541 | 520 | 430 | 545 | 263 | 555 | 184 |
| Rfam | Breast Normal | | | | | | | Breast IDC | | | | | | |
| Read Length | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 365 | 535 | 294 | 52 | 60 | 80 | 33 | 34 | 58 | 30 | 71 | 35 | 126 | 111 |
| 17 | 2014 | 1316 | 2273 | 260 | 327 | 395 | 290 | 1088 | 1361 | 848 | 1382 | 520 | 1196 | 1127 |
| 18 | 3137 | 2253 | 3482 | 261 | 340 | 457 | 383 | 1657 | 1956 | 1434 | 2115 | 583 | 999 | 1103 |
| 19 | 1881 | 1080 | 2323 | 144 | 160 | 223 | 155 | 886 | 820 | 705 | 1022 | 289 | 471 | 503 |
| 20 | 1229 | 779 | 2340 | 82 | 132 | 134 | 95 | 588 | 548 | 409 | 676 | 155 | 259 | 349 |
| 21 | 584 | 420 | 1068 | 82 | 66 | 114 | 47 | 252 | 279 | 234 | 390 | 74 | 119 | 395 |
| 22 | 397 | 315 | 759 | 63 | 50 | 69 | 27 | 125 | 146 | 109 | 173 | 27 | 48 | 347 |
| 23 | 190 | 148 | 399 | 36 | 18 | 24 | 4 | 63 | 59 | 35 | 52 | 7 | 20 | 84 |
| 24 | 110 | 75 | 228 | 2 | 2 | 6 | 1 | 23 | 17 | 15 | 31 | 4 | 4 | 20 |
| 25 | 19 | 23 | 55 | 1 | 1 | 1 | 1 | 9 | 6 | 7 | 6 | 2 | 3 | 2 |
| Total | 9926 | 6944 | 13221 | 983 | 1156 | 1503 | 1036 | 4725 | 5250 | 3826 | 5918 | 1696 | 3245 | 4041 |

| | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chrM | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 89 | 32 | 27 | 19 | 23 | 19 | 10 | 3 | 11 | 11 | 17 | 53 | 30 | 96 |
| 17 | 4038 | 1034 | 1078 | 225 | 260 | 269 | 93 | 396 | 551 | 478 | 702 | 518 | 344 | 632 |
| 18 | 5610 | 669 | 660 | 271 | 225 | 384 | 104 | 759 | 828 | 514 | 1155 | 1096 | 378 | 678 |
| 19 | 1309 | 242 | 297 | 122 | 133 | 117 | 83 | 642 | 497 | 249 | 633 | 1893 | 419 | 451 |
| 20 | 1142 | 362 | 282 | 133 | 159 | 205 | 99 | 748 | 603 | 384 | 1168 | 2355 | 669 | 1330 |
| 21 | 3760 | 1266 | 968 | 204 | 180 | 241 | 118 | 679 | 643 | 520 | 817 | 585 | 338 | 586 |
| 22 | 2482 | 607 | 528 | 127 | 152 | 166 | 72 | 554 | 439 | 322 | 565 | 620 | 311 | 425 |
| 23 | 578 | 133 | 167 | 116 | 97 | 108 | 62 | 435 | 344 | 185 | 484 | 613 | 273 | 327 |
| 24 | 626 | 115 | 143 | 108 | 99 | 98 | 71 | 396 | 372 | 216 | 660 | 872 | 276 | 384 |
| 25 | 332 | 74 | 75 | 47 | 53 | 68 | 38 | 191 | 118 | 100 | 228 | 1130 | 219 | 280 |
| Total | 19966 | 4534 | 4225 | 1372 | 1381 | 1675 | 750 | 4803 | 4406 | 2979 | 6429 | 9735 | 3257 | 5189 |

| Genomic | Breast Normal | | | | | | | Breast IDC | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aligned ReadLength | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 | SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 |
| 16 | 505 | 263 | 273 | 135 | 118 | 146 | 33 | 25 | 42 | 42 | 54 | 38 | 82 | 83 |
| 17 | 8128 | 3206 | 3507 | 744 | 713 | 748 | 408 | 1242 | 1391 | 1055 | 1698 | 603 | 1454 | 1126 |
| 18 | 12387 | 4334 | 6808 | 1378 | 1380 | 1472 | 1040 | 5581 | 5609 | 3908 | 6712 | 1834 | 3281 | 2593 |
| 19 | 9531 | 3685 | 5370 | 1317 | 1206 | 1185 | 994 | 4372 | 3926 | 3013 | 4926 | 2191 | 2763 | 2217 |
| 20 | 8154 | 3013 | 4768 | 1412 | 1296 | 1169 | 1744 | 4909 | 3971 | 3424 | 5102 | 3173 | 3883 | 3742 |
| 21 | 6806 | 2523 | 3884 | 1437 | 1416 | 1222 | 2539 | 4825 | 3916 | 3208 | 5281 | 3658 | 5661 | 6011 |
| 22 | 6192 | 1847 | 2873 | 1310 | 1198 | 1135 | 8481 | 4324 | 3227 | 2963 | 4819 | 9117 | 12647 | 14948 |
| 23 | 3664 | 955 | 1695 | 697 | 744 | 625 | 626 | 3782 | 2997 | 2167 | 4649 | 2032 | 2023 | 1687 |
| 24 | 1627 | 624 | 959 | 329 | 373 | 343 | 264 | 2470 | 2166 | 1459 | 2756 | 1483 | 1261 | 1067 |
| 25 | 954 | 296 | 509 | 218 | 174 | 182 | 117 | 1049 | 825 | 593 | 1212 | 1340 | 823 | 909 |
| Total | 57948 | 20746 | 30646 | 8977 | 8618 | 8227 | 16246 | 32579 | 28070 | 21832 | 37209 | 25469 | 33878 | 34383 |

**Table S4** List of expressed miRNAs

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr13 | 41675154 | 41675236 | hsa-mir-3168 | - | 81 | 14 | 37 | 3 | 6 |
| chr7 | 127847924 | 127847996 | hsa-mir-129-1 | + | 3 | 6 | 0 | 1 | 5 |
| chr14 | 101528386 | 101528455 | hsa-mir-377 | + | 445 | 69 | 19 | 52 | 136 |
| chr16 | 2185977 | 2186130 | hsa-mir-3180-5 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 146271221 | 146271305 | hsa-mir-513c | - | 0 | 0 | 0 | 0 | 0 |
| chr4 | 185859536 | 185859594 | hsa-mir-4455 | - | 3 | 2 | 1 | 0 | 0 |
| chr14 | 107259099 | 107259214 | hsa-mir-5195 | - | 1 | 0 | 0 | 0 | 0 |
| chr1 | 22959750 | 22959859 | hsa-mir-6127 | - | 1 | 0 | 0 | 0 | 0 |
| chr20 | 61162118 | 61162220 | hsa-mir-133a-2 | + | 10 | 4 | 7 | 4 | 2 |
| chr12 | 65016288 | 65016385 | hsa-mir-548c | + | 5 | 0 | 1 | 6 | 2 |
| chr10 | 89263637 | 89263711 | hsa-mir-4678 | + | 1 | 0 | 0 | 0 | 0 |
| chr19 | 40788449 | 40788548 | hsa-mir-641 | - | 0 | 0 | 0 | 0 | 0 |
| chr16 | 18496034 | 18496128 | hsa-mir-3180-3 | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 65016288 | 65016385 | hsa-mir-548z | - | 8 | 1 | 2 | 13 | 3 |
| chr22 | 23228446 | 23228559 | hsa-mir-5571 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 8090492 | 8090577 | hsa-mir-3676 | + | 86 | 39 | 49 | 9 | 8 |
| chr16 | 2321747 | 2321841 | hsa-mir-940 | + | 13 | 0 | 4 | 0 | 1 |
| chr7 | 27209098 | 27209182 | hsa-mir-196b | - | 74 | 4 | 9 | 47 | 36 |
| chr3 | 16974687 | 16974752 | hsa-mir-3714 | + | 0 | 0 | 0 | 0 | 0 |
| chr7 | 1062568 | 1062662 | hsa-mir-339 | - | 486 | 222 | 194 | 51 | 98 |
| chr7 | 130562217 | 130562298 | hsa-mir-29b-1 | - | 1743 | 46 | 38 | 875 | 382 |
| chr19 | 39900262 | 39900318 | hsa-mir-4530 | - | 0 | 0 | 0 | 0 | 0 |
| chr10 | 687628 | 687718 | hsa-mir-5699 | - | 2 | 3 | 0 | 0 | 3 |
| chr1 | 23384350 | 23384427 | hsa-mir-4419a | - | 1 | 0 | 0 | 0 | 0 |
| chrX | 139006306 | 139006390 | hsa-mir-505 | - | 92 | 37 | 75 | 31 | 57 |
| chr1 | 54519751 | 54519827 | hsa-mir-4781 | + | 0 | 0 | 0 | 0 | 0 |
| chr20 | 34041775 | 34041919 | hsa-mir-1289-1 | - | 0 | 0 | 0 | 0 | 0 |
| chr11 | 2155363 | 2155439 | hsa-mir-483 | - | 201 | 95 | 53 | 4 | 12 |
| chr9 | 95054739 | 95054829 | hsa-mir-3651 | - | 3 | 1 | 3 | 1 | 1 |
| chr17 | 19247818 | 19247887 | hsa-mir-1180 | - | 36 | 10 | 10 | 0 | 2 |
| chr17 | 17717149 | 17717245 | hsa-mir-33b | - | 111 | 215 | 134 | 10 | 7 |
| chr13 | 100295312 | 100295403 | hsa-mir-4306 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54194134 | 54194219 | hsa-mir-520a | + | 3 | 1 | 0 | 0 | 0 |
| chrX | 49768108 | 49768194 | hsa-mir-188 | + | 35 | 8 | 2 | 3 | 4 |
| chr2 | 11680730 | 11680803 | hsa-mir-4429 | - | 6 | 5 | 3 | 1 | 6 |
| chrX | 53583183 | 53583302 | hsa-mir-98 | - | 268 | 25 | 52 | 210 | 240 |
| chr12 | 79813036 | 79813101 | hsa-mir-1252 | + | 1 | 0 | 0 | 0 | 0 |
| chr19 | 54175221 | 54175294 | hsa-mir-1323 | + | 0 | 0 | 0 | 0 | 0 |
| chr3 | 44903379 | 44903473 | hsa-mir-564 | + | 1 | 0 | 1 | 0 | 0 |
| chr13 | 50623108 | 50623197 | hsa-mir-16-1 | - | 217 | 12 | 19 | 1221 | 91 |
| chr17 | 67095704 | 67095773 | hsa-mir-4524a | - | 14 | 4 | 12 | 1 | 0 |
| chr9 | 96581638 | 96581703 | hsa-mir-4291 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 224444705 | 224444843 | hsa-mir-320b-2 | - | 218 | 214 | 251 | 34 | 71 |
| chr9 | 131154899 | 131154987 | hsa-mir-2964a | + | 6 | 2 | 0 | 2 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 7 | 0 | 1 | 6 | 0 | 2 | 0 | 2 | 1 | 6.72E-07 | 5.69E-05 |
| 2 | 5 | 2 | 2 | 5 | 10 | 3 | 1 | 1 | 0.16649341 | 0.40145979 |
| 385 | 109 | 185 | 229 | 133 | 135 | 29 | 25 | 130 | 0.02417139 | 0.10830643 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 0.68634275 | 0.90275157 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00181102 | 0.01896485 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 6 | 12 | 8 | 13 | 3 | 0 | 2 | 0 | 0.14359115 | 0.35878538 |
| 3 | 2 | 13 | 13 | 12 | 19 | 11 | 10 | 10 | 0.50965236 | 0.75276704 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 2 | 4 | 3 | 5 | 4 | 12 | 2 | 0.00542294 | 0.03828921 |
| 0 | 0 | 1 | 0 | 0 | 2 | 0 | 3 | 0 | 0.3028855 | 0.5346388 |
| 7 | 4 | 23 | 20 | 20 | 34 | 15 | 21 | 14 | 0.66154329 | 0.89746484 |
| 0 | 0 | 1 | 19 | 0 | 18 | 4 | 7 | 0 | 0.00193865 | 0.01985519 |
| 19 | 9 | 40 | 27 | 44 | 48 | 1 | 39 | 6 | 0.09924628 | 0.28286708 |
| 1 | 0 | 12 | 20 | 12 | 35 | 0 | 4 | 3 | 0.75005609 | 0.92101749 |
| 44 | 50 | 239 | 190 | 246 | 330 | 7 | 44 | 117 | 0.69122654 | 0.90275157 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 57 | 75 | 449 | 536 | 659 | 804 | 115 | 90 | 265 | 0.78263557 | 0.94239839 |
| 608 | 440 | 4355 | 4883 | 1250 | 5761 | 2146 | 1652 | 1234 | 0.39319863 | 0.64178831 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 2 | 5 | 9 | 11 | 9 | 6 | 9 | 6 | 0.41644619 | 0.67150147 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 31 | 35 | 171 | 147 | 145 | 259 | 127 | 161 | 36 | 0.97803347 | 1 |
| 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0.45250807 | 0.70097152 |
| 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 16 | 15 | 35 | 5 | 6 | 6 | 1 | 2 | 5 | 2.69E-05 | 0.00108376 |
| 0 | 0 | 2 | 2 | 1 | 2 | 0 | 0 | 0 | 0.1296928 | 0.33576026 |
| 5 | 5 | 17 | 12 | 16 | 18 | 1 | 5 | 3 | 0.10099661 | 0.28400472 |
| 27 | 7 | 133 | 51 | 59 | 246 | 45 | 17 | 54 | 0.27934706 | 0.51554744 |
| 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0.71923557 | 0.90275157 |
| 0 | 0 | 2 | 0 | 0 | 0 | 1 | 1 | 0 | 0.22611245 | 0.5029518 |
| 15 | 5 | 25 | 63 | 31 | 40 | 90 | 39 | 16 | 0.54894436 | 0.78710176 |
| 1 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0.00022938 | 0.00464749 |
| 200 | 233 | 1320 | 1463 | 1048 | 983 | 637 | 526 | 265 | 0.41537456 | 0.67150147 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 0.45250807 | 0.70097152 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.69927681 | 0.90275157 |
| 138 | 473 | 6124 | 4184 | 1137 | 3333 | 1080 | 2765 | 2998 | 0.04182451 | 0.16204427 |
| 3 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0.00134621 | 0.01548974 |
| 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0.71923557 | 0.90275157 |
| 64 | 24 | 48 | 42 | 99 | 40 | 23 | 19 | 12 | 0.00223906 | 0.02173755 |
| 1 | 0 | 2 | 8 | 6 | 8 | 19 | 5 | 2 | 0.50948422 | 0.75276704 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr16 | 2156669 | 2156754 | hsa-mir-6511b-1 | - | 5 | 2 | 3 | 1 | 1 |
| chr4 | 7461754 | 7461845 | hsa-mir-4274 | + | 1 | 0 | 0 | 0 | 0 |
| chr16 | 15737150 | 15737229 | hsa-mir-484 | + | 300 | 21 | 77 | 56 | 26 |
| chr2 | 56227848 | 56227930 | hsa-mir-216b | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 143424140 | 143424215 | hsa-mir-3118-3 | - | 0 | 0 | 0 | 0 | 0 |
| chr2 | 207647957 | 207648032 | hsa-mir-3130-2 | + | 1 | 1 | 2 | 1 | 1 |
| chrX | 146360764 | 146360862 | hsa-mir-514a-1 | - | 2 | 0 | 0 | 1 | 1 |
| chr12 | 54427733 | 54427829 | hsa-mir-615 | + | 50 | 8 | 11 | 0 | 2 |
| chr17 | 35391041 | 35391110 | hsa-mir-2909 | + | 7 | 5 | 2 | 0 | 0 |
| chr19 | 54291958 | 54292027 | hsa-mir-373 | + | 2 | 0 | 0 | 0 | 0 |
| chr3 | 49057580 | 49057667 | hsa-mir-425 | - | 419 | 94 | 195 | 300 | 243 |
| chr3 | 50712510 | 50712594 | hsa-mir-4787 | + | 1 | 0 | 0 | 0 | 0 |
| chr13 | 50570550 | 50570637 | hsa-mir-3613 | - | 29 | 2 | 4 | 60 | 20 |
| chr14 | 31483851 | 31483948 | hsa-mir-624 | - | 3 | 0 | 0 | 4 | 2 |
| chrX | 55477927 | 55478015 | hsa-mir-4536-2 | + | 3 | 0 | 0 | 1 | 0 |
| chr12 | 48526579 | 48526650 | hsa-mir-6505 | + | 2 | 0 | 0 | 0 | 0 |
| chr14 | 101531783 | 101531874 | hsa-mir-412 | + | 6 | 1 | 3 | 0 | 1 |
| chr2 | 136422966 | 136423048 | hsa-mir-128-1 | + | 50 | 7 | 9 | 31 | 20 |
| chr2 | 103048748 | 103048826 | hsa-mir-4772 | + | 1 | 2 | 0 | 0 | 1 |
| chr10 | 56367633 | 56367717 | hsa-mir-548f-1 | - | 1 | 0 | 1 | 4 | 1 |
| chr14 | 74946835 | 74946907 | hsa-mir-4709 | - | 2 | 0 | 2 | 0 | 0 |
| chr19 | 13985688 | 13985825 | hsa-mir-181d | + | 78 | 10 | 11 | 20 | 39 |
| chr16 | 69599710 | 69599771 | hsa-mir-1538 | - | 1 | 0 | 0 | 0 | 0 |
| chr15 | 99327654 | 99327731 | hsa-mir-4714 | + | 1 | 1 | 0 | 0 | 0 |
| chr9 | 96938238 | 96938318 | hsa-let-7a-1 | + | 4645 | 391 | 983 | 5358 | 4656 |
| chr7 | 7106594 | 7106676 | hsa-mir-3683 | - | 0 | 0 | 0 | 0 | 0 |
| chr17 | 11985215 | 11985313 | hsa-mir-744 | + | 437 | 102 | 68 | 21 | 21 |
| chr6 | 52013720 | 52013839 | hsa-mir-133b | + | 1 | 1 | 0 | 0 | 0 |
| chr18 | 37256684 | 37256743 | hsa-mir-5583-1 | + | 1 | 0 | 0 | 0 | 0 |
| chr5 | 179442302 | 179442397 | hsa-mir-340 | - | 5 | 1 | 0 | 59 | 1 |
| chr3 | 49058050 | 49058142 | hsa-mir-191 | - | 5147 | 1151 | 696 | 1208 | 2063 |
| chr9 | 69002238 | 69002321 | hsa-mir-1299 | - | 4 | 22 | 7 | 1 | 17 |
| chr11 | 70130060 | 70130176 | hsa-mir-548k | + | 6 | 0 | 0 | 4 | 2 |
| chr18 | 19263470 | 19263558 | hsa-mir-320c-1 | + | 159 | 191 | 198 | 18 | 62 |
| chr16 | 16462732 | 16462799 | hsa-mir-6511a-3 | + | 2 | 1 | 0 | 0 | 0 |
| chr6 | 132113311 | 132113371 | hsa-mir-548h-5 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54204480 | 54204541 | hsa-mir-520b | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 110141514 | 110141589 | hsa-mir-197 | + | 210 | 81 | 73 | 25 | 61 |
| chr8 | 12584740 | 12584813 | hsa-mir-3926-1 | - | 4 | 2 | 1 | 0 | 0 |
| chrX | 85090784 | 85090863 | hsa-mir-1321 | + | 1 | 0 | 5 | 0 | 0 |
| chr17 | 57918626 | 57918698 | hsa-mir-21 | + | 12169 | 990 | 1178 | 18277 | 19071 |
| chr16 | 15248706 | 15248859 | hsa-mir-3180-4 | - | 0 | 0 | 0 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 2 | 2 | 4 | 5 | 0 | 1 | 1 | 0.24968918 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 25 | 75 | 767 | 421 | 565 | 655 | 38 | 172 | 154 | 0.6271678 | 0.86085477 |
| 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 3 | 1 | 3 | 7 | 1 | 1 | 0 | 1 | 1 |
| 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 0.18973507 | 0.44319069 |
| 1 | 6 | 129 | 63 | 149 | 305 | 1 | 16 | 35 | 0.20040044 | 0.45777748 |
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00044648 | 0.00717452 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 229 | 229 | 1495 | 970 | 2499 | 4349 | 8071 | 1038 | 5266 | 0.0057233 | 0.0401061 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 15 | 47 | 181 | 223 | 213 | 220 | 56 | 89 | 77 | 0.36117722 | 0.60743291 |
| 0 | 0 | 26 | 11 | 3 | 28 | 2 | 8 | 5 | 0.17072659 | 0.40718703 |
| 0 | 0 | 1 | 0 | 3 | 1 | 1 | 0 | 1 | 0.43214339 | 0.69083642 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |
| 1 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.03E-05 | 0.00060263 |
| 21 | 33 | 279 | 295 | 152 | 473 | 65 | 164 | 108 | 0.15574563 | 0.38198664 |
| 1 | 0 | 1 | 1 | 1 | 1 | 9 | 6 | 2 | 0.52579776 | 0.76330765 |
| 0 | 1 | 2 | 7 | 4 | 8 | 7 | 7 | 1 | 0.53946813 | 0.77470615 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0.56462384 | 0.80217899 |
| 22 | 38 | 138 | 64 | 77 | 338 | 25 | 204 | 53 | 0.73440876 | 0.91141008 |
| 0 | 0 | 1 | 0 | 1 | 3 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0.70450643 | 0.90275157 |
| 3651 | 6226 | 33870 | 21963 | 22633 | 21357 | 5103 | 5069 | 5009 | 0.45894773 | 0.70097152 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 39 | 41 | 199 | 306 | 344 | 432 | 43 | 145 | 57 | 0.49443494 | 0.7408575 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.23689045 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 5 | 28 | 954 | 731 | 131 | 916 | 357 | 728 | 503 | 0.00028393 | 0.00494969 |
| 649 | 1242 | 6109 | 1713 | 9241 | 11364 | 9881 | 2481 | 9930 | 0.50478062 | 0.74913303 |
| 9 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5.83E-09 | 1.09E-06 |
| 1 | 2 | 6 | 20 | 10 | 17 | 12 | 18 | 13 | 0.27224233 | 0.51361322 |
| 41 | 16 | 38 | 22 | 92 | 30 | 15 | 6 | 14 | 0.00242108 | 0.02302498 |
| 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0.45613637 | 0.70097152 |
| 0 | 0 | 2 | 3 | 2 | 3 | 1 | 2 | 2 | 0.02188 | 0.10019112 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 64 | 35 | 340 | 414 | 500 | 570 | 97 | 261 | 87 | 0.66430531 | 0.89990196 |
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0.02148756 | 0.09963387 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00260922 | 0.02360966 |
| 10632 | 15740 | 495500 | 343646 | 410418 | 428303 | 307645 | 309635 | 710432 | 0.00297229 | 0.02588947 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0.71923557 | 0.90275157 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr17 | 15154943 | 15155013 | hsa-mir-4731 | - | 2 | 0 | 1 | 0 | 0 |
| chr5 | 148808480 | 148808586 | hsa-mir-143 | + | 44495 | 8261 | 8195 | 10777 | 4056 |
| chrX | 151127049 | 151127130 | hsa-mir-224 | - | 666 | 9 | 36 | 173 | 120 |
| chr2 | 25551508 | 25551590 | hsa-mir-1301 | - | 35 | 10 | 9 | 1 | 8 |
| chr3 | 48357849 | 48357949 | hsa-mir-2115 | - | 27 | 15 | 0 | 0 | 0 |
| chrX | 146331668 | 146331748 | hsa-mir-514b | - | 0 | 0 | 0 | 0 | 0 |
| chr14 | 106325652 | 106325722 | hsa-mir-4537 | - | 0 | 0 | 0 | 0 | 0 |
| chr5 | 57825869 | 57825936 | hsa-mir-548ae-2 | - | 2 | 0 | 0 | 1 | 1 |
| chr2 | 189162218 | 189162315 | hsa-mir-561 | + | 2 | 0 | 0 | 0 | 1 |
| chrX | 45605584 | 45605694 | hsa-mir-221 | - | 3816 | 408 | 472 | 1016 | 1427 |
| chr21 | 24451605 | 24451714 | hsa-mir-6130 | + | 1 | 0 | 0 | 0 | 1 |
| chr16 | 26036557 | 26036631 | hsa-mir-548w | + | 4 | 1 | 0 | 1 | 3 |
| chr1 | 220373879 | 220373961 | hsa-mir-664a | - | 56 | 36 | 45 | 18 | 17 |
| chrX | 73438381 | 73438453 | hsa-mir-374b | - | 33 | 9 | 3 | 305 | 98 |
| chr1 | 85599476 | 85599556 | hsa-mir-4423 | + | 1 | 0 | 1 | 0 | 0 |
| chr13 | 92003318 | 92003389 | hsa-mir-20a | + | 100 | 16 | 5 | 1264 | 217 |
| chr7 | 43190493 | 43190593 | hsa-mir-3943 | + | 1 | 0 | 0 | 0 | 0 |
| chr16 | 2324620 | 2324692 | hsa-mir-4717 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 149112387 | 149112453 | hsa-mir-378a | + | 8867 | 4096 | 7011 | 1053 | 2691 |
| chr17 | 27717679 | 27717748 | hsa-mir-4523 | + | 1 | 1 | 0 | 0 | 0 |
| chr3 | 185485634 | 185485692 | hsa-mir-548aq | - | 2 | 1 | 0 | 2 | 2 |
| chr4 | 87463634 | 87463705 | hsa-mir-4452 | - | 1 | 0 | 0 | 0 | 1 |
| chr3 | 160122375 | 160122473 | hsa-mir-15b | + | 178 | 14 | 52 | 634 | 56 |
| chr19 | 50004041 | 50004125 | hsa-mir-150 | - | 530 | 328 | 301 | 200 | 232 |
| chr8 | 9760897 | 9760982 | hsa-mir-124-1 | - | 17 | 0 | 0 | 4 | 1 |
| chr11 | 75046135 | 75046230 | hsa-mir-326 | - | 106 | 45 | 71 | 10 | 18 |
| chr11 | 111218481 | 111218549 | hsa-mir-4491 | + | 1 | 0 | 0 | 0 | 0 |
| chr14 | 101340829 | 101340922 | hsa-mir-337 | + | 13 | 5 | 1 | 25 | 30 |
| chr22 | 46486923 | 46487006 | hsa-mir-3619 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 207975787 | 207975868 | hsa-mir-29b-2 | - | 1814 | 56 | 57 | 913 | 403 |
| chr17 | 37882747 | 37882814 | hsa-mir-4728 | + | 0 | 0 | 0 | 0 | 0 |
| chr11 | 10529816 | 10529873 | hsa-mir-4485 | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 154166140 | 154166219 | hsa-mir-190b | - | 41 | 1 | 20 | 9 | 7 |
| chr14 | 101506026 | 101506092 | hsa-mir-376c | + | 1564 | 1580 | 112 | 248 | 900 |
| chr18 | 19408964 | 19409049 | hsa-mir-1-2 | - | 6 | 0 | 6 | 46 | 24 |
| chrX | 133674370 | 133674461 | hsa-mir-450a-1 | - | 36 | 4 | 4 | 19 | 24 |
| chr5 | 54466473 | 54466570 | hsa-mir-449b | - | 0 | 0 | 0 | 0 | 0 |
| chr18 | 21901649 | 21901699 | hsa-mir-320c-2 | + | 44 | 64 | 59 | 7 | 23 |
| chr5 | 85916313 | 85916392 | hsa-mir-3607 | + | 100 | 166 | 42 | 3 | 41 |
| chr8 | 41517958 | 41518026 | hsa-mir-486 | - | 800 | 158 | 1223 | 2070 | 166 |
| chr2 | 208133998 | 208134148 | hsa-mir-1302-4 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 71591120 | 71591240 | hsa-mir-1284 | - | 0 | 0 | 0 | 0 | 0 |
| chr16 | 11400296 | 11400384 | hsa-mir-548h-2 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 135633036 | 135633119 | hsa-mir-934 | + | 48 | 5 | 1 | 12 | 8 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0141554 | 0.07411706 |
| 13171 | 27859 | 38906 | 29920 | 31599 | 32008 | 14654 | 14160 | 15856 | 0.1321433 | 0.34021425 |
| 51 | 76 | 430 | 4300 | 1495 | 159 | 9 | 145 | 34 | 0.43568604 | 0.69411861 |
| 7 | 6 | 35 | 107 | 117 | 231 | 36 | 63 | 10 | 0.20373828 | 0.46247833 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3.35E-08 | 4.47E-06 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 7 | 0 | 1 | 2 | 0 | 0 | 0.07464642 | 0.25390681 |
| 2 | 1 | 12 | 5 | 4 | 6 | 6 | 3 | 3 | 0.42202495 | 0.67815043 |
| 0 | 2 | 9 | 3 | 1 | 2 | 2 | 1 | 0 | 1 | 1 |
| 2394 | 914 | 2839 | 2428 | 3224 | 3201 | 1121 | 1156 | 1028 | 0.17661571 | 0.41879956 |
| 3 | 6 | 3 | 2 | 1 | 4 | 9 | 2 | 2 | 0.67817645 | 0.90275157 |
| 3 | 4 | 5 | 11 | 11 | 8 | 3 | 4 | 4 | 0.94785414 | 1 |
| 24 | 32 | 169 | 186 | 99 | 210 | 46 | 25 | 56 | 0.83601652 | 0.98753789 |
| 76 | 175 | 1113 | 974 | 294 | 1414 | 4004 | 918 | 467 | 0.00966283 | 0.05810166 |
| 3 | 0 | 1 | 3 | 1 | 4 | 2 | 0 | 0 | 0.84990024 | 0.9976159 |
| 83 | 851 | 7913 | 1988 | 921 | 1062 | 699 | 503 | 1070 | 0.31607398 | 0.54856788 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 2 | 0 | 2 | 2 | 0 | 0 | 0 | 0.3028855 | 0.5346388 |
| 1024 | 2209 | 752 | 796 | 1065 | 1969 | 720 | 943 | 571 | 1.67E-06 | 0.00012973 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 1 | 0 | 6 | 4 | 4 | 13 | 6 | 7 | 1 | 0.45586615 | 0.70097152 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0.68904704 | 0.90275157 |
| 77 | 236 | 2497 | 1587 | 448 | 1303 | 415 | 2403 | 1427 | 0.09177009 | 0.2762803 |
| 116 | 167 | 996 | 2385 | 239 | 388 | 401 | 260 | 342 | 0.94517997 | 1 |
| 0 | 0 | 6 | 0 | 2 | 1 | 0 | 0 | 1 | 0.00211681 | 0.02098795 |
| 23 | 13 | 127 | 71 | 55 | 187 | 31 | 13 | 15 | 0.43839659 | 0.69611088 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 42 | 23 | 36 | 25 | 20 | 18 | 9 | 4 | 16 | 0.0822486 | 0.27056701 |
| 0 | 0 | 1 | 0 | 3 | 2 | 0 | 2 | 0 | 0.09986078 | 0.28307267 |
| 639 | 448 | 4625 | 5129 | 1345 | 6117 | 2251 | 1663 | 1273 | 0.38800422 | 0.6368218 |
| 0 | 0 | 1 | 27 | 60 | 37 | 0 | 0 | 2 | 0.00010059 | 0.00260416 |
| 0 | 0 | 1 | 1 | 1 | 2 | 0 | 0 | 1 | 0.3028855 | 0.5346388 |
| 5 | 10 | 1316 | 128 | 23 | 1242 | 862 | 3 | 255 | 0.00049045 | 0.00774742 |
| 1891 | 854 | 365 | 314 | 349 | 253 | 89 | 117 | 390 | 5.32E-05 | 0.00179823 |
| 17 | 33 | 92 | 49 | 46 | 26 | 19 | 19 | 5 | 0.52960541 | 0.76763957 |
| 41 | 45 | 136 | 43 | 50 | 74 | 88 | 63 | 45 | 0.98005501 | 1 |
| 0 | 0 | 4 | 11 | 5 | 8 | 1 | 1 | 19 | 0.00170712 | 0.01828775 |
| 12 | 7 | 15 | 12 | 41 | 12 | 7 | 2 | 6 | 0.00929175 | 0.05660073 |
| 14 | 3 | 1 | 3 | 1 | 5 | 8 | 1 | 2 | 3.58E-07 | 3.33E-05 |
| 82 | 197 | 253 | 90 | 78 | 52 | 5 | 13 | 20 | 2.48E-07 | 2.57E-05 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0.71923557 | 0.90275157 |
| 0 | 0 | 3 | 2 | 1 | 3 | 0 | 0 | 5 | 0.03326823 | 0.13480867 |
| 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 1 | 0.07026364 | 0.24526482 |
| 4 | 20 | 18 | 5 | 20 | 19 | 2 | 1248 | 2 | 0.02046083 | 0.09679945 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr9 | 127455988 | 127456077 | hsa-mir-181b-2 | + | 159 | 42 | 38 | 46 | 150 |
| chrX | 6301946 | 6302004 | hsa-mir-4770 | - | 2 | 0 | 0 | 0 | 0 |
| chr10 | 18134035 | 18134122 | hsa-mir-511-2 | + | 50 | 8 | 4 | 9 | 7 |
| chr19 | 14640354 | 14640452 | hsa-mir-639 | + | 1 | 0 | 1 | 0 | 0 |
| chr2 | 233415183 | 233415283 | hsa-mir-5001 | - | 3 | 0 | 1 | 1 | 0 |
| chr16 | 70064248 | 70064325 | hsa-mir-1972-2 | + | 1 | 1 | 1 | 0 | 0 |
| chrX | 146312237 | 146312361 | hsa-mir-506 | - | 0 | 0 | 0 | 0 | 0 |
| chr17 | 65467604 | 65467701 | hsa-mir-548aa-2 | + | 2 | 0 | 1 | 3 | 1 |
| chr22 | 46156403 | 46156478 | hsa-mir-4762 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 198828172 | 198828282 | hsa-mir-181a-1 | - | 1817 | 787 | 721 | 395 | 1336 |
| chr8 | 125834219 | 125834300 | hsa-mir-4662b | - | 2 | 0 | 0 | 1 | 1 |
| chr19 | 54169932 | 54170016 | hsa-mir-512-1 | + | 5 | 0 | 0 | 1 | 1 |
| chr9 | 95290265 | 95290340 | hsa-mir-4670 | - | 2 | 0 | 2 | 0 | 0 |
| chr8 | 145625475 | 145625559 | hsa-mir-1234 | - | 2 | 0 | 0 | 0 | 0 |
| chr17 | 27188388 | 27188456 | hsa-mir-451b | + | 13127 | 781 | 1856 | 163887 | 5476 |
| chr20 | 37145205 | 37145275 | hsa-mir-548o-2 | + | 3 | 1 | 1 | 5 | 1 |
| chr17 | 40666205 | 40666325 | hsa-mir-5010 | + | 9 | 1 | 2 | 2 | 2 |
| chr11 | 122017229 | 122017301 | hsa-let-7a-2 | - | 4626 | 353 | 939 | 5327 | 4636 |
| chr14 | 101493436 | 101493520 | hsa-mir-329-2 | + | 22 | 1 | 3 | 2 | 7 |
| chr1 | 178646883 | 178646969 | hsa-mir-4424 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 47365707 | 47365816 | hsa-mir-6129 | - | 2 | 0 | 0 | 1 | 1 |
| chr2 | 33643582 | 33643631 | hsa-mir-4430 | + | 2 | 0 | 0 | 0 | 0 |
| chr14 | 101521023 | 101521096 | hsa-mir-134 | + | 306 | 152 | 147 | 15 | 79 |
| chr5 | 71465293 | 71465367 | hsa-mir-4803 | + | 0 | 0 | 0 | 0 | 0 |
| chr16 | 85775226 | 85775306 | hsa-mir-1910 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 184970997 | 184971060 | hsa-mir-5588 | - | 1 | 0 | 0 | 2 | 2 |
| chr6 | 30552108 | 30552194 | hsa-mir-877 | + | 8 | 1 | 3 | 1 | 2 |
| chr9 | 125873824 | 125873922 | hsa-mir-600 | - | 0 | 0 | 0 | 0 | 0 |
| chr15 | 89151337 | 89151428 | hsa-mir-1179 | + | 6 | 1 | 0 | 1 | 3 |
| chr2 | 56216084 | 56216194 | hsa-mir-216a | - | 2 | 1 | 1 | 1 | 3 |
| chr8 | 1765396 | 1765473 | hsa-mir-596 | + | 0 | 0 | 0 | 0 | 0 |
| chr2 | 35696470 | 35696552 | hsa-mir-548ad | + | 1 | 0 | 1 | 6 | 2 |
| chr10 | 53059332 | 53059415 | hsa-mir-605 | + | 1 | 0 | 0 | 0 | 0 |
| chr17 | 29887014 | 29887102 | hsa-mir-193a | + | 2754 | 779 | 1825 | 379 | 338 |
| chr17 | 29902287 | 29902377 | hsa-mir-4725 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 6416420 | 6416522 | hsa-mir-3940 | - | 2 | 0 | 1 | 0 | 0 |
| chr2 | 207974710 | 207974797 | hsa-mir-2355 | - | 17 | 0 | 1 | 3 | 4 |
| chr22 | 31127543 | 31127628 | hsa-mir-3200 | + | 2 | 1 | 0 | 3 | 0 |
| chr14 | 95604255 | 95604323 | hsa-mir-3173 | - | 4 | 2 | 0 | 0 | 0 |
| chr13 | 92003445 | 92003532 | hsa-mir-19b-1 | + | 7383 | 2173 | 530 | 1557 | 2244 |
| chr2 | 112528637 | 112528711 | hsa-mir-4771-2 | - | 0 | 0 | 0 | 0 | 0 |
| chr17 | 57228496 | 57228582 | hsa-mir-301a | - | 52 | 0 | 2 | 84 | 25 |
| chr13 | 115009979 | 115010036 | hsa-mir-548ar | + | 2 | 0 | 0 | 1 | 1 |
| chr15 | 64163128 | 64163218 | hsa-mir-422a | - | 53 | 22 | 43 | 1 | 5 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 70 | 114 | 411 | 195 | 446 | 406 | 89 | 874 | 443 | 0.46259348 | 0.70332321 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |
| 13 | 5 | 26 | 16 | 11 | 13 | 3 | 14 | 6 | 0.0559964 | 0.20228156 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 1 | 4 | 5 | 1 | 9 | 3 | 2 | 0 | 0.7047163 | 0.90275157 |
| 1 | 0 | 2 | 3 | 2 | 6 | 2 | 2 | 1 | 0.53899964 | 0.77470615 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 1 | 2 | 15 | 12 | 7 | 13 | 12 | 25 | 12 | 0.09642618 | 0.27737408 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 676 | 803 | 3158 | 1459 | 2953 | 2036 | 462 | 2653 | 2653 | 0.75981099 | 0.92688984 |
| 0 | 0 | 12 | 4 | 1 | 93 | 2 | 2 | 1 | 0.01705688 | 0.08455859 |
| 2 | 1 | 4 | 0 | 2 | 0 | 1 | 2 | 0 | 0.1022153 | 0.28437213 |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00289228 | 0.0256724 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 3409 | 8109 | 6793 | 2032 | 2083 | 1434 | 2664 | 566 | 806 | 7.01E-18 | 3.27E-15 |
| 3 | 1 | 22 | 10 | 9 | 22 | 14 | 16 | 9 | 0.23130614 | 0.50963907 |
| 1 | 0 | 3 | 0 | 3 | 5 | 1 | 0 | 2 | 0.05378474 | 0.19657794 |
| 3625 | 6232 | 33644 | 21905 | 22588 | 21226 | 5100 | 5045 | 4927 | 0.45651308 | 0.70097152 |
| 13 | 4 | 11 | 6 | 6 | 11 | 1 | 2 | 7 | 0.04232522 | 0.16204427 |
| 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0.68634274 | 0.90275157 |
| 0 | 2 | 1 | 0 | 1 | 1 | 1 | 2 | 0 | 0.15381403 | 0.37824452 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00513831 | 0.0368377 |
| 106 | 43 | 24 | 41 | 70 | 32 | 25 | 18 | 37 | 0.00100446 | 0.01245739 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0.71923557 | 0.90275157 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00398986 | 0.03124828 |
| 2 | 1 | 4 | 7 | 9 | 19 | 9 | 28 | 4 | 0.53867996 | 0.77470615 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 2 | 0 | 1 | 0 | 2 | 0 | 0 | 0.01224127 | 0.06956623 |
| 1 | 1 | 6 | 1 | 2 | 0 | 0 | 1 | 0 | 0.14279757 | 0.35872596 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 2 | 7 | 5 | 3 | 11 | 4 | 4 | 6 | 0.94394499 | 1 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 529 | 890 | 2435 | 1267 | 1240 | 2527 | 838 | 1355 | 897 | 0.22343615 | 0.49818777 |
| 0 | 0 | 1 | 2 | 4 | 5 | 0 | 1 | 0 | 0.0489372 | 0.18317057 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0141554 | 0.07411706 |
| 2 | 1 | 15 | 20 | 27 | 24 | 28 | 27 | 4 | 0.43310351 | 0.69118574 |
| 3 | 1 | 9 | 28 | 35 | 57 | 12 | 21 | 6 | 0.02347959 | 0.10571486 |
| 1 | 0 | 1 | 4 | 2 | 4 | 0 | 1 | 0 | 0.38759415 | 0.6368218 |
| 1646 | 1677 | 8262 | 5344 | 2344 | 2105 | 5066 | 1470 | 2691 | 0.24645472 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 26 | 39 | 516 | 221 | 177 | 250 | 157 | 609 | 265 | 0.08296394 | 0.27130663 |
| 0 | 0 | 1 | 2 | 1 | 4 | 1 | 2 | 0 | 1 | 1 |
| 4 | 5 | 1 | 6 | 5 | 11 | 0 | 2 | 0 | 0.00010367 | 0.00261146 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr17 | 26687676 | 26687757 | hsa-mir-4723 | + | 1 | 0 | 1 | 0 | 0 |
| chr14 | 101488402 | 101488469 | hsa-mir-379 | + | 156 | 13 | 50 | 27 | 127 |
| chr7 | 73605527 | 73605624 | hsa-mir-590 | + | 28 | 2 | 0 | 41 | 14 |
| chr9 | 97848295 | 97848376 | hsa-mir-3074 | - | 17270 | 4678 | 4271 | 1979 | 4685 |
| chr2 | 189842818 | 189842887 | hsa-mir-1245b | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 100583661 | 100583727 | hsa-mir-1827 | + | 3 | 0 | 0 | 1 | 1 |
| chr19 | 4445974 | 4446045 | hsa-mir-4746 | + | 1 | 0 | 0 | 0 | 0 |
| chrX | 85158640 | 85158712 | hsa-mir-361 | - | 280 | 96 | 111 | 136 | 159 |
| chr10 | 12172758 | 12172815 | hsa-mir-548ak | - | 1 | 1 | 1 | 3 | 1 |
| chr19 | 1816157 | 1816237 | hsa-mir-1909 | - | 1 | 0 | 0 | 0 | 0 |
| chr14 | 100774195 | 100774293 | hsa-mir-345 | + | 225 | 66 | 40 | 73 | 78 |
| chr6 | 32137806 | 32137893 | hsa-mir-6721 | - | 1 | 0 | 0 | 0 | 0 |
| chr19 | 54188262 | 54188345 | hsa-mir-515-2 | + | 3 | 0 | 1 | 1 | 1 |
| chr15 | 79502129 | 79502213 | hsa-mir-184 | + | 2 | 1 | 3 | 0 | 2 |
| chr15 | 21145580 | 21145662 | hsa-mir-5701-1 | + | 1 | 3 | 0 | 0 | 1 |
| chr2 | 241395417 | 241395506 | hsa-mir-149 | + | 64 | 18 | 21 | 12 | 15 |
| chr3 | 10371912 | 10371969 | hsa-mir-378b | + | 21 | 15 | 23 | 3 | 3 |
| chr17 | 1953201 | 1953302 | hsa-mir-132 | - | 122 | 21 | 11 | 27 | 186 |
| chr19 | 54228695 | 54228780 | hsa-mir-516b-2 | + | 1 | 0 | 0 | 0 | 0 |
| chr2 | 177015030 | 177015140 | hsa-mir-10b | + | 3050 | 172 | 290 | 2240 | 2492 |
| chr17 | 35974975 | 35975084 | hsa-mir-378j | - | 10 | 2 | 4 | 0 | 1 |
| chr19 | 54209505 | 54209590 | hsa-mir-526a-1 | + | 0 | 1 | 0 | 0 | 0 |
| chr3 | 188406568 | 188406654 | hsa-mir-28 | + | 1096 | 215 | 228 | 323 | 635 |
| chr4 | 61788336 | 61788402 | hsa-mir-548ag-1 | + | 0 | 0 | 0 | 0 | 0 |
| chr11 | 62560173 | 62560243 | hsa-mir-6514 | - | 0 | 0 | 0 | 0 | 0 |
| chr6 | 72113253 | 72113324 | hsa-mir-30a | - | 5210 | 788 | 536 | 2074 | 4187 |
| chr1 | 161196975 | 161197051 | hsa-mir-5187 | + | 1 | 3 | 4 | 0 | 1 |
| chr21 | 26946291 | 26946356 | hsa-mir-155 | + | 13 | 2 | 3 | 44 | 17 |
| chr8 | 14710946 | 14711019 | hsa-mir-383 | - | 10 | 37 | 0 | 1 | 6 |
| chr19 | 52196506 | 52196592 | hsa-mir-125a | + | 5842 | 2652 | 4042 | 1163 | 1677 |
| chr1 | 171070868 | 171070947 | hsa-mir-1295a | - | 1 | 2 | 0 | 1 | 1 |
| chr11 | 18409333 | 18409407 | hsa-mir-3159 | + | 4 | 5 | 1 | 0 | 2 |
| chr22 | 41209886 | 41209962 | hsa-mir-4766 | - | 2 | 0 | 0 | 0 | 1 |
| chr19 | 54254464 | 54254551 | hsa-mir-522 | + | 1 | 2 | 0 | 1 | 1 |
| chr2 | 110844009 | 110844100 | hsa-mir-4436b-1 | - | 0 | 0 | 0 | 0 | 0 |
| chr4 | 13629488 | 13629581 | hsa-mir-5091 | + | 0 | 0 | 0 | 0 | 0 |
| chr12 | 66644861 | 66644937 | hsa-mir-6502 | + | 2 | 0 | 0 | 0 | 1 |
| chr22 | 46509565 | 46509648 | hsa-let-7b | + | 28839 | 8701 | 20323 | 5542 | 9399 |
| chr14 | 101521594 | 101521660 | hsa-mir-668 | + | 1 | 0 | 0 | 0 | 0 |
| chrX | 137749871 | 137749954 | hsa-mir-504 | - | 47 | 10 | 16 | 7 | 13 |
| chr6 | 107231999 | 107232095 | hsa-mir-587 | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 100191048 | 100191117 | hsa-mir-4685 | - | 3 | 0 | 0 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0.30692987 | 0.53593028 |
| 55 | 118 | 112 | 100 | 145 | 80 | 22 | 23 | 73 | 0.09992587 | 0.28307267 |
| 13 | 28 | 243 | 239 | 58 | 155 | 112 | 120 | 331 | 0.08244746 | 0.27056701 |
| 2702 | 3894 | 18180 | 17419 | 28589 | 38150 | 7392 | 12558 | 8786 | 0.83921735 | 0.98881235 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 0.28679032 | 0.52204801 |
| 5 | 0 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 0.12263158 | 0.32014743 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 90 | 147 | 2046 | 1124 | 484 | 1160 | 474 | 556 | 524 | 0.25938293 | 0.51361322 |
| 1 | 3 | 5 | 8 | 3 | 6 | 2 | 7 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 69 | 69 | 177 | 433 | 686 | 1232 | 1042 | 334 | 228 | 0.19030035 | 0.44339981 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 4 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 0 | 0.0277968 | 0.11775735 |
| 1 | 3 | 138 | 385 | 213 | 23795 | 10 | 56 | 13 | 3.07E-12 | 9.53E-10 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0087498 | 0.05436539 |
| 11 | 15 | 859 | 183 | 405 | 538 | 80 | 77 | 216 | 0.03717003 | 0.14730381 |
| 3 | 18 | 2 | 0 | 3 | 7 | 5 | 3 | 1 | 0.00067211 | 0.00958327 |
| 124 | 208 | 201 | 135 | 103 | 150 | 31 | 48 | 140 | 0.14787942 | 0.36557988 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1607 | 3130 | 4084 | 2280 | 1671 | 1476 | 2333 | 941 | 2796 | 0.10147379 | 0.28400472 |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6.71E-06 | 0.0004466 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 467 | 606 | 2052 | 1348 | 1457 | 2005 | 839 | 2199 | 1269 | 0.89898923 | 1 |
| 0 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0.68634275 | 0.90275157 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2200 | 5631 | 31834 | 5115 | 3075 | 2298 | 10546 | 5129 | 1749 | 0.97334299 | 1 |
| 0 | 0 | 1 | 3 | 4 | 1 | 0 | 0 | 1 | 0.27098826 | 0.51361322 |
| 9 | 57 | 410 | 826 | 332 | 393 | 1080 | 463 | 366 | 0.00119746 | 0.01412704 |
| 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.86E-08 | 4.44E-06 |
| 1113 | 1220 | 9373 | 4318 | 4808 | 5501 | 160 | 2623 | 1537 | 0.26479955 | 0.51361322 |
| 1 | 1 | 5 | 3 | 4 | 6 | 1 | 1 | 2 | 1 | 1 |
| 0 | 1 | 10 | 6 | 7 | 11 | 4 | 9 | 3 | 0.65533621 | 0.89085875 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |
| 2 | 0 | 1 | 1 | 15 | 4 | 1 | 107 | 1 | 0.01038029 | 0.06157765 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00490233 | 0.03605237 |
| 6055 | 6114 | 53328 | 32675 | 57678 | 15657 | 3232 | 4106 | 4477 | 0.37547054 | 0.62266644 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 15 | 8 | 91 | 15 | 44 | 90 | 77 | 28 | 5 | 0.99230885 | 1 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr1 | 51525508 | 51525577 | hsa-mir-4421 | + | 6 | 0 | 0 | 0 | 0 |
| chr4 | 5925001 | 5925055 | hsa-mir-378d-1 | - | 166 | 130 | 161 | 29 | 41 |
| chrX | 133674214 | 133674292 | hsa-mir-450b | - | 36 | 1 | 2 | 7 | 10 |
| chr10 | 104196268 | 104196341 | hsa-mir-146b | + | 896 | 409 | 250 | 204 | 993 |
| chr11 | 72326106 | 72326174 | hsa-mir-139 | - | 329 | 38 | 136 | 110 | 65 |
| chr19 | 54216600 | 54216688 | hsa-mir-519d | + | 2 | 0 | 0 | 1 | 0 |
| chr1 | 94312387 | 94312467 | hsa-mir-760 | + | 5 | 3 | 3 | 0 | 0 |
| chr1 | 71533313 | 71533399 | hsa-mir-186 | - | 1962 | 405 | 212 | 389 | 1251 |
| chr1 | 10287775 | 10287861 | hsa-mir-1273d | + | 3 | 0 | 1 | 0 | 0 |
| chr19 | 54223349 | 54223436 | hsa-mir-520d | + | 1 | 0 | 1 | 0 | 1 |
| chr16 | 88535325 | 88535439 | hsa-mir-5189 | + | 1 | 0 | 0 | 0 | 0 |
| chr5 | 128732754 | 128732840 | hsa-mir-4460 | - | 3 | 0 | 1 | 1 | 0 |
| chr4 | 160049953 | 160050046 | hsa-mir-3688-1 | - | 2 | 1 | 0 | 1 | 0 |
| chr10 | 35930099 | 35930180 | hsa-mir-4683 | - | 1 | 1 | 1 | 0 | 0 |
| chr5 | 87962670 | 87962757 | hsa-mir-9-2 | - | 34 | 1 | 0 | 16 | 34 |
| chr9 | 131006999 | 131007109 | hsa-mir-199b | - | 3426 | 215 | 1020 | 3538 | 2943 |
| chr11 | 101390550 | 101390636 | hsa-mir-3920 | - | 0 | 0 | 0 | 0 | 0 |
| chr5 | 159912358 | 159912457 | hsa-mir-146a | + | 1343 | 1070 | 1303 | 377 | 854 |
| chr3 | 15738804 | 15738878 | hsa-mir-3134 | - | 0 | 0 | 0 | 0 | 0 |
| chr20 | 50069441 | 50069514 | hsa-mir-3194 | - | 3 | 2 | 0 | 0 | 0 |
| chr19 | 46522189 | 46522307 | hsa-mir-769 | + | 330 | 146 | 110 | 20 | 77 |
| chr7 | 32772592 | 32772689 | hsa-mir-550b-2 | - | 6 | 5 | 4 | 2 | 4 |
| chr2 | 219267368 | 219267445 | hsa-mir-26b | + | 1587 | 90 | 70 | 2640 | 1141 |
| chr3 | 113313722 | 113313789 | hsa-mir-4446 | + | 1 | 0 | 0 | 0 | 0 |
| chr12 | 7072861 | 7072929 | hsa-mir-200c | + | 581 | 33 | 55 | 692 | 1543 |
| chr3 | 35785967 | 35786051 | hsa-mir-128-2 | + | 32 | 4 | 8 | 24 | 15 |
| chr14 | 101512256 | 101512331 | hsa-mir-381 | + | 29 | 2 | 1 | 37 | 33 |
| chr12 | 128729050 | 128729118 | hsa-mir-4419b | + | 2 | 0 | 1 | 0 | 0 |
| chr16 | 67236223 | 67236298 | hsa-mir-328 | - | 268 | 183 | 210 | 25 | 36 |
| chrX | 105883043 | 105883126 | hsa-mir-548an | + | 1 | 0 | 0 | 0 | 1 |
| chr22 | 42319225 | 42319301 | hsa-mir-378i | - | 263 | 196 | 157 | 28 | 104 |
| chr3 | 52328234 | 52328324 | hsa-mir-135a-1 | - | 4 | 0 | 1 | 3 | 2 |
| chrX | 49773571 | 49773636 | hsa-mir-362 | + | 194 | 88 | 89 | 43 | 61 |
| chr11 | 96074601 | 96074690 | hsa-mir-1260b | + | 461 | 153 | 207 | 76 | 106 |
| chr15 | 66789295 | 66789372 | hsa-mir-4512 | - | 204 | 37 | 123 | 8 | 35 |
| chr7 | 93113247 | 93113331 | hsa-mir-489 | - | 11 | 0 | 0 | 2 | 0 |
| chr9 | 36864250 | 36864305 | hsa-mir-4540 | - | 0 | 0 | 0 | 0 | 0 |
| chr2 | 207647957 | 207648032 | hsa-mir-3130-1 | - | 1 | 1 | 2 | 1 | 1 |
| chr8 | 135817118 | 135817188 | hsa-mir-30d | - | 2835 | 388 | 425 | 1206 | 1342 |
| chr10 | 124176480 | 124176583 | hsa-mir-3941 | + | 1 | 0 | 0 | 0 | 1 |
| chr16 | 18437869 | 18437936 | hsa-mir-6511a-4 | - | 2 | 1 | 0 | 0 | 0 |
| chr9 | 97847726 | 97847823 | hsa-mir-27b | + | 3346 | 379 | 621 | 1495 | 1225 |
| chrX | 20035205 | 20035305 | hsa-mir-23c | - | 8 | 2 | 0 | 6 | 6 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00095745 | 0.01205865 |
| 23 | 36 | 13 | 15 | 15 | 36 | 15 | 7 | 5 | 0.00014083 | 0.00329431 |
| 23 | 32 | 50 | 26 | 33 | 55 | 45 | 104 | 108 | 0.68002703 | 0.90275157 |
| 771 | 1198 | 1303 | 1686 | 758 | 1349 | 1692 | 1624 | 1909 | 0.70740329 | 0.90275157 |
| 38 | 165 | 214 | 78 | 53 | 46 | 2 | 17 | 17 | 0.00672966 | 0.04512258 |
| 3 | 0 | 1 | 0 | 0 | 1 | 0 | 7 | 1 | 0.45922556 | 0.70097152 |
| 1 | 56 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3.06E-09 | 7.13E-07 |
| 542 | 582 | 2334 | 1837 | 1175 | 2887 | 2274 | 1512 | 1152 | 0.77242468 | 0.93372218 |
| 1 | 1 | 5 | 6 | 2 | 4 | 3 | 2 | 3 | 0.72178144 | 0.90295342 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00154916 | 0.01684238 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00321004 | 0.02744727 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00491272 | 0.03605237 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02636789 | 0.11377256 |
| 79 | 37 | 43 | 18 | 39 | 264 | 27 | 10 | 22 | 0.5119816 | 0.75501085 |
| 5008 | 3224 | 14929 | 10338 | 11871 | 11192 | 2327 | 2931 | 4531 | 0.99493753 | 1 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 578 | 935 | 1103 | 2105 | 956 | 1449 | 2273 | 587 | 604 | 0.26508899 | 0.51361322 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0.71923557 | 0.90275157 |
| 0 | 0 | 1 | 0 | 0 | 1 | 2 | 3 | 4 | 0.85688826 | 1 |
| 50 | 56 | 392 | 204 | 166 | 280 | 693 | 393 | 96 | 0.94021726 | 1 |
| 2 | 2 | 9 | 5 | 8 | 12 | 4 | 7 | 3 | 0.58836073 | 0.82088653 |
| 692 | 3245 | 16763 | 7508 | 4051 | 6152 | 5104 | 2815 | 3596 | 0.37685158 | 0.62384667 |
| 0 | 0 | 7 | 9 | 7 | 11 | 5 | 0 | 0 | 0.01834782 | 0.08961841 |
| 586 | 1289 | 9198 | 7884 | 5298 | 9616 | 4186 | 6940 | 5262 | 0.03434349 | 0.13856336 |
| 13 | 20 | 191 | 210 | 116 | 322 | 47 | 124 | 67 | 0.14198853 | 0.3576576 |
| 78 | 77 | 196 | 95 | 118 | 103 | 15 | 21 | 91 | 0.75111055 | 0.92109872 |
| 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0.45805662 | 0.70097152 |
| 32 | 15 | 135 | 94 | 205 | 120 | 19 | 5 | 2 | 0.06034705 | 0.21632095 |
| 1 | 1 | 2 | 0 | 2 | 3 | 0 | 0 | 7 | 0.85769997 | 1 |
| 33 | 67 | 23 | 18 | 30 | 86 | 24 | 20 | 21 | 0.00068373 | 0.00958327 |
| 4 | 7 | 25 | 3 | 0 | 1 | 2 | 6 | 3 | 0.46760359 | 0.70862854 |
| 31 | 34 | 182 | 161 | 168 | 199 | 545 | 84 | 98 | 0.99468894 | 1 |
| 176 | 102 | 782 | 827 | 632 | 1942 | 229 | 254 | 287 | 0.73857183 | 0.91507608 |
| 22 | 2 | 20 | 16 | 34 | 26 | 20 | 10 | 5 | 0.001032 | 0.01245739 |
| 1 | 2 | 1 | 4 | 2 | 5 | 0 | 0 | 0 | 0.02084234 | 0.09761335 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 3 | 1 | 3 | 7 | 1 | 1 | 0 | 1 | 1 |
| 933 | 1741 | 14213 | 6141 | 9320 | 12162 | 7699 | 3055 | 4954 | 0.16670058 | 0.40145979 |
| 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0.71348377 | 0.90275157 |
| 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0.45613637 | 0.70097152 |
| 1342 | 1872 | 7720 | 8572 | 8984 | 15433 | 3357 | 7298 | 4647 | 0.28191101 | 0.51822695 |
| 5 | 1 | 36 | 37 | 18 | 48 | 12 | 27 | 16 | 0.23477737 | 0.51361322 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr15 | 89155077 | 89155155 | hsa-mir-3529 | - | 77 | 3 | 20 | 30 | 30 |
| chr9 | 35608090 | 35608156 | hsa-mir-4667 | + | 1 | 1 | 0 | 0 | 0 |
| chr16 | 593276 | 593366 | hsa-mir-3176 | + | 0 | 0 | 0 | 0 | 0 |
| chr7 | 150935506 | 150935624 | hsa-mir-671 | + | 255 | 247 | 138 | 16 | 46 |
| chr18 | 56118311 | 56118384 | hsa-mir-3591 | - | 6 | 0 | 5 | 0 | 3 |
| chr7 | 129410222 | 129410332 | hsa-mir-182 | - | 325 | 37 | 194 | 73 | 126 |
| chr14 | 101489661 | 101489757 | hsa-mir-411 | + | 39 | 1 | 0 | 19 | 18 |
| chr6 | 33175611 | 33175721 | hsa-mir-219-1 | + | 18 | 8 | 5 | 1 | 4 |
| chrX | 32601772 | 32601869 | hsa-mir-3915 | - | 1 | 0 | 0 | 0 | 0 |
| chr19 | 52196038 | 52196117 | hsa-let-7e | + | 1265 | 70 | 174 | 640 | 676 |
| chr1 | 243509477 | 243509557 | hsa-mir-4677 | + | 1 | 0 | 1 | 0 | 0 |
| chrX | 83480759 | 83480836 | hsa-mir-548i-4 | - | 0 | 0 | 0 | 0 | 0 |
| chr19 | 47212549 | 47212602 | hsa-mir-320e | - | 32 | 40 | 25 | 2 | 11 |
| chrX | 73438383 | 73438453 | hsa-mir-374c | + | 33 | 9 | 3 | 305 | 98 |
| chr8 | 104166841 | 104166917 | hsa-mir-3151 | + | 1 | 0 | 0 | 0 | 0 |
| chr14 | 101500091 | 101500173 | hsa-mir-495 | + | 83 | 5 | 6 | 28 | 84 |
| chr7 | 151130574 | 151130725 | hsa-mir-3907 | - | 1 | 0 | 4 | 0 | 0 |
| chr11 | 68850643 | 68850726 | hsa-mir-3164 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 100152185 | 100152269 | hsa-mir-548p | - | 1 | 0 | 0 | 2 | 0 |
| chrX | 109298556 | 109298654 | hsa-mir-652 | + | 583 | 89 | 336 | 134 | 174 |
| chr15 | 22156765 | 22156847 | hsa-mir-5701-2 | + | 1 | 3 | 0 | 0 | 1 |
| chr14 | 101532248 | 101532328 | hsa-mir-410 | + | 9 | 0 | 1 | 2 | 5 |
| chr14 | 100575991 | 100576090 | hsa-mir-342 | + | 477 | 79 | 140 | 101 | 199 |
| chr9 | 96941115 | 96941202 | hsa-let-7d | + | 2752 | 306 | 971 | 1372 | 1495 |
| chr2 | 189997761 | 189997837 | hsa-mir-3129 | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 162312335 | 162312430 | hsa-mir-556 | + | 3 | 1 | 0 | 1 | 6 |
| chr16 | 50776215 | 50776288 | hsa-mir-3181 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 77680984 | 77681058 | hsa-mir-4739 | - | 1 | 0 | 0 | 0 | 0 |
| chr9 | 28888876 | 28888953 | hsa-mir-873 | - | 7 | 2 | 8 | 0 | 1 |
| chr22 | 45596834 | 45596900 | hsa-mir-1249 | - | 28 | 8 | 12 | 0 | 3 |
| chr8 | 10892715 | 10892812 | hsa-mir-598 | - | 232 | 144 | 379 | 82 | 18 |
| chr6 | 16141786 | 16141855 | hsa-mir-4639 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 7126615 | 7126698 | hsa-mir-324 | - | 449 | 197 | 259 | 92 | 112 |
| chr5 | 9053927 | 9054007 | hsa-mir-4636 | - | 2 | 0 | 0 | 1 | 2 |
| chr17 | 65467604 | 65467701 | hsa-mir-548d-2 | - | 1 | 0 | 0 | 2 | 1 |
| chr1 | 53394345 | 53394444 | hsa-mir-1273f | + | 5 | 2 | 1 | 0 | 3 |
| chr17 | 54968630 | 54968716 | hsa-mir-3614 | - | 1 | 0 | 1 | 2 | 0 |
| chr10 | 65132716 | 65132808 | hsa-mir-1296 | - | 11 | 2 | 5 | 4 | 1 |
| chr15 | 75081012 | 75081098 | hsa-mir-4513 | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 172107947 | 172108028 | hsa-mir-3120 | + | 375 | 93 | 86 | 60 | 80 |
| chr14 | 101531934 | 101532004 | hsa-mir-369 | + | 46 | 1 | 2 | 56 | 70 |
| chr13 | 42142421 | 42142531 | hsa-mir-5006 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 146318430 | 146318545 | hsa-mir-508 | - | 0 | 0 | 0 | 0 | 0 |
| chr5 | 153726665 | 153726807 | hsa-mir-1294 | + | 1 | 1 | 1 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 17 | 20 | 359 | 322 | 480 | 2243 | 869 | 429 | 256 | 0.00284625 | 0.02550681 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 5 | 1 | 6 | 4 | 2 | 3 | 3 | 0.00860239 | 0.05436539 |
| 51 | 3 | 145 | 85 | 200 | 142 | 37 | 21 | 12 | 0.09081228 | 0.2762803 |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.11E-05 | 0.00093731 |
| 93 | 169 | 1350 | 1405 | 2728 | 4265 | 1013 | 1542 | 1232 | 0.01876746 | 0.09110039 |
| 19 | 74 | 149 | 96 | 42 | 105 | 8 | 37 | 96 | 0.98315097 | 1 |
| 3 | 3 | 8 | 15 | 14 | 12 | 12 | 3 | 10 | 0.4734852 | 0.71405858 |
| 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 539 | 1110 | 5706 | 5101 | 5093 | 3431 | 717 | 3150 | 1171 | 0.32075469 | 0.55462593 |
| 0 | 1 | 2 | 7 | 1 | 4 | 2 | 8 | 6 | 0.1562316 | 0.38217284 |
| 0 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 0.68634275 | 0.90275157 |
| 3 | 3 | 11 | 4 | 15 | 3 | 1 | 2 | 2 | 0.00232086 | 0.02229942 |
| 76 | 175 | 1113 | 974 | 294 | 1414 | 4004 | 918 | 467 | 0.00966283 | 0.05810166 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 58 | 67 | 44 | 37 | 39 | 33 | 22 | 17 | 50 | 0.03159401 | 0.13029034 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00902398 | 0.05533126 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 4 | 4 | 2 | 2 | 1 | 4 | 2 | 0.32687267 | 0.56415802 |
| 84 | 119 | 360 | 311 | 229 | 425 | 333 | 136 | 111 | 0.19804961 | 0.45463605 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0087498 | 0.05436539 |
| 4 | 20 | 27 | 18 | 4 | 21 | 5 | 2 | 10 | 0.53293924 | 0.77007654 |
| 172 | 133 | 6209 | 1104 | 555 | 2737 | 694 | 367 | 465 | 0.10673556 | 0.29258101 |
| 1277 | 1120 | 5708 | 7124 | 7250 | 4750 | 2440 | 1785 | 1372 | 0.86450822 | 1 |
| 0 | 0 | 3 | 7 | 2 | 4 | 2 | 0 | 2 | 0.01643535 | 0.08279864 |
| 3 | 11 | 51 | 10 | 7 | 11 | 42 | 9 | 17 | 0.28033749 | 0.51635286 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 1 | 0 | 1 | 4 | 16 | 0 | 5 | 0 | 2 | 0.38878928 | 0.6368218 |
| 1 | 0 | 8 | 21 | 24 | 20 | 2 | 3 | 7 | 0.36302631 | 0.60852611 |
| 50 | 37 | 95 | 67 | 54 | 84 | 7 | 25 | 25 | 0.00443557 | 0.03388486 |
| 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 123 | 94 | 354 | 292 | 389 | 529 | 282 | 290 | 77 | 0.43114509 | 0.69083642 |
| 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | 0 | 0.05419784 | 0.19731401 |
| 1 | 1 | 12 | 10 | 6 | 11 | 10 | 21 | 10 | 0.04546696 | 0.17155953 |
| 0 | 0 | 7 | 3 | 3 | 3 | 1 | 4 | 2 | 0.59292944 | 0.8235622 |
| 0 | 0 | 6 | 8 | 0 | 2 | 19 | 1 | 3 | 0.11990313 | 0.3139037 |
| 1 | 9 | 39 | 63 | 39 | 87 | 8 | 9 | 9 | 0.22015417 | 0.49204721 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 84 | 86 | 243 | 248 | 307 | 266 | 42 | 167 | 241 | 0.40593342 | 0.6591114 |
| 83 | 90 | 203 | 138 | 126 | 111 | 59 | 18 | 89 | 0.59258355 | 0.8235622 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 0 | 0 | 2 | 0 | 0 | 1 | 4 | 0 | 2 | 0.07025388 | 0.24526482 |
| 1 | 0 | 1 | 4 | 7 | 7 | 0 | 4 | 1 | 0.35314621 | 0.59625411 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chrX | 153246548 | 153246627 | hsa-mir-3202-2 | - | 1 | 0 | 0 | 0 | 0 |
| chr3 | 195426271 | 195426368 | hsa-mir-570 | + | 1 | 1 | 0 | 4 | 2 |
| chr19 | 50357847 | 50357908 | hsa-mir-4749 | + | 1 | 0 | 0 | 0 | 0 |
| chr1 | 220291498 | 220291583 | hsa-mir-194-1 | - | 36 | 6 | 4 | 28 | 16 |
| chr3 | 69098108 | 69098186 | hsa-mir-3136 | - | 1 | 0 | 0 | 0 | 0 |
| chrX | 37883147 | 37883239 | hsa-mir-548aj-2 | - | 2 | 0 | 0 | 2 | 2 |
| chr2 | 111042429 | 111042520 | hsa-mir-4436b-2 | + | 0 | 0 | 0 | 0 | 0 |
| chr11 | 74110281 | 74110378 | hsa-mir-548al | + | 2 | 1 | 0 | 0 | 2 |
| chr11 | 79133212 | 79133270 | hsa-mir-5579 | - | 0 | 0 | 0 | 0 | 0 |
| chr13 | 92003567 | 92003645 | hsa-mir-92a-1 | + | 215 | 57 | 79 | 409 | 209 |
| chr12 | 97957589 | 97957689 | hsa-mir-135a-2 | + | 9 | 2 | 3 | 3 | 4 |
| chr19 | 54177450 | 54177574 | hsa-mir-498 | + | 1 | 0 | 0 | 0 | 0 |
| chr21 | 17911408 | 17911489 | hsa-mir-99a | + | 13414 | 6547 | 3467 | 3951 | 7922 |
| chr3 | 19356339 | 19356423 | hsa-mir-4791 | - | 1 | 1 | 2 | 0 | 0 |
| chr9 | 127454720 | 127454830 | hsa-mir-181a-2 | + | 1830 | 791 | 725 | 392 | 1324 |
| chr17 | 1617196 | 1617281 | hsa-mir-22 | - | 59075 | 20967 | 45709 | 7861 | 11657 |
| chr14 | 65937819 | 65937904 | hsa-mir-625 | + | 59 | 11 | 25 | 16 | 6 |
| chr11 | 74431312 | 74431382 | hsa-mir-4696 | - | 1 | 0 | 0 | 0 | 0 |
| chr14 | 104583754 | 104583840 | hsa-mir-203b | - | 88 | 4 | 21 | 47 | 124 |
| chr11 | 71783273 | 71783348 | hsa-mir-3165 | - | 0 | 0 | 0 | 0 | 0 |
| chr20 | 57392669 | 57392749 | hsa-mir-296 | - | 62 | 37 | 45 | 4 | 8 |
| chr9 | 97848302 | 97848370 | hsa-mir-24-1 | + | 8799 | 2377 | 2169 | 1017 | 2380 |
| chr14 | 23426158 | 23426238 | hsa-mir-4707 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 7065900 | 7065978 | hsa-mir-4767 | + | 0 | 0 | 0 | 0 | 0 |
| chr21 | 17912147 | 17912231 | hsa-let-7c | + | 7631 | 952 | 2074 | 2566 | 3950 |
| chr8 | 22102474 | 22102556 | hsa-mir-320a | - | 10074 | 6821 | 8313 | 1362 | 3652 |
| chrX | 53584152 | 53584235 | hsa-let-7f-2 | - | 4107 | 300 | 1442 | 5288 | 7422 |
| chrX | 133675370 | 133675467 | hsa-mir-542 | - | 641 | 24 | 39 | 35 | 239 |
| chr1 | 9211726 | 9211836 | hsa-mir-34a | - | 975 | 117 | 127 | 231 | 241 |
| chr15 | 22049273 | 22049348 | hsa-mir-3118-6 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 109849529 | 109849616 | hsa-mir-548f-3 | - | 1 | 0 | 0 | 0 | 1 |
| chr13 | 41301963 | 41302011 | hsa-mir-320d-1 | - | 25 | 45 | 31 | 4 | 15 |
| chr17 | 46657199 | 46657309 | hsa-mir-10a | - | 909 | 90 | 110 | 643 | 580 |
| chr9 | 92785722 | 92785817 | hsa-mir-4290 | - | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101348222 | 101348315 | hsa-mir-433 | + | 27 | 4 | 1 | 2 | 5 |
| chr10 | 23682333 | 23682396 | hsa-mir-1254-2 | + | 2 | 2 | 3 | 1 | 0 |
| chr8 | 120337410 | 120337505 | hsa-mir-548az | + | 2 | 1 | 0 | 4 | 1 |
| chrX | 133303700 | 133303796 | hsa-mir-19b-2 | - | 7376 | 2161 | 527 | 1558 | 2240 |
| chr1 | 19209695 | 19209769 | hsa-mir-4695 | - | 1 | 0 | 0 | 0 | 0 |
| chr5 | 54466359 | 54466450 | hsa-mir-449a | - | 2 | 0 | 0 | 0 | 0 |
| chrX | 140008336 | 140008384 | hsa-mir-320d-2 | - | 24 | 45 | 31 | 4 | 15 |
| chr14 | 101513657 | 101513735 | hsa-mir-539 | + | 29 | 0 | 0 | 7 | 12 |
| chr12 | 81329514 | 81329612 | hsa-mir-618 | - | 2 | 4 | 3 | 3 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 4 | 4 | 7 | 4 | 11 | 8 | 14 | 10 | 0.44674348 | 0.70097152 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 13 | 20 | 122 | 75 | 75 | 257 | 149 | 55 | 113 | 0.23414495 | 0.51361322 |
| 0 | 1 | 2 | 1 | 2 | 8 | 0 | 1 | 2 | 0.26778987 | 0.51361322 |
| 2 | 0 | 8 | 5 | 1 | 11 | 3 | 4 | 2 | 0.63380131 | 0.86868062 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 2 | 7 | 1 | 2 | 0 | 0.85609113 | 1 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 115 | 417 | 6129 | 1826 | 735 | 1540 | 203 | 424 | 463 | 0.18674037 | 0.44061272 |
| 5 | 8 | 29 | 3 | 1 | 3 | 4 | 6 | 3 | 0.27420549 | 0.51361322 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 6958 | 5595 | 6273 | 4268 | 3615 | 4186 | 610 | 1604 | 3831 | 0.0002366 | 0.00469175 |
| 1 | 2 | 2 | 3 | 1 | 2 | 6 | 3 | 1 | 1 | 1 |
| 661 | 818 | 3071 | 1397 | 2907 | 2042 | 357 | 2689 | 2455 | 0.71619731 | 0.90275157 |
| 5572 | 7349 | 14209 | 21608 | 34851 | 27999 | 11867 | 10505 | 15841 | 0.00175633 | 0.01860114 |
| 41 | 6 | 171 | 144 | 51 | 322 | 152 | 42 | 132 | 0.30601537 | 0.53593028 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.34426564 | 0.58337378 |
| 78 | 193 | 1170 | 2044 | 4435 | 1672 | 1177 | 159 | 1351 | 0.00257074 | 0.02360966 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 17 | 9 | 63 | 24 | 23 | 29 | 6 | 5 | 8 | 0.08364559 | 0.27257933 |
| 1384 | 1994 | 9391 | 8940 | 14651 | 19581 | 3799 | 6418 | 4537 | 0.83551167 | 0.98753789 |
| 0 | 0 | 2 | 0 | 7 | 19 | 0 | 1 | 1 | 0.00893971 | 0.05517756 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 2901 | 3632 | 6812 | 5327 | 4748 | 4400 | 405 | 1238 | 801 | 0.03471838 | 0.13947213 |
| 2864 | 1160 | 2095 | 1746 | 3159 | 2426 | 821 | 842 | 694 | 1.24E-05 | 0.00063984 |
| 5518 | 7258 | 23493 | 21135 | 17903 | 15621 | 32109 | 9860 | 10275 | 0.55403944 | 0.79073 |
| 261 | 203 | 282 | 104 | 232 | 158 | 514 | 339 | 183 | 0.16791677 | 0.40334647 |
| 157 | 375 | 3414 | 1943 | 2595 | 7210 | 886 | 498 | 729 | 0.16976208 | 0.40673074 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2 | 0 | 1 | 3 | 0 | 1 | 0 | 0 | 0 | 0.26416253 | 0.51361322 |
| 11 | 6 | 9 | 6 | 22 | 7 | 6 | 2 | 3 | 0.00614173 | 0.04208889 |
| 429 | 729 | 2461 | 7194 | 2438 | 2627 | 603 | 588 | 1407 | 0.42122847 | 0.67803961 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 15 | 5 | 17 | 14 | 14 | 14 | 1 | 4 | 2 | 0.09526949 | 0.27628247 |
| 2 | 0 | 3 | 2 | 1 | 1 | 1 | 2 | 1 | 0.23568817 | 0.51361322 |
| 2 | 2 | 5 | 1 | 2 | 13 | 10 | 3 | 5 | 0.88077094 | 1 |
| 1632 | 1673 | 8323 | 5363 | 2337 | 2119 | 5114 | 1469 | 2714 | 0.254266 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 2 | 15 | 245 | 92 | 40 | 4 | 2 | 246 | 7.25E-06 | 0.00045065 |
| 11 | 6 | 9 | 6 | 22 | 7 | 6 | 2 | 3 | 0.00635862 | 0.04325719 |
| 31 | 15 | 23 | 24 | 18 | 12 | 9 | 10 | 10 | 0.1150536 | 0.30549845 |
| 2 | 0 | 17 | 8 | 2 | 1 | 5 | 0 | 0 | 0.87597952 | 1 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr9 | 94398532 | 94398643 | hsa-mir-3910-1 | + | 3 | 1 | 0 | 1 | 0 |
| chr2 | 161264320 | 161264393 | hsa-mir-4785 | - | 2 | 1 | 2 | 0 | 1 |
| chr3 | 32547774 | 32547881 | hsa-mir-548ay | - | 2 | 0 | 0 | 3 | 1 |
| chr7 | 97592969 | 97593038 | hsa-mir-5692a-1 | + | 0 | 0 | 0 | 0 | 0 |
| chr11 | 87909669 | 87909761 | hsa-mir-3166 | + | 2 | 2 | 0 | 0 | 0 |
| chr9 | 21512113 | 21512184 | hsa-mir-31 | - | 102 | 22 | 72 | 23 | 40 |
| chr17 | 57215118 | 57215233 | hsa-mir-454 | - | 5 | 2 | 1 | 25 | 3 |
| chr7 | 130561505 | 130561569 | hsa-mir-29a | - | 7865 | 583 | 478 | 4294 | 3534 |
| chr15 | 89911247 | 89911337 | hsa-mir-9-3 | + | 31 | 1 | 0 | 16 | 33 |
| chr17 | 56408592 | 56408679 | hsa-mir-142 | - | 1040 | 165 | 575 | 2542 | 567 |
| chr10 | 135061014 | 135061124 | hsa-mir-202 | - | 3 | 3 | 0 | 10 | 0 |
| chrX | 49767753 | 49767844 | hsa-mir-532 | + | 305 | 64 | 53 | 82 | 77 |
| chr19 | 54244566 | 54244661 | hsa-mir-517c | + | 2 | 0 | 0 | 0 | 0 |
| chr14 | 101507118 | 101507186 | hsa-mir-376a-1 | + | 533 | 251 | 36 | 119 | 230 |
| chr4 | 8007027 | 8007108 | hsa-mir-95 | - | 12 | 0 | 2 | 6 | 9 |
| chr5 | 170813659 | 170813764 | hsa-mir-3912 | - | 2 | 1 | 0 | 3 | 0 |
| chr5 | 169455491 | 169455570 | hsa-mir-378e | + | 14 | 11 | 7 | 0 | 4 |
| chr14 | 101491353 | 101491414 | hsa-mir-380 | + | 3 | 1 | 0 | 1 | 1 |
| chrX | 146341169 | 146341244 | hsa-mir-509-3 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 49779205 | 49779291 | hsa-mir-502 | + | 100 | 50 | 33 | 31 | 36 |
| chr14 | 101526091 | 101526175 | hsa-mir-154 | + | 38 | 1 | 2 | 3 | 20 |
| chr8 | 128972878 | 128972941 | hsa-mir-1205 | + | 1 | 0 | 0 | 0 | 0 |
| chr14 | 101335396 | 101335485 | hsa-mir-493 | + | 13 | 1 | 1 | 2 | 7 |
| chr12 | 58218391 | 58218475 | hsa-mir-26a-2 | - | 4163 | 378 | 334 | 3322 | 2742 |
| chr19 | 54210706 | 54210793 | hsa-mir-520c | + | 1 | 0 | 0 | 1 | 0 |
| chrX | 11336733 | 11336806 | hsa-mir-548ax | - | 1 | 0 | 0 | 1 | 1 |
| chr11 | 28078361 | 28078457 | hsa-mir-610 | + | 1 | 0 | 0 | 0 | 0 |
| chrX | 108297771 | 108297820 | hsa-mir-6087 | + | 14 | 9 | 20 | 0 | 1 |
| chr19 | 20579239 | 20579322 | hsa-mir-1270-2 | - | 4 | 1 | 2 | 1 | 1 |
| chr18 | 33514050 | 33514105 | hsa-mir-3929 | - | 2 | 1 | 0 | 1 | 2 |
| chr17 | 57443443 | 57443515 | hsa-mir-4729 | + | 0 | 0 | 0 | 0 | 0 |
| chr11 | 79113065 | 79113153 | hsa-mir-708 | - | 744 | 50 | 57 | 107 | 93 |
| chr7 | 102046188 | 102046302 | hsa-mir-548o | - | 1 | 0 | 0 | 1 | 1 |
| chr3 | 114462291 | 114462372 | hsa-mir-4796 | - | 0 | 0 | 0 | 0 | 0 |
| chr22 | 33832567 | 33832655 | hsa-mir-4764 | - | 1 | 0 | 0 | 0 | 0 |
| chr20 | 3898140 | 3898218 | hsa-mir-103a-2 | + | 3006 | 460 | 344 | 1185 | 2097 |
| chr8 | 144815252 | 144815323 | hsa-mir-4664 | - | 1 | 0 | 0 | 0 | 1 |
| chr17 | 79106995 | 79107108 | hsa-mir-1250 | - | 1 | 0 | 0 | 0 | 0 |
| chr8 | 125834226 | 125834293 | hsa-mir-4662a | + | 2 | 0 | 0 | 1 | 1 |
| chr14 | 48230197 | 48230307 | hsa-mir-548y | - | 1 | 1 | 3 | 1 | 3 |
| chr1 | 3477258 | 3477354 | hsa-mir-551a | - | 3 | 1 | 1 | 1 | 0 |
| chr8 | 9599181 | 9599278 | hsa-mir-597 | + | 2 | 0 | 1 | 1 | 1 |
| chr4 | 148265780 | 148265869 | hsa-mir-548g | - | 2 | 0 | 0 | 3 | 2 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00355881 | 0.02935229 |
| 0 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 0 | 0.19778847 | 0.45463605 |
| 2 | 2 | 11 | 7 | 4 | 11 | 7 | 14 | 7 | 0.3061858 | 0.53593028 |
| 0 | 0 | 3 | 0 | 1 | 3 | 1 | 0 | 1 | 0.11092121 | 0.30003945 |
| 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 0.58122758 | 0.81581943 |
| 48 | 44 | 30 | 24 | 42 | 34 | 21 | 3 | 11 | 0.00782708 | 0.05061433 |
| 6 | 6 | 311 | 100 | 41 | 63 | 45 | 96 | 83 | 0.02673983 | 0.11484571 |
| 3085 | 2325 | 21820 | 22796 | 12054 | 27941 | 3140 | 6503 | 5196 | 0.46959136 | 0.71048562 |
| 80 | 37 | 45 | 17 | 39 | 265 | 26 | 9 | 22 | 0.52416426 | 0.76330157 |
| 500 | 1089 | 8547 | 23327 | 1909 | 7516 | 27952 | 4111 | 4007 | 0.01506782 | 0.07673885 |
| 2 | 3 | 23 | 13 | 0 | 1 | 12 | 2 | 1 | 0.82816001 | 0.9832422 |
| 71 | 98 | 310 | 255 | 287 | 407 | 884 | 232 | 94 | 0.7956863 | 0.95441394 |
| 0 | 0 | 1 | 0 | 0 | 1 | 2 | 23 | 1 | 0.07503755 | 0.25419227 |
| 590 | 264 | 170 | 120 | 156 | 109 | 24 | 23 | 73 | 0.00087951 | 0.0113848 |
| 2 | 6 | 168 | 73 | 34 | 139 | 53 | 29 | 22 | 0.04144446 | 0.16204427 |
| 0 | 3 | 12 | 2 | 4 | 1 | 7 | 4 | 2 | 0.78607726 | 0.9441031 |
| 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1.93E-05 | 0.00093731 |
| 3 | 3 | 5 | 5 | 0 | 1 | 0 | 0 | 1 | 0.14770097 | 0.36557988 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| 41 | 24 | 87 | 134 | 117 | 173 | 212 | 82 | 37 | 0.96044324 | 1 |
| 21 | 21 | 36 | 26 | 24 | 40 | 3 | 5 | 18 | 0.19221721 | 0.44453211 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 12 | 10 | 54 | 66 | 68 | 97 | 10 | 3 | 39 | 0.24166016 | 0.51361322 |
| 1565 | 4927 | 20607 | 11540 | 6520 | 22627 | 4988 | 8182 | 10370 | 0.37227288 | 0.62067679 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01496683 | 0.07664331 |
| 1 | 1 | 9 | 9 | 5 | 7 | 1 | 7 | 4 | 0.20284475 | 0.46222814 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 2 | 1 | 2 | 2 | 0 | 2 | 0 | 0.00053138 | 0.00811876 |
| 2 | 1 | 1 | 1 | 7 | 16 | 1 | 2 | 1 | 0.70882537 | 0.90275157 |
| 1 | 0 | 2 | 3 | 2 | 6 | 0 | 1 | 0 | 0.5239792 | 0.76330157 |
| 0 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0.68634275 | 0.90275157 |
| 92 | 81 | 452 | 574 | 816 | 914 | 206 | 337 | 735 | 0.98631416 | 1 |
| 1 | 1 | 13 | 5 | 2 | 8 | 9 | 8 | 5 | 0.11968387 | 0.3139037 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 907 | 1273 | 10470 | 5736 | 7739 | 8270 | 8296 | 3905 | 3919 | 0.30138079 | 0.5346388 |
| 0 | 0 | 17 | 32 | 18 | 44 | 14 | 0 | 3 | 0.00155413 | 0.01684238 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 12 | 4 | 1 | 93 | 2 | 2 | 1 | 0.01705688 | 0.08455859 |
| 2 | 0 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 0.03511004 | 0.14044017 |
| 0 | 0 | 4 | 6 | 3 | 5 | 1 | 23 | 0 | 0.23766177 | 0.51361322 |
| 2 | 1 | 5 | 8 | 2 | 8 | 1 | 2 | 1 | 1 | 1 |
| 0 | 0 | 6 | 7 | 2 | 11 | 4 | 5 | 2 | 0.45806301 | 0.70097152 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr4 | 102251458 | 102251571 | hsa-mir-1255a | - | 2 | 0 | 1 | 0 | 1 |
| chr15 | 42491767 | 42491864 | hsa-mir-627 | - | 29 | 33 | 30 | 7 | 9 |
| chr10 | 12767252 | 12767352 | hsa-mir-548q | - | 2 | 2 | 0 | 4 | 3 |
| chr19 | 54197646 | 54197729 | hsa-mir-526b | + | 2 | 2 | 1 | 0 | 0 |
| chr7 | 129414744 | 129414854 | hsa-mir-183 | - | 126 | 15 | 35 | 34 | 51 |
| chr15 | 31357234 | 31357344 | hsa-mir-211 | - | 1 | 1 | 1 | 0 | 0 |
| chr14 | 101526909 | 101527011 | hsa-mir-496 | + | 2 | 1 | 0 | 0 | 0 |
| chr8 | 124360273 | 124360370 | hsa-mir-548d-1 | - | 1 | 0 | 0 | 2 | 1 |
| chr8 | 141742662 | 141742752 | hsa-mir-151a | - | 2115 | 382 | 245 | 640 | 916 |
| chr10 | 100684255 | 100684325 | hsa-mir-6507 | - | 1 | 0 | 1 | 0 | 3 |
| chr13 | 66540461 | 66540561 | hsa-mir-548x-2 | - | 1 | 0 | 0 | 0 | 0 |
| chr1 | 156905922 | 156906036 | hsa-mir-765 | - | 2 | 0 | 4 | 0 | 0 |
| chr19 | 54191734 | 54191821 | hsa-mir-1283-1 | + | 1 | 0 | 0 | 0 | 0 |
| chr12 | 109029585 | 109029646 | hsa-mir-4496 | + | 0 | 0 | 0 | 0 | 0 |
| chr4 | 1988110 | 1988204 | hsa-mir-943 | - | 1 | 1 | 0 | 0 | 0 |
| chr1 | 201688635 | 201688755 | hsa-mir-5191 | + | 1 | 0 | 0 | 0 | 1 |
| chr3 | 49843569 | 49843678 | hsa-mir-5193 | - | 3 | 0 | 0 | 0 | 0 |
| chr1 | 193105632 | 193105713 | hsa-mir-1278 | + | 0 | 0 | 0 | 0 | 0 |
| chr22 | 22007592 | 22007674 | hsa-mir-130b | + | 98 | 15 | 36 | 48 | 64 |
| chr13 | 54886106 | 54886183 | hsa-mir-1297 | - | 0 | 0 | 0 | 0 | 0 |
| chr5 | 154065335 | 154065421 | hsa-mir-1303 | + | 235 | 38 | 129 | 20 | 44 |
| chr7 | 32772592 | 32772689 | hsa-mir-550a-2 | + | 4 | 4 | 3 | 1 | 3 |
| chr17 | 6921229 | 6921341 | hsa-mir-497 | - | 6964 | 2655 | 633 | 823 | 4807 |
| chr10 | 17887106 | 17887193 | hsa-mir-511-1 | + | 50 | 8 | 4 | 9 | 7 |
| chr8 | 12576640 | 12576699 | hsa-mir-5692a-2 | + | 0 | 0 | 0 | 0 | 0 |
| chr20 | 44333740 | 44333819 | hsa-mir-3617 | - | 2 | 0 | 0 | 0 | 0 |
| chrX | 133303567 | 133303642 | hsa-mir-92a-2 | - | 157 | 38 | 59 | 372 | 183 |
| chr14 | 101515886 | 101515983 | hsa-mir-655 | + | 3 | 0 | 1 | 13 | 9 |
| chr9 | 94398545 | 94398627 | hsa-mir-3910-2 | - | 3 | 1 | 0 | 1 | 0 |
| chr16 | 14403141 | 14403228 | hsa-mir-365a | + | 94 | 8 | 10 | 91 | 44 |
| chr20 | 33578202 | 33578275 | hsa-mir-499b | - | 39 | 6 | 0 | 11 | 16 |
| chr9 | 68415307 | 68415388 | hsa-mir-4477a | - | 0 | 0 | 0 | 0 | 0 |
| chr5 | 10478148 | 10478257 | hsa-mir-6131 | + | 9 | 2 | 3 | 6 | 1 |
| chr19 | 54205990 | 54206073 | hsa-mir-518b | + | 5 | 0 | 0 | 1 | 1 |
| chr19 | 49812053 | 49812125 | hsa-mir-4324 | - | 3 | 0 | 1 | 1 | 0 |
| chr14 | 101518782 | 101518862 | hsa-mir-487a | + | 24 | 1 | 3 | 3 | 6 |
| chr6 | 72086662 | 72086734 | hsa-mir-30c-2 | - | 895 | 85 | 116 | 533 | 593 |
| chr14 | 101351038 | 101351120 | hsa-mir-136 | + | 1356 | 197 | 26 | 102 | 133 |
| chr8 | 26906369 | 26906480 | hsa-mir-548h-4 | - | 3 | 1 | 0 | 6 | 3 |
| chr14 | 101512791 | 101512875 | hsa-mir-487b | + | 67 | 12 | 3 | 15 | 36 |
| chr4 | 20529897 | 20530007 | hsa-mir-218-1 | + | 194 | 30 | 34 | 102 | 87 |
| chr1 | 41220026 | 41220118 | hsa-mir-30e | + | 3350 | 704 | 485 | 1932 | 2388 |
| chr19 | 13985512 | 13985622 | hsa-mir-181c | + | 168 | 50 | 21 | 72 | 99 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 1 | 0 | 1 | 2 | 2 | 6 | 7 | 0 | 1 | 0.77954278 | 0.93988858 |
| 6 | 2 | 21 | 29 | 17 | 26 | 8 | 11 | 7 | 0.16518919 | 0.39988655 |
| 2 | 2 | 2 | 3 | 3 | 1 | 2 | 3 | 3 | 0.22780647 | 0.50551342 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.07554856 | 0.25419227 |
| 50 | 58 | 856 | 1044 | 1692 | 3479 | 412 | 969 | 486 | 0.00315828 | 0.02725475 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02636789 | 0.11377256 |
| 1 | 2 | 2 | 2 | 1 | 6 | 1 | 0 | 1 | 0.60736431 | 0.8380582 |
| 1 | 1 | 9 | 7 | 4 | 10 | 6 | 21 | 9 | 0.07690294 | 0.2578185 |
| 613 | 1137 | 4515 | 3854 | 4743 | 5531 | 3157 | 3482 | 5161 | 0.33076629 | 0.5656407 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01043913 | 0.06157765 |
| 0 | 0 | 3 | 3 | 5 | 4 | 1 | 2 | 3 | 0.06096652 | 0.21694137 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00416763 | 0.03236859 |
| 0 | 0 | 1 | 1 | 4 | 1 | 0 | 100 | 1 | 0.00028679 | 0.00494969 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |
| 0 | 0 | 2 | 1 | 1 | 2 | 0 | 0 | 0 | 0.3028855 | 0.5346388 |
| 41 | 44 | 274 | 388 | 389 | 510 | 271 | 323 | 168 | 0.23119984 | 0.50963907 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 21 | 7 | 82 | 68 | 71 | 107 | 91 | 57 | 24 | 0.11910969 | 0.3139037 |
| 1 | 2 | 6 | 3 | 6 | 8 | 3 | 5 | 2 | 0.5557169 | 0.79073 |
| 3070 | 2488 | 3361 | 2498 | 3228 | 2707 | 634 | 1079 | 1174 | 0.00495653 | 0.03608976 |
| 13 | 5 | 26 | 16 | 11 | 13 | 3 | 14 | 6 | 0.0559964 | 0.20228156 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0.68634275 | 0.90275157 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 104 | 351 | 5157 | 1604 | 637 | 1227 | 168 | 365 | 347 | 0.19492368 | 0.44967541 |
| 18 | 42 | 45 | 27 | 18 | 25 | 22 | 9 | 30 | 0.60786196 | 0.8380582 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00355881 | 0.02935229 |
| 20 | 51 | 483 | 158 | 130 | 461 | 102 | 102 | 162 | 0.53109252 | 0.76859973 |
| 6 | 27 | 74 | 29 | 28 | 115 | 18 | 26 | 96 | 0.81886662 | 0.97469181 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 3 | 6 | 6 | 11 | 2 | 1 | 0 | 0.19877011 | 0.4551689 |
| 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7.29E-05 | 0.00205793 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0.03991804 | 0.15697727 |
| 4 | 2 | 6 | 11 | 12 | 5 | 2 | 2 | 12 | 0.11376291 | 0.3038024 |
| 586 | 442 | 4121 | 2225 | 1344 | 2308 | 914 | 1033 | 555 | 0.71879721 | 0.90275157 |
| 534 | 160 | 276 | 341 | 157 | 214 | 43 | 50 | 282 | 0.00465436 | 0.03498275 |
| 2 | 3 | 7 | 8 | 8 | 19 | 7 | 9 | 8 | 0.72601013 | 0.90581183 |
| 55 | 44 | 56 | 66 | 53 | 40 | 8 | 13 | 35 | 0.13658952 | 0.34877105 |
| 126 | 123 | 283 | 152 | 120 | 124 | 116 | 27 | 58 | 0.19178711 | 0.44453211 |
| 1197 | 3309 | 7908 | 7213 | 4603 | 10618 | 9566 | 5164 | 3824 | 0.67442319 | 0.90275157 |
| 61 | 144 | 192 | 69 | 120 | 325 | 40 | 548 | 259 | 0.85460682 | 1 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr21 | 20058407 | 20058482 | hsa-mir-548x | - | 1 | 0 | 0 | 2 | 1 |
| chr9 | 96938628 | 96938715 | hsa-let-7f-1 | + | 4010 | 314 | 1423 | 5189 | 7268 |
| chr12 | 62654139 | 62654235 | hsa-mir-6125 | + | 1 | 0 | 0 | 0 | 0 |
| chr6 | 99572484 | 99572572 | hsa-mir-548ai | + | 3 | 1 | 0 | 2 | 3 |
| chr16 | 69966983 | 69967083 | hsa-mir-140 | + | 2893 | 714 | 1352 | 814 | 1056 |
| chr9 | 97847489 | 97847586 | hsa-mir-23b | + | 1104 | 114 | 103 | 1186 | 911 |
| chrX | 73506938 | 73507044 | hsa-mir-545 | - | 9 | 0 | 0 | 8 | 2 |
| chrX | 17444003 | 17444077 | hsa-mir-4768 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 220291194 | 220291304 | hsa-mir-215 | - | 29 | 4 | 3 | 12 | 20 |
| chr7 | 99691182 | 99691266 | hsa-mir-25 | - | 748 | 105 | 167 | 1140 | 494 |
| chr4 | 53578848 | 53578914 | hsa-mir-4449 | + | 2 | 0 | 0 | 1 | 0 |
| chr19 | 46178189 | 46178266 | hsa-mir-642b | - | 9 | 1 | 3 | 3 | 2 |
| chr20 | 33578178 | 33578300 | hsa-mir-499a | + | 39 | 6 | 0 | 11 | 16 |
| chrX | 146294980 | 146295109 | hsa-mir-513a-1 | - | 0 | 0 | 0 | 0 | 0 |
| chr8 | 105496596 | 105496693 | hsa-mir-548a-3 | - | 0 | 0 | 0 | 0 | 0 |
| chr19 | 13947400 | 13947473 | hsa-mir-23a | - | 6952 | 804 | 721 | 4635 | 3842 |
| chr11 | 93466839 | 93466930 | hsa-mir-1304 | - | 3 | 0 | 1 | 2 | 2 |
| chr3 | 111831647 | 111831745 | hsa-mir-567 | + | 0 | 0 | 0 | 0 | 0 |
| chr22 | 35731632 | 35731751 | hsa-mir-3909 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 35836415 | 35836530 | hsa-mir-5196 | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 112748683 | 112748771 | hsa-mir-548e | + | 10 | 1 | 2 | 6 | 2 |
| chr19 | 813583 | 813653 | hsa-mir-3187 | + | 0 | 0 | 0 | 0 | 0 |
| chr6 | 6169566 | 6169642 | hsa-mir-5683 | + | 31 | 0 | 0 | 3 | 8 |
| chr13 | 88270919 | 88270995 | hsa-mir-4500 | - | 1 | 0 | 1 | 0 | 0 |
| chr1 | 142667288 | 142667363 | hsa-mir-3118-1 | + | 0 | 0 | 0 | 0 | 0 |
| chr2 | 87421908 | 87421982 | hsa-mir-4771-1 | + | 0 | 0 | 0 | 0 | 0 |
| chr12 | 57912945 | 57913042 | hsa-mir-616 | - | 2 | 1 | 1 | 3 | 0 |
| chr2 | 56210101 | 56210211 | hsa-mir-217 | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 1103242 | 1103332 | hsa-mir-200a | + | 655 | 38 | 48 | 456 | 838 |
| chr5 | 179225277 | 179225346 | hsa-mir-1229 | - | 0 | 0 | 0 | 0 | 0 |
| chr15 | 49461266 | 49461350 | hsa-mir-4716 | - | 4 | 0 | 2 | 0 | 0 |
| chr12 | 66417399 | 66417506 | hsa-mir-6074 | - | 1 | 0 | 0 | 0 | 0 |
| chr15 | 22513228 | 22513280 | hsa-mir-1268a | - | 8 | 8 | 13 | 0 | 2 |
| chr8 | 96085141 | 96085221 | hsa-mir-3150a | + | 1 | 0 | 0 | 0 | 0 |
| chr12 | 124020955 | 124021081 | hsa-mir-3908 | + | 1 | 1 | 0 | 2 | 0 |
| chr22 | 46508628 | 46508702 | hsa-let-7a-3 | + | 4651 | 392 | 980 | 5349 | 4655 |
| chr1 | 166123979 | 166124035 | hsa-mir-921 | - | 1 | 0 | 0 | 0 | 0 |
| chr3 | 5291861 | 5291940 | hsa-mir-4790 | - | 1 | 0 | 0 | 0 | 0 |
| chr15 | 59463381 | 59463461 | hsa-mir-2116 | - | 5 | 3 | 1 | 0 | 5 |
| chr17 | 29902429 | 29902540 | hsa-mir-365b | + | 88 | 5 | 10 | 92 | 44 |
| chr14 | 101280827 | 101280911 | hsa-mir-2392 | + | 5 | 0 | 0 | 0 | 1 |
| chr19 | 4770681 | 4770791 | hsa-mir-7-3 | + | 23 | 1 | 7 | 10 | 9 |
| chr18 | 70520555 | 70520617 | hsa-mir-548av | - | 0 | 0 | 0 | 0 | 0 |
| chr11 | 118889653 | 118889722 | hsa-mir-3656 | + | 2 | 3 | 2 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 5 | 4 | 1 | 9 | 3 | 4 | 1 | 0.30923209 | 0.5376946 |
| 5408 | 7103 | 22838 | 20436 | 17374 | 14986 | 31112 | 9597 | 9667 | 0.56930502 | 0.80609294 |
| 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 2 | 3 | 3 | 1 | 4 | 5 | 3 | 2 | 0.58071752 | 0.81581943 |
| 831 | 801 | 1636 | 2504 | 1540 | 1720 | 1180 | 3091 | 1164 | 0.26912197 | 0.51361322 |
| 820 | 682 | 2493 | 3159 | 3042 | 5519 | 2861 | 2758 | 1616 | 0.47646995 | 0.71624193 |
| 2 | 3 | 18 | 36 | 4 | 35 | 4 | 48 | 12 | 0.26841255 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 8 | 12 | 4 | 8 | 11 | 16 | 6 | 14 | 0.071396 | 0.2464484 |
| 561 | 1001 | 3604 | 2997 | 2635 | 3820 | 1583 | 3847 | 2529 | 0.33319443 | 0.56874946 |
| 0 | 0 | 2 | 0 | 1 | 2 | 0 | 0 | 0 | 0.37454865 | 0.62266644 |
| 1 | 1 | 3 | 11 | 3 | 19 | 0 | 1 | 0 | 0.38442023 | 0.63412329 |
| 6 | 27 | 74 | 29 | 28 | 115 | 18 | 26 | 96 | 0.81886662 | 0.97469181 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 0 | 0 | 2 | 2 | 1 | 11 | 3 | 1 | 1 | 0.01137532 | 0.06626125 |
| 3006 | 2895 | 13209 | 13285 | 9922 | 20978 | 4213 | 5754 | 4756 | 0.92758477 | 1 |
| 0 | 2 | 5 | 4 | 1 | 18 | 2 | 12 | 0 | 0.69268365 | 0.90275157 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 0 | 0 | 19 | 35 | 17 | 33 | 4 | 5 | 4 | 0.00013441 | 0.00329431 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2 | 6 | 21 | 22 | 9 | 41 | 46 | 33 | 10 | 0.23803316 | 0.51361322 |
| 0 | 0 | 2 | 2 | 5 | 5 | 0 | 1 | 0 | 0.03272762 | 0.13437066 |
| 0 | 12 | 19 | 3 | 1 | 5 | 2 | 0 | 1 | 0.00473693 | 0.03531855 |
| 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0.46181056 | 0.70328014 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 2 | 9 | 3 | 1 | 22 | 3 | 2 | 3 | 0.82519285 | 0.98096905 |
| 0 | 0 | 7 | 6 | 5 | 2 | 3 | 0 | 0 | 0.01148723 | 0.06649751 |
| 377 | 1314 | 13177 | 4718 | 3979 | 19492 | 590 | 2289 | 1331 | 0.03139967 | 0.13006443 |
| 0 | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 0.28463513 | 0.52015675 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 0.09879355 | 0.28286708 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 2 | 5 | 9 | 7 | 6 | 1 | 2 | 4 | 0.17265601 | 0.41049846 |
| 0 | 0 | 1 | 2 | 17 | 1 | 2 | 0 | 0 | 0.05307151 | 0.19473482 |
| 0 | 0 | 2 | 2 | 1 | 1 | 1 | 1 | 3 | 1 | 1 |
| 3653 | 6234 | 33731 | 22015 | 22725 | 21365 | 5118 | 5074 | 4924 | 0.45954249 | 0.70097152 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1 | 4 | 5 | 3 | 4 | 3 | 1 | 0 | 2 | 0.10661877 | 0.29258101 |
| 21 | 51 | 479 | 162 | 131 | 480 | 95 | 101 | 164 | 0.50257068 | 0.74823623 |
| 0 | 1 | 13 | 6 | 3 | 1 | 1 | 0 | 3 | 1 | 1 |
| 6 | 6 | 124 | 107 | 161 | 752 | 293 | 146 | 91 | 0.00398527 | 0.03124828 |
| 0 | 0 | 3 | 1 | 1 | 1 | 0 | 1 | 1 | 0.09906778 | 0.28286708 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0025879 | 0.02360966 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chrX | 153996870 | 153996931 | hsa-mir-664b | + | 6 | 1 | 2 | 1 | 2 |
| chr3 | 48238053 | 48238106 | hsa-mir-4443 | + | 1 | 0 | 0 | 0 | 0 |
| chr18 | 47652868 | 47652933 | hsa-mir-4320 | - | 29 | 18 | 63 | 11 | 6 |
| chr10 | 132760850 | 132760931 | hsa-mir-378c | - | 794 | 462 | 544 | 83 | 292 |
| chr5 | 92956401 | 92956494 | hsa-mir-2277 | - | 11 | 3 | 2 | 1 | 1 |
| chr1 | 205417429 | 205417526 | hsa-mir-135b | - | 28 | 0 | 0 | 60 | 7 |
| chrX | 133680643 | 133680741 | hsa-mir-424 | - | 5358 | 265 | 144 | 790 | 2145 |
| chr15 | 35664456 | 35664565 | hsa-mir-3942 | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 7073259 | 7073354 | hsa-mir-141 | + | 3614 | 291 | 164 | 1615 | 3914 |
| chr10 | 24564613 | 24564710 | hsa-mir-603 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 117214370 | 117214449 | hsa-mir-320b-1 | + | 212 | 209 | 251 | 32 | 72 |
| chr21 | 15017095 | 15017171 | hsa-mir-3118-5 | - | 0 | 0 | 0 | 0 | 0 |
| chr9 | 28863623 | 28863704 | hsa-mir-876 | - | 7 | 1 | 1 | 0 | 0 |
| chr22 | 38240278 | 38240378 | hsa-mir-658 | - | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101514237 | 101514316 | hsa-mir-889 | + | 0 | 0 | 0 | 0 | 0 |
| chr4 | 40504056 | 40504136 | hsa-mir-4802 | - | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54185412 | 54185499 | hsa-mir-520f | + | 1 | 0 | 1 | 1 | 1 |
| chrX | 49773038 | 49773122 | hsa-mir-500a | + | 108 | 38 | 27 | 28 | 31 |
| chr20 | 62550777 | 62550894 | hsa-mir-941-1 | + | 9 | 1 | 1 | 2 | 4 |
| chr16 | 16418444 | 16418511 | hsa-mir-6511a-2 | + | 2 | 1 | 0 | 0 | 0 |
| chr11 | 568088 | 568198 | hsa-mir-210 | - | 313 | 157 | 261 | 34 | 45 |
| chr2 | 242417319 | 242417397 | hsa-mir-3133 | + | 0 | 0 | 0 | 0 | 0 |
| chr11 | 3877291 | 3877371 | hsa-mir-4687 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 45939911 | 45939962 | hsa-mir-6088 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 53371347 | 53371413 | hsa-mir-4459 | - | 1 | 0 | 1 | 0 | 1 |
| chr14 | 101350819 | 101350913 | hsa-mir-432 | + | 39 | 1 | 2 | 5 | 14 |
| chr6 | 166922841 | 166922921 | hsa-mir-1913 | - | 1 | 0 | 0 | 0 | 0 |
| chr4 | 174189310 | 174189384 | hsa-mir-548t | + | 2 | 0 | 0 | 2 | 1 |
| chr16 | 2140195 | 2140285 | hsa-mir-1225 | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 172107937 | 172108047 | hsa-mir-214 | - | 377 | 93 | 86 | 61 | 80 |
| chr12 | 121160995 | 121161069 | hsa-mir-4700 | + | 35 | 39 | 11 | 0 | 15 |
| chr19 | 54233091 | 54233179 | hsa-mir-518e | + | 6 | 1 | 2 | 2 | 1 |
| chr7 | 99691390 | 99691470 | hsa-mir-93 | - | 1433 | 195 | 170 | 1579 | 1520 |
| chr19 | 54219847 | 54219934 | hsa-mir-521-2 | + | 1 | 0 | 0 | 0 | 0 |
| chr1 | 1104384 | 1104467 | hsa-mir-429 | + | 58 | 2 | 4 | 79 | 63 |
| chr20 | 60639857 | 60639941 | hsa-mir-3195 | + | 158 | 35 | 414 | 6 | 6 |
| chr5 | 148441875 | 148441972 | hsa-mir-584 | - | 6 | 0 | 0 | 7 | 0 |
| chr16 | 1784985 | 1785067 | hsa-mir-3177 | + | 1 | 0 | 0 | 0 | 0 |
| chr19 | 46142251 | 46142345 | hsa-mir-330 | - | 40 | 8 | 4 | 5 | 6 |
| chr6 | 132436331 | 132436403 | hsa-mir-548aj-1 | - | 0 | 0 | 0 | 0 | 0 |
| chr11 | 56511348 | 56511457 | hsa-mir-6128 | + | 2 | 2 | 1 | 1 | 2 |
| chr18 | 13459945 | 13460025 | hsa-mir-5190 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54198466 | 54198547 | hsa-mir-519b | + | 0 | 0 | 0 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 1 | 2 | 14 | 9 | 5 | 20 | 5 | 3 | 1 | 0.81094821 | 0.96897914 |
| 1 | 0 | 5 | 12 | 8 | 27 | 3 | 2 | 3 | 0.01984478 | 0.09520546 |
| 4 | 0 | 11 | 5 | 4 | 7 | 0 | 0 | 0 | 0.00028508 | 0.00494969 |
| 94 | 201 | 73 | 52 | 90 | 179 | 119 | 80 | 65 | 0.00029239 | 0.00495475 |
| 0 | 1 | 10 | 11 | 9 | 6 | 2 | 2 | 1 | 0.58409656 | 0.81861352 |
| 5 | 19 | 106 | 918 | 16 | 441 | 7 | 1637 | 40 | 0.001917 | 0.01985156 |
| 3579 | 2052 | 3392 | 2290 | 2318 | 3063 | 3623 | 4987 | 2046 | 0.2145665 | 0.48303377 |
| 0 | 0 | 1 | 2 | 1 | 2 | 3 | 1 | 0 | 0.07416326 | 0.25390681 |
| 2710 | 3305 | 29926 | 11345 | 13326 | 22841 | 9244 | 17863 | 13985 | 0.09625306 | 0.27737408 |
| 0 | 0 | 1 | 0 | 0 | 3 | 2 | 0 | 1 | 0.28777831 | 0.52282532 |
| 63 | 25 | 52 | 41 | 104 | 44 | 19 | 21 | 14 | 0.00292117 | 0.0256842 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 1 | 2 | 1 | 2 | 0 | 2 | 0 | 3 | 0.12846743 | 0.33351433 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 13 | 14 | 3 | 16 | 5 | 3 | 12 | 0.00065694 | 0.00958327 |
| 0 | 0 | 9 | 12 | 4 | 24 | 2 | 1 | 5 | 0.00147889 | 0.01660629 |
| 0 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0.57083835 | 0.80609294 |
| 31 | 22 | 106 | 140 | 92 | 160 | 227 | 85 | 46 | 0.93931291 | 1 |
| 2 | 2 | 23 | 27 | 21 | 32 | 7 | 10 | 3 | 0.40067484 | 0.65254564 |
| 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0.45613637 | 0.70097152 |
| 46 | 24 | 277 | 406 | 3601 | 2116 | 91 | 559 | 319 | 0.11358037 | 0.3038024 |
| 0 | 0 | 2 | 2 | 0 | 3 | 1 | 0 | 0 | 0.09545381 | 0.27628247 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 2 | 0 | 3 | 3 | 1 | 2 | 1 | 0.8371202 | 0.98758991 |
| 15 | 20 | 17 | 18 | 10 | 36 | 0 | 1 | 7 | 0.04424452 | 0.16830976 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1 | 1 | 13 | 6 | 5 | 10 | 9 | 3 | 5 | 0.23654493 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 86 | 86 | 246 | 249 | 311 | 266 | 42 | 168 | 242 | 0.4076117 | 0.6606854 |
| 3 | 0 | 1 | 1 | 17 | 5 | 2 | 1 | 0 | 0.00077485 | 0.01017123 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 1 | 0.01434122 | 0.0746705 |
| 574 | 884 | 6220 | 3791 | 5437 | 7114 | 2573 | 4078 | 3239 | 0.32941976 | 0.5656407 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 24 | 123 | 2588 | 1306 | 872 | 5290 | 116 | 359 | 315 | 0.0021834 | 0.02142033 |
| 20 | 0 | 11 | 20 | 7 | 11 | 0 | 2 | 1 | 2.01E-06 | 0.00014435 |
| 2 | 6 | 5 | 3 | 0 | 12 | 2 | 23 | 8 | 0.87010575 | 1 |
| 0 | 0 | 1 | 4 | 3 | 9 | 1 | 0 | 1 | 0.06979643 | 0.24526482 |
| 9 | 5 | 53 | 35 | 31 | 110 | 27 | 61 | 14 | 0.64604003 | 0.88027676 |
| 0 | 0 | 2 | 0 | 1 | 1 | 0 | 1 | 1 | 0.3028855 | 0.5346388 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00102414 | 0.01245739 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 15 | 1 | 0.02247518 | 0.10168384 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chrX | 49774329 | 49774413 | hsa-mir-501 | + | 99 | 33 | 27 | 16 | 12 |
| chr22 | 26951177 | 26951289 | hsa-mir-548j | - | 1 | 0 | 1 | 2 | 1 |
| chr1 | 62544457 | 62544531 | hsa-mir-3116-1 | + | 1 | 0 | 0 | 0 | 0 |
| chrX | 153246547 | 153246628 | hsa-mir-3202-1 | + | 1 | 0 | 0 | 0 | 0 |
| chr20 | 61151512 | 61151583 | hsa-mir-1-1 | + | 5 | 0 | 5 | 41 | 23 |
| chr22 | 28316512 | 28316600 | hsa-mir-3199-1 | - | 6 | 0 | 9 | 1 | 1 |
| chr10 | 105154009 | 105154158 | hsa-mir-1307 | - | 855 | 88 | 134 | 53 | 84 |
| chr22 | 20073580 | 20073665 | hsa-mir-1306 | + | 20 | 4 | 7 | 1 | 3 |
| chr14 | 101349315 | 101349412 | hsa-mir-127 | + | 2918 | 978 | 645 | 113 | 728 |
| chrX | 73438211 | 73438296 | hsa-mir-421 | - | 7 | 1 | 1 | 7 | 2 |
| chr14 | 104583741 | 104583851 | hsa-mir-203a | + | 88 | 4 | 21 | 47 | 124 |
| chr12 | 81226311 | 81226408 | hsa-mir-617 | - | 0 | 0 | 0 | 0 | 0 |
| chr16 | 56892429 | 56892513 | hsa-mir-138-2 | + | 12 | 17 | 12 | 2 | 9 |
| chrX | 55477927 | 55478015 | hsa-mir-4536-1 | - | 3 | 0 | 0 | 1 | 0 |
| chr2 | 213290986 | 213291084 | hsa-mir-548f-2 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 117520356 | 117520434 | hsa-mir-1277 | + | 2 | 1 | 1 | 18 | 2 |
| chr16 | 15019793 | 15019860 | hsa-mir-6511a-1 | + | 2 | 1 | 0 | 0 | 0 |
| chr8 | 135812762 | 135812850 | hsa-mir-30b | - | 2472 | 152 | 336 | 1154 | 1147 |
| chr5 | 36147993 | 36148090 | hsa-mir-580 | - | 1 | 0 | 0 | 1 | 1 |
| chrX | 146363460 | 146363548 | hsa-mir-514a-2 | - | 2 | 0 | 0 | 1 | 1 |
| chr11 | 34963383 | 34963467 | hsa-mir-1343 | + | 2 | 0 | 3 | 0 | 0 |
| chr19 | 54251889 | 54251976 | hsa-mir-521-1 | + | 1 | 0 | 0 | 0 | 0 |
| chr5 | 153975571 | 153975632 | hsa-mir-3141 | - | 1 | 4 | 3 | 0 | 0 |
| chr7 | 102111915 | 102111978 | hsa-mir-4467 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54225419 | 54225509 | hsa-mir-520g | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 32394483 | 32394581 | hsa-mir-579 | - | 1 | 1 | 0 | 0 | 0 |
| chr14 | 50766572 | 50766664 | hsa-mir-4504 | - | 1 | 0 | 0 | 0 | 0 |
| chr19 | 54214255 | 54214342 | hsa-mir-524 | + | 6 | 0 | 0 | 0 | 0 |
| chr14 | 101522555 | 101522637 | hsa-mir-323b | + | 1 | 0 | 0 | 0 | 1 |
| chr16 | 30886586 | 30886644 | hsa-mir-4519 | - | 0 | 0 | 0 | 0 | 0 |
| chr8 | 17539086 | 17539166 | hsa-mir-548v | - | 1 | 0 | 0 | 1 | 2 |
| chr16 | 15227922 | 15228007 | hsa-mir-6511b-1 | - | 5 | 2 | 3 | 1 | 1 |
| chr7 | 75544514 | 75544587 | hsa-mir-4651 | + | 1 | 0 | 1 | 0 | 0 |
| chr1 | 62544460 | 62544528 | hsa-mir-3116-2 | - | 1 | 0 | 0 | 0 | 0 |
| chr19 | 54224329 | 54224396 | hsa-mir-517b | + | 2 | 0 | 0 | 1 | 0 |
| chr20 | 3898148 | 3898210 | hsa-mir-103b-2 | - | 2878 | 437 | 324 | 1256 | 2004 |
| chr14 | 101531636 | 101531715 | hsa-mir-409 | + | 357 | 94 | 46 | 31 | 172 |
| chr1 | 1102483 | 1102578 | hsa-mir-200b | + | 585 | 86 | 61 | 440 | 1645 |
| chr1 | 21314806 | 21314925 | hsa-mir-1256 | - | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101492068 | 101492154 | hsa-mir-323a | + | 12 | 0 | 0 | 1 | 3 |
| chr12 | 95702195 | 95702289 | hsa-mir-331 | + | 349 | 262 | 219 | 63 | 94 |
| chr3 | 186504460 | 186504566 | hsa-mir-1248 | + | 4 | 2 | 2 | 1 | 1 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 35 | 13 | 69 | 71 | 62 | 124 | 163 | 33 | 14 | 0.74320796 | 0.91695911 |
| 0 | 2 | 2 | 5 | 4 | 5 | 6 | 1 | 0 | 0.9072973 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 17 | 32 | 87 | 47 | 42 | 23 | 18 | 14 | 5 | 0.51439607 | 0.75617844 |
| 0 | 4 | 4 | 1 | 3 | 4 | 5 | 0 | 4 | 0.1629151 | 0.39540852 |
| 62 | 80 | 637 | 401 | 537 | 967 | 42 | 525 | 168 | 0.58982635 | 0.82170128 |
| 2 | 0 | 26 | 16 | 40 | 24 | 2 | 6 | 2 | 0.97606651 | 1 |
| 773 | 543 | 472 | 553 | 689 | 483 | 54 | 88 | 354 | 0.00069921 | 0.00958327 |
| 4 | 4 | 41 | 54 | 18 | 171 | 52 | 80 | 37 | 0.0157085 | 0.07956698 |
| 78 | 193 | 1170 | 2044 | 4435 | 1673 | 1177 | 159 | 1351 | 0.00256837 | 0.02360966 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 29 | 3 | 8 | 24 | 3 | 49 | 6 | 2 | 8 | 0.21606247 | 0.48522944 |
| 0 | 0 | 1 | 0 | 3 | 1 | 1 | 0 | 1 | 0.43214339 | 0.69083642 |
| 0 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 0 | 0.45250806 | 0.70097152 |
| 0 | 7 | 206 | 118 | 23 | 224 | 94 | 87 | 79 | 0.00423267 | 0.03260203 |
| 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0.45613637 | 0.70097152 |
| 1332 | 764 | 7787 | 5429 | 5895 | 6762 | 4606 | 2258 | 2818 | 0.38816561 | 0.6368218 |
| 1 | 0 | 2 | 3 | 1 | 1 | 0 | 0 | 1 | 0.80348063 | 0.96128876 |
| 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 0.18973507 | 0.44319069 |
| 1 | 0 | 2 | 4 | 6 | 6 | 1 | 1 | 1 | 0.89752288 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0.0213194 | 0.09934842 |
| 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0.71923557 | 0.90275157 |
| 0 | 0 | 2 | 1 | 0 | 0 | 1 | 1 | 0 | 0.45250806 | 0.70097152 |
| 0 | 0 | 2 | 13 | 2 | 5 | 3 | 14 | 5 | 0.02551378 | 0.11163775 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00017196 | 0.00372712 |
| 0 | 6 | 3 | 1 | 2 | 4 | 3 | 4 | 5 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 0 | 1 | 6 | 10 | 3 | 15 | 8 | 5 | 2 | 0.1451948 | 0.36182235 |
| 0 | 0 | 2 | 2 | 4 | 5 | 0 | 1 | 1 | 0.24968918 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 3 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0.0882563 | 0.2762803 |
| 866 | 1219 | 9930 | 5416 | 7521 | 7817 | 8513 | 3773 | 3437 | 0.30706735 | 0.53593028 |
| 293 | 123 | 106 | 111 | 154 | 100 | 27 | 17 | 72 | 0.00871072 | 0.05436539 |
| 475 | 1235 | 12080 | 5623 | 4419 | 14422 | 712 | 1525 | 2614 | 0.07783882 | 0.26002073 |
| 0 | 0 | 2 | 1 | 3 | 7 | 1 | 0 | 1 | 0.03311354 | 0.13476775 |
| 11 | 5 | 4 | 12 | 7 | 12 | 2 | 1 | 2 | 0.13997904 | 0.35381202 |
| 79 | 54 | 919 | 733 | 607 | 848 | 1047 | 191 | 235 | 0.51364582 | 0.75617844 |
| 2 | 1 | 13 | 3 | 1 | 5 | 2 | 3 | 3 | 0.72765898 | 0.9066553 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chrX | 65238711 | 65238821 | hsa-mir-223 | + | 200 | 10 | 21 | 454 | 71 |
| chr1 | 53400601 | 53400689 | hsa-mir-5095 | + | 3 | 4 | 0 | 0 | 1 |
| chr12 | 69978501 | 69978603 | hsa-mir-3913-1 | - | 7 | 4 | 2 | 0 | 3 |
| chr11 | 121970464 | 121970552 | hsa-mir-125b-1 | - | 18569 | 8831 | 13470 | 4792 | 5261 |
| chr2 | 54076258 | 54076342 | hsa-mir-3682 | - | 0 | 0 | 0 | 0 | 0 |
| chr5 | 136983260 | 136983338 | hsa-mir-874 | - | 361 | 162 | 211 | 25 | 52 |
| chr10 | 121718024 | 121718104 | hsa-mir-4682 | + | 0 | 0 | 0 | 0 | 0 |
| chr8 | 7946462 | 7946611 | hsa-mir-548i-3 | - | 0 | 0 | 0 | 0 | 0 |
| chr14 | 60113679 | 60113738 | hsa-mir-5586 | - | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54240098 | 54240188 | hsa-mir-516b-1 | + | 3 | 0 | 0 | 0 | 0 |
| chr15 | 45725247 | 45725327 | hsa-mir-147b | + | 32 | 20 | 31 | 7 | 6 |
| chr20 | 2633422 | 2633488 | hsa-mir-1292 | + | 2 | 1 | 3 | 0 | 1 |
| chr9 | 139565053 | 139565138 | hsa-mir-126 | + | 4652 | 190 | 9812 | 7626 | 4497 |
| chr8 | 65291705 | 65291814 | hsa-mir-124-2 | + | 17 | 0 | 0 | 4 | 1 |
| chr9 | 130452965 | 130453074 | hsa-mir-3911 | - | 1 | 0 | 1 | 0 | 0 |
| chr19 | 13947253 | 13947331 | hsa-mir-27a | - | 4703 | 284 | 263 | 2605 | 1848 |
| chrX | 49777848 | 49777945 | hsa-mir-660 | + | 141 | 10 | 9 | 62 | 40 |
| chr3 | 38010894 | 38010971 | hsa-mir-26a-1 | + | 4154 | 377 | 332 | 3313 | 2731 |
| chr20 | 62572817 | 62572897 | hsa-mir-1914 | - | 0 | 0 | 0 | 0 | 0 |
| chr11 | 70718374 | 70718473 | hsa-mir-3664 | - | 4 | 1 | 1 | 0 | 1 |
| chr18 | 6374359 | 6374424 | hsa-mir-4317 | - | 2 | 0 | 0 | 0 | 0 |
| chr20 | 57393280 | 57393368 | hsa-mir-298 | - | 2 | 0 | 2 | 0 | 1 |
| chr14 | 101506555 | 101506636 | hsa-mir-654 | + | 29 | 2 | 8 | 10 | 15 |
| chr3 | 175087328 | 175087410 | hsa-mir-4789 | + | 0 | 0 | 0 | 0 | 0 |
| chr6 | 33665904 | 33666011 | hsa-mir-3934 | + | 1 | 0 | 2 | 0 | 2 |
| chr6 | 33967748 | 33967828 | hsa-mir-1275 | - | 2 | 2 | 1 | 0 | 0 |
| chr7 | 25989538 | 25989606 | hsa-mir-148a | - | 9060 | 799 | 1505 | 2662 | 7762 |
| chr19 | 54259994 | 54260084 | hsa-mir-516a-1 | + | 4 | 0 | 0 | 0 | 0 |
| chr6 | 32717688 | 32717756 | hsa-mir-3135b | - | 124 | 22 | 50 | 1 | 4 |
| chr9 | 86584662 | 86584772 | hsa-mir-7-1 | - | 26 | 2 | 9 | 15 | 17 |
| chr22 | 38243684 | 38243781 | hsa-mir-659 | - | 2 | 1 | 0 | 0 | 1 |
| chr1 | 51525689 | 51525775 | hsa-mir-6500 | + | 14 | 3 | 1 | 4 | 3 |
| chr12 | 57588286 | 57588359 | hsa-mir-1228 | + | 9 | 0 | 1 | 1 | 1 |
| chr7 | 100802753 | 100802836 | hsa-mir-4653 | + | 1 | 1 | 0 | 0 | 0 |
| chr17 | 27188386 | 27188458 | hsa-mir-451a | - | 13127 | 781 | 1856 | 163887 | 5476 |
| chr14 | 101498323 | 101498401 | hsa-mir-543 | + | 52 | 3 | 4 | 4 | 8 |
| chr6 | 10439949 | 10440027 | hsa-mir-5689 | + | 0 | 0 | 0 | 0 | 0 |
| chr2 | 101925911 | 101925996 | hsa-mir-5696 | + | 1 | 0 | 0 | 1 | 0 |
| chr11 | 2017988 | 2018061 | hsa-mir-675 | - | 8 | 3 | 0 | 3 | 4 |
| chr1 | 155164967 | 155165063 | hsa-mir-92b | + | 18 | 4 | 16 | 8 | 5 |
| chr9 | 111808508 | 111808578 | hsa-mir-32 | - | 30 | 1 | 0 | 115 | 19 |
| chr5 | 8461037 | 8461112 | hsa-mir-4458 | + | 1 | 0 | 0 | 0 | 0 |
| chr4 | 164014725 | 164014780 | hsa-mir-4454 | - | 89 | 33 | 18 | 10 | 7 |
| chr9 | 20716103 | 20716187 | hsa-mir-491 | + | 53 | 65 | 34 | 13 | 17 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 84 | 92 | 441 | 1274 | 221 | 682 | 209 | 149 | 113 | 0.92253477 | 1 |
| 0 | 1 | 6 | 4 | 4 | 10 | 0 | 4 | 0 | 1 | 1 |
| 3 | 9 | 17 | 5 | 13 | 14 | 10 | 8 | 7 | 0.90240954 | 1 |
| 5540 | 3123 | 10308 | 6759 | 4623 | 5243 | 216 | 1595 | 2113 | 0.00014492 | 0.00329431 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 62 | 31 | 86 | 147 | 199 | 132 | 15 | 27 | 39 | 0.04179118 | 0.16204427 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0.71923557 | 0.90275157 |
| 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00066457 | 0.00958327 |
| 2 | 1 | 195 | 37 | 39 | 132 | 178 | 61 | 8 | 0.21947558 | 0.49170972 |
| 1 | 0 | 4 | 2 | 2 | 14 | 2 | 1 | 0 | 1 | 1 |
| 1565 | 5308 | 25456 | 13779 | 5918 | 9210 | 2345 | 2071 | 4555 | 0.36695045 | 0.61289931 |
| 0 | 0 | 6 | 0 | 2 | 1 | 0 | 0 | 1 | 0.00211681 | 0.02098795 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01912825 | 0.09237062 |
| 1509 | 2600 | 18305 | 17911 | 12093 | 26540 | 1970 | 9022 | 6779 | 0.16082169 | 0.39237124 |
| 30 | 35 | 382 | 252 | 114 | 465 | 432 | 95 | 109 | 0.36673536 | 0.61289931 |
| 1554 | 4916 | 20529 | 11501 | 6504 | 22502 | 4952 | 8168 | 10302 | 0.37518995 | 0.62266644 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.71923557 | 0.90275157 |
| 1 | 2 | 4 | 5 | 0 | 12 | 2 | 7 | 6 | 0.86557468 | 1 |
| 0 | 0 | 3 | 2 | 0 | 4 | 0 | 0 | 1 | 0.79793109 | 0.95587632 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0.06280706 | 0.22172795 |
| 34 | 33 | 83 | 62 | 51 | 60 | 2 | 7 | 33 | 0.61757701 | 0.85019464 |
| 0 | 0 | 1 | 2 | 1 | 10 | 1 | 4 | 0 | 0.01674037 | 0.08388186 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00328747 | 0.0277035 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00397825 | 0.03124828 |
| 4122 | 8527 | 8139 | 21710 | 28106 | 28779 | 4444 | 12014 | 13086 | 0.8143389 | 0.9717847 |
| 1 | 2 | 3 | 1 | 10 | 2 | 0 | 460 | 0 | 6.63E-05 | 0.00199388 |
| 9 | 2 | 5 | 16 | 5 | 20 | 5 | 3 | 1 | 0.0002649 | 0.00484292 |
| 10 | 16 | 197 | 165 | 217 | 859 | 349 | 174 | 136 | 0.00726209 | 0.04800188 |
| 0 | 0 | 1 | 1 | 0 | 0 | 3 | 0 | 0 | 0.57074903 | 0.80609294 |
| 5 | 7 | 5 | 0 | 0 | 1 | 2 | 0 | 0 | 0.00025071 | 0.00484292 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.21E-05 | 0.00063984 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 3409 | 8109 | 6793 | 2032 | 2083 | 1434 | 2664 | 566 | 806 | 7.01E-18 | 3.27E-15 |
| 18 | 9 | 14 | 28 | 17 | 14 | 4 | 4 | 8 | 0.04242361 | 0.16204427 |
| 0 | 0 | 1 | 3 | 2 | 0 | 0 | 0 | 0 | 0.3028855 | 0.5346388 |
| 1 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 1 | 0.76186194 | 0.9281769 |
| 3 | 3 | 14 | 5 | 9 | 13 | 0 | 1 | 2 | 0.39116712 | 0.63959255 |
| 8 | 11 | 463 | 198 | 254 | 248 | 14 | 49 | 47 | 0.0247868 | 0.1100062 |
| 19 | 33 | 762 | 606 | 177 | 780 | 346 | 250 | 430 | 0.02044554 | 0.09679945 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 9 | 31 | 104 | 73 | 65 | 152 | 17 | 143 | 76 | 0.93941147 | 1 |
| 12 | 17 | 19 | 45 | 64 | 103 | 34 | 9 | 16 | 0.32064819 | 0.55462593 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr10 | 6194169 | 6194225 | hsa-mir-3155b | - | 1 | 0 | 0 | 0 | 0 |
| chr3 | 183604587 | 183604673 | hsa-mir-4448 | + | 1 | 0 | 0 | 0 | 0 |
| chr10 | 112657847 | 112657913 | hsa-mir-4680 | + | 0 | 0 | 0 | 0 | 0 |
| chrX | 146366158 | 146366246 | hsa-mir-514a-3 | - | 2 | 0 | 0 | 1 | 1 |
| chr6 | 140526388 | 140526463 | hsa-mir-3668 | + | 2 | 0 | 1 | 0 | 0 |
| chr11 | 111384163 | 111384240 | hsa-mir-34c | + | 126 | 11 | 15 | 34 | 62 |
| chr14 | 101377475 | 101377550 | hsa-mir-370 | + | 180 | 68 | 60 | 3 | 36 |
| chr19 | 54255650 | 54255735 | hsa-mir-519a-1 | + | 2 | 0 | 1 | 2 | 0 |
| chr1 | 53405985 | 53406085 | hsa-mir-1273g | + | 2 | 1 | 2 | 1 | 0 |
| chr17 | 1953564 | 1953674 | hsa-mir-212 | - | 18 | 6 | 5 | 3 | 29 |
| chr14 | 101521755 | 101521828 | hsa-mir-485 | + | 79 | 39 | 30 | 4 | 8 |
| chr8 | 10524487 | 10524580 | hsa-mir-4286 | + | 238 | 22 | 39 | 16 | 23 |
| chrX | 133674537 | 133674637 | hsa-mir-450a-2 | - | 35 | 4 | 3 | 19 | 18 |
| chr15 | 86368865 | 86368961 | hsa-mir-548ap | + | 1 | 0 | 0 | 3 | 2 |
| chrX | 78156690 | 78156746 | hsa-mir-4328 | - | 2 | 0 | 1 | 0 | 0 |
| chr13 | 24736554 | 24736643 | hsa-mir-2276 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 79099676 | 79099755 | hsa-mir-3065 | + | 42 | 7 | 11 | 17 | 8 |
| chr15 | 70371710 | 70371807 | hsa-mir-629 | - | 39 | 17 | 92 | 13 | 21 |
| chr12 | 69978502 | 69978602 | hsa-mir-3913-2 | + | 7 | 4 | 2 | 0 | 3 |
| chr9 | 116971713 | 116971809 | hsa-mir-455 | + | 163 | 19 | 19 | 50 | 166 |
| chr18 | 33484780 | 33484889 | hsa-mir-187 | - | 24 | 1 | 1 | 1 | 2 |
| chr8 | 124228027 | 124228103 | hsa-mir-4663 | - | 1 | 0 | 0 | 0 | 0 |
| chrX | 39696814 | 39696867 | hsa-mir-1587 | + | 0 | 0 | 0 | 0 | 0 |
| chr13 | 92002858 | 92002942 | hsa-mir-17 | + | 1225 | 325 | 199 | 1003 | 1651 |
| chr15 | 55665137 | 55665232 | hsa-mir-628 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 133304227 | 133304308 | hsa-mir-106a | - | 72 | 18 | 18 | 94 | 73 |
| chr13 | 92003004 | 92003075 | hsa-mir-18a | + | 89 | 3 | 8 | 164 | 62 |
| chr7 | 44150447 | 44150511 | hsa-mir-4649 | + | 1 | 1 | 0 | 0 | 0 |
| chr5 | 172089167 | 172089266 | hsa-mir-5003 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 8090262 | 8090322 | hsa-mir-4521 | + | 68 | 10 | 1 | 6 | 9 |
| chr15 | 52569313 | 52569397 | hsa-mir-1266 | - | 0 | 0 | 0 | 0 | 0 |
| chr22 | 29729146 | 29729256 | hsa-mir-3653 | - | 35 | 75 | 124 | 1 | 7 |
| chr22 | 42296947 | 42297016 | hsa-mir-33a | + | 196 | 12 | 22 | 122 | 133 |
| chr5 | 53247333 | 53247429 | hsa-mir-581 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 47891044 | 47891119 | hsa-mir-1226 | + | 3 | 0 | 0 | 0 | 0 |
| chrX | 151128099 | 151128184 | hsa-mir-452 | - | 2009 | 145 | 266 | 190 | 278 |
| chr5 | 168195150 | 168195260 | hsa-mir-218-2 | - | 158 | 19 | 18 | 93 | 70 |
| chr10 | 100154974 | 100155064 | hsa-mir-1287 | - | 31 | 11 | 9 | 5 | 5 |
| chr1 | 162126896 | 162126972 | hsa-mir-4654 | + | 0 | 0 | 0 | 0 | 0 |
| chr7 | 30329409 | 30329506 | hsa-mir-550b-1 | - | 6 | 5 | 4 | 2 | 4 |
| chr10 | 97824071 | 97824156 | hsa-mir-3157 | - | 1 | 0 | 0 | 0 | 2 |
| chr17 | 62496891 | 62496957 | hsa-mir-3064 | - | 3 | 1 | 0 | 0 | 0 |
| chr7 | 93112071 | 93112167 | hsa-mir-653 | - | 36 | 3 | 0 | 2 | 3 |
| chrX | 146342049 | 146342143 | hsa-mir-509-1 | - | 0 | 0 | 0 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0.45250806 | 0.70097152 |
| 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 0.18973507 | 0.44319069 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0141554 | 0.07411706 |
| 74 | 61 | 213 | 102 | 182 | 100 | 71 | 38 | 256 | 0.78527632 | 0.9441031 |
| 70 | 21 | 18 | 26 | 46 | 26 | 3 | 3 | 10 | 0.00095736 | 0.01205865 |
| 2 | 0 | 1 | 1 | 10 | 1 | 0 | 589 | 1 | 2.79E-05 | 0.00108376 |
| 1 | 0 | 3 | 3 | 1 | 4 | 5 | 3 | 2 | 1 | 1 |
| 29 | 36 | 24 | 12 | 7 | 17 | 1 | 6 | 12 | 0.018366 | 0.08961841 |
| 19 | 19 | 19 | 12 | 18 | 24 | 0 | 5 | 6 | 0.00451339 | 0.03419903 |
| 102 | 32 | 112 | 301 | 112 | 351 | 93 | 113 | 18 | 0.6001779 | 0.83238958 |
| 35 | 50 | 140 | 44 | 47 | 76 | 85 | 57 | 47 | 0.99072686 | 1 |
| 2 | 1 | 10 | 1 | 1 | 9 | 4 | 4 | 6 | 0.7240809 | 0.90461581 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00734166 | 0.04818608 |
| 0 | 0 | 1 | 0 | 2 | 4 | 1 | 1 | 0 | 0.11118512 | 0.30003945 |
| 22 | 9 | 42 | 147 | 34 | 121 | 12 | 27 | 67 | 0.75797747 | 0.92586501 |
| 10 | 29 | 164 | 144 | 92 | 291 | 129 | 134 | 54 | 0.47517616 | 0.71545103 |
| 3 | 9 | 17 | 5 | 13 | 14 | 10 | 8 | 7 | 0.90240954 | 1 |
| 76 | 148 | 264 | 239 | 274 | 424 | 592 | 1488 | 584 | 0.20655825 | 0.46620335 |
| 2 | 4 | 19 | 12 | 110 | 63 | 1 | 43 | 0 | 0.28462724 | 0.52015675 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1017 | 1130 | 5512 | 3664 | 929 | 1716 | 1580 | 1190 | 1563 | 0.77560382 | 0.93635072 |
| 0 | 0 | 23 | 14 | 15 | 36 | 13 | 10 | 17 | 5.75E-05 | 0.00184834 |
| 42 | 43 | 134 | 132 | 38 | 268 | 89 | 85 | 768 | 0.51704327 | 0.75887296 |
| 26 | 108 | 922 | 636 | 144 | 512 | 105 | 465 | 222 | 0.27614055 | 0.51369858 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 27 | 10 | 4 | 7 | 11 | 5 | 3 | 5 | 2 | 0.00026501 | 0.00484292 |
| 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0.45250806 | 0.70097152 |
| 1 | 1 | 10 | 7 | 8 | 8 | 1 | 2 | 2 | 9.57E-05 | 0.0025479 |
| 126 | 189 | 1496 | 1032 | 1821 | 1404 | 441 | 893 | 314 | 0.09284684 | 0.2762803 |
| 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |
| 84 | 223 | 116 | 1077 | 306 | 83 | 21 | 259 | 54 | 0.00540776 | 0.03828921 |
| 100 | 117 | 265 | 129 | 113 | 118 | 107 | 25 | 53 | 0.24428548 | 0.51361322 |
| 12 | 9 | 43 | 31 | 23 | 89 | 8 | 9 | 9 | 0.7655127 | 0.92898156 |
| 0 | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 0 | 0.3036156 | 0.53491443 |
| 2 | 2 | 9 | 5 | 8 | 12 | 4 | 7 | 3 | 0.58836073 | 0.82088653 |
| 0 | 0 | 1 | 1 | 0 | 6 | 1 | 2 | 0 | 1 | 1 |
| 2 | 0 | 2 | 4 | 2 | 1 | 1 | 1 | 3 | 0.871296 | 1 |
| 5 | 6 | 26 | 31 | 17 | 40 | 2 | 5 | 3 | 0.44256046 | 0.70097152 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr10 | 114059369 | 114059446 | hsa-mir-6715b | - | 1 | 2 | 0 | 0 | 0 |
| chr5 | 154209017 | 154209100 | hsa-mir-378h | + | 19 | 6 | 9 | 0 | 2 |
| chr19 | 12897941 | 12898006 | hsa-mir-5684 | + | 2 | 0 | 0 | 0 | 0 |
| chr7 | 157367027 | 157367114 | hsa-mir-153-2 | - | 43 | 9 | 6 | 7 | 16 |
| chr10 | 106028093 | 106028163 | hsa-mir-4482 | - | 0 | 0 | 0 | 0 | 0 |
| chr16 | 15005076 | 15005170 | hsa-mir-3180-1 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 209605477 | 209605587 | hsa-mir-205 | + | 12375 | 5320 | 3291 | 666 | 3033 |
| chr17 | 27188550 | 27188636 | hsa-mir-144 | - | 1133 | 2 | 47 | 10519 | 156 |
| chr13 | 50623254 | 50623337 | hsa-mir-15a | - | 639 | 45 | 20 | 901 | 191 |
| chr11 | 64658826 | 64658911 | hsa-mir-194-2 | - | 49 | 13 | 7 | 40 | 24 |
| chr3 | 124870308 | 124870396 | hsa-mir-5092 | - | 0 | 0 | 0 | 0 | 0 |
| chr16 | 15104177 | 15104254 | hsa-mir-1972-1 | - | 1 | 1 | 1 | 0 | 0 |
| chr8 | 96085138 | 96085224 | hsa-mir-3150b | - | 1 | 0 | 0 | 0 | 0 |
| chr14 | 101347343 | 101347457 | hsa-mir-431 | + | 27 | 3 | 1 | 1 | 0 |
| chr17 | 28444096 | 28444190 | hsa-mir-423 | + | 4731 | 1532 | 2536 | 363 | 564 |
| chr19 | 54215521 | 54215608 | hsa-mir-517a | + | 2 | 0 | 0 | 1 | 0 |
| chr8 | 94928249 | 94928347 | hsa-mir-378d-2 | - | 213 | 162 | 200 | 30 | 47 |
| chrX | 8095005 | 8095102 | hsa-mir-651 | + | 12 | 0 | 5 | 4 | 1 |
| chr14 | 101492356 | 101492444 | hsa-mir-758 | + | 36 | 1 | 0 | 2 | 1 |
| chr8 | 117886966 | 117887039 | hsa-mir-3610 | - | 0 | 0 | 0 | 0 | 0 |
| chr4 | 166307393 | 166307489 | hsa-mir-578 | + | 1 | 0 | 0 | 0 | 0 |
| chr18 | 37256682 | 37256741 | hsa-mir-5583-2 | - | 1 | 0 | 0 | 0 | 0 |
| chr19 | 13031133 | 13031218 | hsa-mir-5695 | + | 8 | 4 | 2 | 1 | 0 |
| chr19 | 54238130 | 54238217 | hsa-mir-518d | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 148810208 | 148810296 | hsa-mir-145 | + | 26340 | 10898 | 10123 | 5855 | 5978 |
| chr20 | 62551084 | 62551201 | hsa-mir-941-3 | + | 18 | 3 | 1 | 3 | 9 |
| chr5 | 167987900 | 167987978 | hsa-mir-103a-1 | - | 2962 | 457 | 341 | 1168 | 2083 |
| chr14 | 101493121 | 101493201 | hsa-mir-329-1 | + | 22 | 1 | 3 | 2 | 7 |
| chr8 | 124360273 | 124360370 | hsa-mir-548aa-1 | + | 2 | 0 | 1 | 3 | 1 |
| chr14 | 77732560 | 77732633 | hsa-mir-1260a | + | 682 | 198 | 310 | 102 | 162 |
| chr19 | 54201638 | 54201725 | hsa-mir-523 | + | 1 | 0 | 0 | 0 | 0 |
| chr2 | 66585380 | 66585460 | hsa-mir-4778 | - | 2 | 0 | 0 | 0 | 0 |
| chrX | 133680357 | 133680428 | hsa-mir-503 | - | 125 | 5 | 7 | 10 | 26 |
| chr17 | 67095682 | 67095797 | hsa-mir-4524b | + | 14 | 4 | 12 | 1 | 0 |
| chr12 | 54730999 | 54731098 | hsa-mir-148b | + | 1224 | 101 | 81 | 345 | 420 |
| chr3 | 125509246 | 125509395 | hsa-mir-548i-1 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 69242706 | 69242773 | hsa-mir-676 | + | 1 | 0 | 1 | 1 | 0 |
| chrX | 63005881 | 63005967 | hsa-mir-1468 | - | 6 | 3 | 11 | 4 | 2 |
| chr17 | 28444103 | 28444178 | hsa-mir-3184 | - | 4731 | 1532 | 2536 | 363 | 564 |
| chr11 | 113320744 | 113320810 | hsa-mir-4301 | - | 1 | 0 | 1 | 0 | 0 |
| chrX | 113887129 | 113887198 | hsa-mir-1264 | + | 1 | 0 | 0 | 0 | 0 |
| chr19 | 50391431 | 50391487 | hsa-mir-4750 | + | 2 | 0 | 1 | 0 | 0 |
| chr7 | 126698141 | 126698238 | hsa-mir-592 | - | 2 | 0 | 0 | 0 | 0 |
| chr3 | 49311552 | 49311619 | hsa-mir-4271 | + | 1 | 0 | 0 | 0 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02809079 | 0.11793068 |
| 1 | 1 | 1 | 0 | 1 | 3 | 0 | 0 | 0 | 5.40E-05 | 0.00179823 |
| 0 | 0 | 3 | 4 | 1 | 4 | 0 | 2 | 0 | 0.34620086 | 0.58558839 |
| 10 | 25 | 27 | 10 | 5 | 56 | 2 | 8 | 12 | 0.0846992 | 0.274096 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 2 | 0 | 3 | 0 | 0.3028855 | 0.5346388 |
| 1899 | 1393 | 5742 | 18555 | 14289 | 11276 | 351 | 421 | 1427 | 0.36172239 | 0.60743291 |
| 116 | 451 | 789 | 256 | 168 | 163 | 108 | 80 | 115 | 7.18E-08 | 8.37E-06 |
| 235 | 324 | 1764 | 1805 | 933 | 1781 | 376 | 1008 | 1819 | 0.6075246 | 0.8380582 |
| 19 | 27 | 173 | 109 | 104 | 415 | 211 | 83 | 253 | 0.17704617 | 0.41879956 |
| 0 | 0 | 1 | 2 | 0 | 2 | 0 | 0 | 0 | 0.45250807 | 0.70097152 |
| 1 | 0 | 2 | 3 | 2 | 6 | 2 | 2 | 1 | 0.53899964 | 0.77470615 |
| 0 | 0 | 1 | 2 | 17 | 1 | 2 | 0 | 0 | 0.05307151 | 0.19473482 |
| 1 | 2 | 11 | 16 | 19 | 18 | 1 | 5 | 11 | 0.52385319 | 0.76330157 |
| 480 | 553 | 3702 | 1682 | 3010 | 7074 | 1329 | 1094 | 411 | 0.33031344 | 0.5656407 |
| 3 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0.0882563 | 0.2762803 |
| 24 | 40 | 20 | 19 | 22 | 43 | 15 | 12 | 3 | 0.00015176 | 0.00336753 |
| 0 | 6 | 31 | 26 | 10 | 111 | 140 | 11 | 10 | 0.03616385 | 0.14403722 |
| 9 | 6 | 5 | 14 | 17 | 11 | 3 | 2 | 3 | 0.04947959 | 0.1844599 |
| 0 | 0 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0.45250807 | 0.70097152 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.43E-05 | 0.00102893 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA |
| 6344 | 7499 | 17092 | 12068 | 12260 | 16835 | 1424 | 1878 | 3519 | 0.00837615 | 0.05346969 |
| 3 | 5 | 41 | 50 | 42 | 58 | 13 | 17 | 6 | 0.46723884 | 0.70862854 |
| 886 | 1257 | 10368 | 5700 | 7714 | 8219 | 8225 | 3881 | 3857 | 0.29840218 | 0.5346388 |
| 13 | 4 | 11 | 6 | 6 | 11 | 1 | 2 | 7 | 0.04232522 | 0.16204427 |
| 1 | 2 | 12 | 10 | 5 | 12 | 8 | 25 | 11 | 0.14008223 | 0.35381202 |
| 222 | 120 | 887 | 1046 | 742 | 1875 | 257 | 324 | 306 | 0.99841691 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 42 | 36 | 208 | 86 | 200 | 437 | 78 | 228 | 90 | 0.5034922 | 0.74841264 |
| 3 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0.00134621 | 0.01548974 |
| 397 | 682 | 3034 | 2377 | 2296 | 10230 | 687 | 1291 | 1371 | 0.25441002 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0.71923557 | 0.90275157 |
| 0 | 1 | 3 | 2 | 2 | 2 | 0 | 1 | 0 | 1 | 1 |
| 1 | 3 | 5 | 4 | 0 | 9 | 4 | 1 | 0 | 0.07380484 | 0.2538233 |
| 480 | 553 | 3702 | 1681 | 3010 | 7074 | 1329 | 1094 | 411 | 0.33029826 | 0.5656407 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1045419 | 0.28826346 |
| 1 | 0 | 9 | 8 | 9 | 16 | 3 | 7 | 3 | 0.04214093 | 0.16204427 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr1 | 207975196 | 207975284 | hsa-mir-29c | - | 3150 | 171 | 167 | 1994 | 567 |
| chr6 | 2854264 | 2854341 | hsa-mir-4645 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 44155703 | 44155802 | hsa-mir-138-1 | + | 20 | 21 | 20 | 10 | 12 |
| chr19 | 52785049 | 52785146 | hsa-mir-643 | + | 0 | 0 | 0 | 0 | 0 |
| chr7 | 130135951 | 130136045 | hsa-mir-335 | + | 1611 | 68 | 36 | 369 | 343 |
| chrX | 113949649 | 113949761 | hsa-mir-1298 | + | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101506772 | 101506872 | hsa-mir-376b | + | 89 | 39 | 0 | 35 | 81 |
| chr1 | 228284963 | 228285042 | hsa-mir-3620 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 41222955 | 41223044 | hsa-mir-30c-1 | + | 829 | 67 | 104 | 512 | 576 |
| chr1 | 117637264 | 117637350 | hsa-mir-942 | + | 3 | 2 | 6 | 5 | 3 |
| chr19 | 54211988 | 54212089 | hsa-mir-518c | + | 2 | 0 | 1 | 2 | 0 |
| chr6 | 18572014 | 18572111 | hsa-mir-548a-1 | + | 2 | 0 | 0 | 3 | 1 |
| chr20 | 61918159 | 61918218 | hsa-mir-4326 | + | 3 | 0 | 1 | 1 | 0 |
| chrX | 146312501 | 146312595 | hsa-mir-507 | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 110271152 | 110271241 | hsa-mir-4497 | + | 1 | 0 | 1 | 0 | 0 |
| chr4 | 2251803 | 2251883 | hsa-mir-4800 | - | 1 | 0 | 0 | 0 | 0 |
| chr20 | 45795608 | 45795700 | hsa-mir-3616 | + | 5 | 2 | 0 | 0 | 1 |
| chr11 | 64658608 | 64658718 | hsa-mir-192 | - | 264 | 19 | 13 | 70 | 66 |
| chr4 | 110409853 | 110409951 | hsa-mir-576 | + | 32 | 39 | 17 | 20 | 14 |
| chr12 | 49165757 | 49165820 | hsa-mir-4701 | - | 1 | 0 | 0 | 0 | 0 |
| chr8 | 101036209 | 101036312 | hsa-mir-1273a | - | 4 | 2 | 1 | 0 | 0 |
| chr1 | 33797993 | 33798093 | hsa-mir-3605 | - | 9 | 3 | 7 | 2 | 4 |
| chr14 | 101491900 | 101491988 | hsa-mir-1197 | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 114059369 | 114059448 | hsa-mir-6715a | + | 1 | 2 | 0 | 0 | 0 |
| chr19 | 54189722 | 54189809 | hsa-mir-519c | + | 1 | 0 | 0 | 0 | 0 |
| chr8 | 12584745 | 12584808 | hsa-mir-3926-2 | + | 3 | 1 | 1 | 0 | 0 |
| chr2 | 185243701 | 185243771 | hsa-mir-548ae-1 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 58999431 | 58999529 | hsa-mir-582 | - | 11 | 1 | 0 | 12 | 9 |
| chr19 | 54230175 | 54230240 | hsa-mir-526a-2 | + | 0 | 0 | 0 | 0 | 0 |
| chr17 | 72744751 | 72744838 | hsa-mir-3615 | + | 26 | 8 | 13 | 3 | 2 |
| chr2 | 219866366 | 219866430 | hsa-mir-375 | - | 308 | 77 | 72 | 13 | 81 |
| chr7 | 91833328 | 91833412 | hsa-mir-1285-1 | - | 14 | 5 | 4 | 2 | 1 |
| chr15 | 93447628 | 93447705 | hsa-mir-3175 | + | 0 | 0 | 0 | 0 | 0 |
| chr6 | 30858659 | 30858749 | hsa-mir-4640 | + | 1 | 0 | 1 | 0 | 1 |
| chr12 | 112475402 | 112475519 | hsa-mir-3657 | - | 0 | 0 | 0 | 0 | 0 |
| chr2 | 70480049 | 70480137 | hsa-mir-1285-2 | - | 7 | 2 | 3 | 2 | 0 |
| chrX | 133303407 | 133303482 | hsa-mir-363 | - | 159 | 14 | 31 | 270 | 44 |
| chr7 | 98479272 | 98479352 | hsa-mir-3609 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 567704 | 567793 | hsa-mir-6723 | - | 2 | 0 | 0 | 1 | 0 |
| chr18 | 19405658 | 19405746 | hsa-mir-133a-1 | - | 10 | 4 | 7 | 4 | 2 |
| chr5 | 175794948 | 175795034 | hsa-mir-1271 | + | 38 | 8 | 9 | 2 | 17 |
| chr4 | 160049956 | 160050043 | hsa-mir-3688-2 | + | 2 | 0 | 0 | 1 | 0 |
| chr19 | 54245765 | 54245853 | hsa-mir-520h | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 127508308 | 127508391 | hsa-mir-4484 | + | 2 | 0 | 3 | 0 | 1 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 637 | 563 | 15849 | 8508 | 2283 | 8664 | 3621 | 1293 | 3201 | 0.25525845 | 0.51361322 |
| 0 | 0 | 3 | 1 | 1 | 4 | 0 | 0 | 0 | 0.06030698 | 0.21632095 |
| 49 | 5 | 13 | 32 | 6 | 75 | 7 | 3 | 9 | 0.14672717 | 0.36466592 |
| 0 | 0 | 1 | 3 | 3 | 8 | 3 | 2 | 1 | 0.01472814 | 0.0758377 |
| 245 | 896 | 596 | 349 | 202 | 192 | 54 | 204 | 308 | 0.00329945 | 0.0277035 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0.71923557 | 0.90275157 |
| 199 | 80 | 55 | 51 | 40 | 44 | 11 | 14 | 20 | 0.00582937 | 0.04024425 |
| 0 | 0 | 1 | 3 | 0 | 1 | 1 | 1 | 1 | 0.10364886 | 0.28750219 |
| 554 | 408 | 3936 | 2172 | 1326 | 2296 | 897 | 1004 | 552 | 0.67531324 | 0.90275157 |
| 1 | 1 | 7 | 16 | 8 | 10 | 6 | 4 | 1 | 0.91402591 | 1 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 7 | 1 | 0.84845436 | 0.99717461 |
| 0 | 1 | 4 | 2 | 1 | 9 | 8 | 1 | 4 | 0.60310163 | 0.83520167 |
| 0 | 0 | 2 | 15 | 0 | 0 | 13 | 0 | 0 | 0.32975901 | 0.5656407 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0.71923557 | 0.90275157 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.28374581 | 0.52015675 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 2 | 0 | 0.06098566 | 0.21694137 |
| 60 | 46 | 285 | 257 | 232 | 732 | 633 | 253 | 647 | 0.30767953 | 0.53599499 |
| 13 | 12 | 33 | 37 | 13 | 61 | 57 | 41 | 19 | 0.55446857 | 0.79073 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 3 | 9 | 7 | 6 | 9 | 4 | 4 | 0 | 0.74379945 | 0.91695911 |
| 2 | 1 | 5 | 6 | 7 | 7 | 5 | 2 | 3 | 0.23966663 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 0.68634275 | 0.90275157 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02809079 | 0.11793068 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0.76516442 | 0.92898156 |
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0.05212685 | 0.19278662 |
| 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0.68634275 | 0.90275157 |
| 10 | 8 | 25 | 51 | 14 | 84 | 25 | 97 | 176 | 0.08424785 | 0.27358536 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | NA | NA |
| 3 | 3 | 79 | 25 | 12 | 31 | 0 | 3 | 1 | 0.75772881 | 0.92586501 |
| 68 | 117 | 11449 | 443 | 312 | 9400 | 2208 | 374 | 2460 | 0.00025614 | 0.00484292 |
| 1 | 2 | 11 | 8 | 8 | 18 | 6 | 11 | 6 | 0.73024426 | 0.9074502 |
| 0 | 0 | 1 | 0 | 2 | 2 | 0 | 0 | 0 | 0.45250807 | 0.70097152 |
| 0 | 0 | 1 | 1 | 0 | 2 | 2 | 1 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 | 0 | 2 | 1 | 1 | 0 | 0.3028855 | 0.5346388 |
| 1 | 1 | 5 | 3 | 2 | 10 | 2 | 4 | 2 | 0.43843035 | 0.69611088 |
| 35 | 59 | 184 | 208 | 127 | 1090 | 196 | 244 | 2337 | 0.13198122 | 0.34021425 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.71923557 | 0.90275157 |
| 0 | 1 | 6 | 4 | 1 | 2 | 4 | 0 | 0 | 0.7420326 | 0.91695911 |
| 3 | 6 | 12 | 8 | 13 | 3 | 0 | 2 | 0 | 0.14359115 | 0.35878538 |
| 16 | 11 | 38 | 15 | 37 | 16 | 3 | 24 | 8 | 0.22871552 | 0.5063251 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00578835 | 0.04024425 |
| 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0.71923557 | 0.90275157 |
| 0 | 0 | 3 | 5 | 2 | 1 | 2 | 0 | 3 | 1 | 1 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr3 | 52302293 | 52302377 | hsa-let-7g | - | 3726 | 183 | 391 | 4501 | 4343 |
| chr10 | 70519074 | 70519171 | hsa-mir-1254-1 | + | 5 | 3 | 3 | 1 | 1 |
| chrX | 45606420 | 45606530 | hsa-mir-222 | - | 1402 | 206 | 174 | 298 | 340 |
| chr19 | 54257271 | 54257356 | hsa-mir-527 | + | 1 | 0 | 0 | 0 | 0 |
| chr13 | 41384901 | 41384997 | hsa-mir-621 | + | 1 | 0 | 0 | 0 | 1 |
| chr11 | 59976543 | 59976629 | hsa-mir-6503 | - | 11 | 3 | 4 | 4 | 1 |
| chr16 | 14397823 | 14397906 | hsa-mir-193b | + | 2798 | 1015 | 1330 | 233 | 349 |
| chr15 | 62635227 | 62635337 | hsa-mir-6085 | + | 1 | 0 | 0 | 0 | 0 |
| chr19 | 54234259 | 54234344 | hsa-mir-518a-1 | + | 1 | 0 | 1 | 1 | 0 |
| chr1 | 38554902 | 38555001 | hsa-mir-3659 | + | 0 | 0 | 0 | 0 | 0 |
| chr9 | 97572243 | 97572339 | hsa-mir-2278 | + | 1 | 0 | 1 | 0 | 0 |
| chr1 | 65524116 | 65524191 | hsa-mir-101-1 | - | 2969 | 267 | 185 | 5239 | 1660 |
| chr17 | 46114526 | 46114613 | hsa-mir-152 | - | 1349 | 86 | 135 | 373 | 616 |
| chr10 | 103361173 | 103361254 | hsa-mir-3158-2 | - | 12 | 8 | 4 | 1 | 7 |
| chr17 | 78072626 | 78072676 | hsa-mir-1268b | + | 9 | 12 | 18 | 0 | 3 |
| chr14 | 101509313 | 101509399 | hsa-mir-1185-1 | + | 28 | 0 | 0 | 4 | 12 |
| chr19 | 13947100 | 13947173 | hsa-mir-24-2 | - | 8749 | 2354 | 2156 | 1035 | 2408 |
| chr20 | 60907542 | 60907613 | hsa-mir-4758 | - | 0 | 0 | 0 | 0 | 0 |
| chr7 | 5535449 | 5535548 | hsa-mir-589 | - | 22 | 17 | 9 | 7 | 1 |
| chr12 | 54385521 | 54385631 | hsa-mir-196a-2 | + | 75 | 12 | 63 | 33 | 50 |
| chr20 | 61809851 | 61809938 | hsa-mir-124-3 | + | 17 | 0 | 0 | 4 | 1 |
| chr5 | 140027428 | 140027511 | hsa-mir-3655 | + | 1 | 0 | 0 | 0 | 0 |
| chr10 | 103361173 | 103361254 | hsa-mir-3158-1 | + | 12 | 8 | 4 | 1 | 7 |
| chr1 | 198828001 | 198828111 | hsa-mir-181b-1 | - | 121 | 36 | 34 | 37 | 116 |
| chr1 | 37966535 | 37966595 | hsa-mir-5581 | - | 0 | 0 | 0 | 0 | 0 |
| chr17 | 12820584 | 12820659 | hsa-mir-1269b | - | 0 | 0 | 0 | 0 | 0 |
| chr22 | 49176106 | 49176165 | hsa-mir-4535 | + | 1 | 0 | 0 | 0 | 0 |
| chr7 | 30329409 | 30329506 | hsa-mir-550a-1 | + | 4 | 4 | 3 | 1 | 3 |
| chr19 | 54242586 | 54242673 | hsa-mir-518a-2 | + | 2 | 0 | 1 | 1 | 0 |
| chr7 | 147075108 | 147075213 | hsa-mir-548f-4 | - | 0 | 0 | 0 | 0 | 0 |
| chr17 | 79099682 | 79099749 | hsa-mir-338 | - | 42 | 7 | 11 | 17 | 8 |
| chr9 | 68415307 | 68415388 | hsa-mir-4477b | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 10928101 | 10928172 | hsa-mir-199a-1 | - | 5218 | 994 | 1545 | 3191 | 3740 |
| chr21 | 34922967 | 34923034 | hsa-mir-6501 | + | 0 | 0 | 0 | 0 | 0 |
| chr19 | 52195864 | 52195934 | hsa-mir-99b | + | 1286 | 203 | 291 | 293 | 569 |
| chr8 | 144895126 | 144895212 | hsa-mir-937 | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 49048226 | 49048313 | hsa-mir-1291 | - | 2 | 4 | 0 | 0 | 4 |
| chr5 | 72174417 | 72174490 | hsa-mir-4804 | + | 1 | 0 | 0 | 0 | 1 |
| chr12 | 62997465 | 62997549 | hsa-let-7i | + | 6515 | 799 | 2428 | 4125 | 5739 |
| chr9 | 96357122 | 96357176 | hsa-mir-548au | + | 2 | 0 | 0 | 3 | 1 |
| chr4 | 115577914 | 115578010 | hsa-mir-577 | + | 0 | 0 | 0 | 0 | 0 |
| chr9 | 123007256 | 123007328 | hsa-mir-147a | - | 0 | 0 | 0 | 0 | 0 |
| chr1 | 98511625 | 98511727 | hsa-mir-137 | - | 0 | 0 | 0 | 0 | 0 |
| chr2 | 47604813 | 47604909 | hsa-mir-559 | + | 2 | 1 | 0 | 5 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 2422 | 5493 | 14007 | 13920 | 11639 | 9386 | 6500 | 4753 | 3378 | 0.9779933 | 1 |
| 2 | 0 | 5 | 5 | 1 | 6 | 1 | 4 | 2 | 0.40118954 | 0.65254564 |
| 723 | 320 | 1144 | 1571 | 1170 | 1279 | 764 | 1039 | 890 | 0.65571792 | 0.89085875 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 0 | 0.13738238 | 0.34983709 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 6 | 1 | 8 | 8 | 3 | 9 | 6 | 1 | 5 | 0.25324401 | 0.51361322 |
| 241 | 151 | 2498 | 476 | 1535 | 1348 | 631 | 221 | 431 | 0.12339371 | 0.32123726 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.34426564 | 0.58337378 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 1 | 1 | 1 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01835236 | 0.08961841 |
| 1391 | 3052 | 14592 | 4351 | 3291 | 4235 | 13260 | 1337 | 2078 | 0.9809745 | 1 |
| 492 | 715 | 1187 | 1198 | 2740 | 1485 | 118 | 955 | 1005 | 0.6241024 | 0.85791067 |
| 0 | 5 | 6 | 4 | 5 | 12 | 2 | 1 | 6 | 0.11170997 | 0.30003945 |
| 3 | 4 | 8 | 11 | 3 | 8 | 0 | 2 | 6 | 0.08012372 | 0.26480604 |
| 8 | 11 | 36 | 32 | 12 | 24 | 6 | 6 | 16 | 0.45530154 | 0.70097152 |
| 1431 | 2001 | 9312 | 8923 | 14526 | 19617 | 3785 | 6414 | 4481 | 0.84215899 | 0.99102548 |
| 0 | 0 | 1 | 1 | 3 | 2 | 0 | 0 | 0 | 0.27846304 | 0.51493562 |
| 7 | 5 | 17 | 29 | 28 | 25 | 8 | 7 | 16 | 0.57911963 | 0.81581943 |
| 26 | 88 | 3048 | 526 | 2305 | 3329 | 48 | 203 | 987 | 0.00153566 | 0.01684238 |
| 0 | 0 | 6 | 0 | 2 | 1 | 0 | 0 | 1 | 0.00211681 | 0.02098795 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 5 | 6 | 4 | 5 | 12 | 2 | 1 | 6 | 0.11170997 | 0.30003945 |
| 56 | 96 | 339 | 177 | 382 | 303 | 70 | 609 | 375 | 0.48844377 | 0.73305892 |
| 0 | 0 | 1 | 0 | 1 | 3 | 4 | 1 | 0 | 0.07546743 | 0.25419227 |
| 0 | 0 | 3 | 0 | 1 | 1 | 19 | 0 | 1 | 0.00671994 | 0.04512258 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1 | 2 | 6 | 3 | 6 | 8 | 3 | 5 | 2 | 0.5557169 | 0.79073 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 1 | 0.73932649 | 0.91507608 |
| 0 | 0 | 2 | 2 | 2 | 0 | 2 | 4 | 0 | 0.04995035 | 0.18547302 |
| 22 | 9 | 42 | 147 | 34 | 121 | 12 | 27 | 67 | 0.75797747 | 0.92586501 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 5195 | 3015 | 15510 | 13363 | 16017 | 14951 | 2613 | 3860 | 6636 | 0.90257111 | 1 |
| 0 | 0 | 3 | 0 | 1 | 1 | 0 | 3 | 0 | 0.10133222 | 0.28400472 |
| 471 | 471 | 2438 | 2010 | 2900 | 3625 | 781 | 2257 | 932 | 0.58518669 | 0.81890991 |
| 0 | 0 | 1 | 4 | 6 | 7 | 3 | 0 | 3 | 0.01178326 | 0.06773004 |
| 1 | 0 | 1 | 0 | 0 | 3 | 0 | 1 | 0 | 0.03291175 | 0.134534 |
| 0 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0.77031617 | 0.93359515 |
| 4236 | 5151 | 13561 | 15812 | 16536 | 13157 | 6033 | 8231 | 3494 | 0.83408671 | 0.98753789 |
| 1 | 1 | 5 | 8 | 4 | 5 | 3 | 3 | 5 | 0.63537502 | 0.86955877 |
| 0 | 0 | 1 | 4 | 0 | 0 | 1 | 4 | 0 | 0.07445672 | 0.25390681 |
| 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0.68634275 | 0.90275157 |
| 0 | 0 | 1 | 44 | 82 | 2 | 1 | 2 | 5 | 5.20E-05 | 0.00179823 |
| 0 | 1 | 7 | 4 | 0 | 1 | 5 | 5 | 1 | 0.91382477 | 1 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr4 | 38869652 | 38869748 | hsa-mir-574 | + | 874 | 409 | 780 | 159 | 204 |
| chr19 | 54261485 | 54261572 | hsa-mir-1283-2 | + | 1 | 0 | 0 | 0 | 0 |
| chr8 | 145619363 | 145619445 | hsa-mir-939 | - | 6 | 0 | 3 | 0 | 0 |
| chr15 | 21038123 | 21038198 | hsa-mir-3118-4 | + | 0 | 0 | 0 | 0 | 0 |
| chrX | 73507120 | 73507192 | hsa-mir-374a | - | 23 | 8 | 11 | 388 | 45 |
| chr10 | 74480786 | 74480858 | hsa-mir-4676 | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 115933863 | 115933938 | hsa-mir-2110 | - | 47 | 26 | 42 | 4 | 3 |
| chr11 | 65403780 | 65403840 | hsa-mir-4690 | + | 1 | 0 | 1 | 0 | 0 |
| chr3 | 151283663 | 151283783 | hsa-mir-5186 | - | 8 | 2 | 0 | 0 | 0 |
| chr22 | 31556047 | 31556105 | hsa-mir-3928 | - | 7 | 3 | 1 | 1 | 1 |
| chr10 | 43493010 | 43493129 | hsa-mir-5100 | + | 33 | 23 | 15 | 3 | 3 |
| chr14 | 101495970 | 101496051 | hsa-mir-494 | + | 126 | 11 | 4 | 7 | 26 |
| chr11 | 122022936 | 122023016 | hsa-mir-100 | - | 2757 | 866 | 1139 | 1212 | 1498 |
| chr17 | 39673415 | 39673469 | hsa-mir-6510 | - | 41 | 11 | 2 | 2 | 8 |
| chr5 | 1063010 | 1063089 | hsa-mir-4635 | - | 1 | 1 | 0 | 0 | 0 |
| chr3 | 160122532 | 160122613 | hsa-mir-16-2 | + | 214 | 70 | 105 | 1348 | 94 |
| chr9 | 73424890 | 73425000 | hsa-mir-204 | - | 52 | 22 | 10 | 16 | 14 |
| chr14 | 101490130 | 101490193 | hsa-mir-299 | + | 96 | 11 | 12 | 14 | 24 |
| chr19 | 54178964 | 54179051 | hsa-mir-520e | + | 1 | 0 | 0 | 0 | 0 |
| chr18 | 56118305 | 56118390 | hsa-mir-122 | + | 6 | 0 | 5 | 0 | 3 |
| chr11 | 133768398 | 133768476 | hsa-mir-4697 | - | 2 | 0 | 0 | 1 | 0 |
| chr1 | 67094122 | 67094200 | hsa-mir-3117 | + | 1 | 0 | 0 | 1 | 0 |
| chr9 | 91360750 | 91360820 | hsa-mir-4289 | - | 0 | 0 | 0 | 0 | 0 |
| chrX | 49775279 | 49775358 | hsa-mir-500b | + | 14 | 3 | 6 | 3 | 3 |
| chr14 | 65511405 | 65511487 | hsa-mir-4706 | + | 0 | 0 | 0 | 0 | 0 |
| chrX | 133304070 | 133304141 | hsa-mir-18b | - | 3 | 0 | 2 | 8 | 0 |
| chr5 | 167987908 | 167987970 | hsa-mir-103b-1 | + | 2854 | 436 | 323 | 1242 | 1996 |
| chr2 | 97464014 | 97464090 | hsa-mir-3127 | + | 0 | 0 | 0 | 0 | 0 |
| chr6 | 44221942 | 44222022 | hsa-mir-4647 | - | 1 | 0 | 0 | 0 | 0 |
| chr15 | 83736086 | 83736167 | hsa-mir-4515 | + | 1 | 0 | 0 | 0 | 0 |
| chr3 | 189547710 | 189547798 | hsa-mir-944 | + | 0 | 0 | 0 | 0 | 0 |
| chr6 | 52009146 | 52009232 | hsa-mir-206 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 134263728 | 134263802 | hsa-mir-4461 | + | 1 | 0 | 0 | 0 | 0 |
| chr15 | 36219056 | 36219124 | hsa-mir-4510 | + | 62 | 6 | 18 | 7 | 23 |
| chr11 | 94199660 | 94199746 | hsa-mir-548l | - | 1 | 1 | 0 | 0 | 1 |
| chr19 | 54182256 | 54182339 | hsa-mir-515-1 | + | 3 | 0 | 1 | 1 | 1 |
| chr15 | 63116155 | 63116240 | hsa-mir-190a | + | 180 | 8 | 7 | 251 | 122 |
| chr7 | 129414531 | 129414609 | hsa-mir-96 | - | 164 | 7 | 2 | 36 | 24 |
| chr4 | 77496703 | 77496779 | hsa-mir-548ah | + | 3 | 1 | 0 | 1 | 0 |
| chr16 | 16403735 | 16403823 | hsa-mir-3180-2 | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 105807846 | 105807944 | hsa-mir-936 | - | 0 | 0 | 0 | 0 | 0 |
| chr13 | 115039302 | 115039383 | hsa-mir-4502 | + | 3 | 1 | 5 | 1 | 0 |
| chr5 | 168690604 | 168690698 | hsa-mir-585 | - | 1 | 0 | 0 | 3 | 0 |
| chr19 | 54265597 | 54265684 | hsa-mir-519a-2 | + | 2 | 0 | 1 | 2 | 0 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 204 | 114 | 614 | 473 | 885 | 989 | 51 | 219 | 135 | 0.19104775 | 0.44403119 |
| 0 | 0 | 1 | 2 | 4 | 1 | 0 | 136 | 1 | 7.60E-05 | 0.00208452 |
| 0 | 0 | 3 | 1 | 4 | 4 | 0 | 1 | 1 | 0.36012063 | 0.60693025 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 39 | 167 | 2960 | 1846 | 265 | 2116 | 4514 | 1977 | 1081 | 0.00104257 | 0.01245739 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 8 | 6 | 6 | 30 | 27 | 53 | 6 | 19 | 2 | 0.15294617 | 0.37710538 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7.18E-05 | 0.00205793 |
| 0 | 0 | 5 | 0 | 4 | 3 | 4 | 0 | 2 | 0.30054425 | 0.5346388 |
| 7 | 3 | 17 | 12 | 19 | 21 | 1 | 27 | 4 | 0.20659011 | 0.46620335 |
| 15 | 19 | 23 | 10 | 63 | 14 | 5 | 8 | 22 | 0.01368558 | 0.07288551 |
| 1946 | 1340 | 1662 | 1437 | 1175 | 1065 | 411 | 891 | 1000 | 0.01184549 | 0.06773004 |
| 6 | 20 | 3 | 6 | 3 | 0 | 0 | 3 | 0 | 3.17E-05 | 0.00118002 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0199196 | 0.09520546 |
| 130 | 475 | 6171 | 4192 | 1158 | 3343 | 1099 | 2746 | 2924 | 0.06170888 | 0.21867936 |
| 42 | 30 | 24 | 11 | 2 | 2 | 1 | 1 | 3 | 0.00021629 | 0.00447959 |
| 40 | 62 | 91 | 60 | 61 | 59 | 13 | 17 | 44 | 0.1876568 | 0.44165691 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.11E-05 | 0.00093731 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0135387 | 0.07269882 |
| 2 | 1 | 7 | 6 | 21 | 6 | 1 | 13 | 8 | 0.07127018 | 0.2464484 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2 | 1 | 20 | 28 | 11 | 28 | 22 | 23 | 11 | 0.51952004 | 0.76130924 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 3 | 19 | 13 | 7 | 121 | 3 | 30 | 67 | 0.02498968 | 0.11038095 |
| 847 | 1210 | 9905 | 5395 | 7506 | 7789 | 8473 | 3768 | 3425 | 0.30191648 | 0.5346388 |
| 0 | 0 | 1 | 1 | 2 | 5 | 1 | 0 | 2 | 0.0462724 | 0.17389465 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.34426564 | 0.58337378 |
| 0 | 0 | 4 | 0 | 2 | 0 | 3 | 0 | 5 | 0.02526028 | 0.11104989 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0.71923557 | 0.90275157 |
| 0 | 0 | 1 | 0 | 1 | 1 | 3 | 0 | 0 | 0.72065147 | 0.90275157 |
| 11 | 60 | 13 | 3 | 13 | 4 | 3 | 27 | 20 | 0.00384351 | 0.03088064 |
| 0 | 0 | 3 | 1 | 0 | 5 | 1 | 0 | 2 | 0.83478372 | 0.98753789 |
| 4 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 0 | 0.0277968 | 0.11775735 |
| 101 | 98 | 338 | 267 | 221 | 260 | 277 | 58 | 47 | 0.49647659 | 0.7427226 |
| 26 | 52 | 2172 | 1517 | 1632 | 3334 | 384 | 713 | 938 | 0.00069314 | 0.00958327 |
| 1 | 0 | 4 | 3 | 2 | 2 | 2 | 9 | 3 | 0.64064147 | 0.8741989 |
| 0 | 0 | 1 | 0 | 0 | 2 | 0 | 3 | 0 | 0.3028855 | 0.5346388 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00044361 | 0.00717452 |
| 1 | 1 | 7 | 13 | 3 | 4 | 6 | 0 | 0 | 0.41618558 | 0.67150147 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00147341 | 0.01660629 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chrX | 146353852 | 146353926 | hsa-mir-510 | - | 1 | 0 | 0 | 0 | 0 |
| chr17 | 6920933 | 6921020 | hsa-mir-195 | - | 768 | 31 | 25 | 560 | 176 |
| chr17 | 27188672 | 27188748 | hsa-mir-4732 | - | 2 | 0 | 2 | 0 | 0 |
| chr1 | 156390132 | 156390221 | hsa-mir-9-1 | - | 34 | 1 | 0 | 16 | 34 |
| chrX | 146307343 | 146307470 | hsa-mir-513a-2 | - | 0 | 0 | 0 | 0 | 0 |
| chr19 | 54485560 | 54485651 | hsa-mir-935 | + | 2 | 1 | 0 | 0 | 1 |
| chr1 | 24255559 | 24255637 | hsa-mir-378f | + | 88 | 68 | 76 | 12 | 39 |
| chr16 | 820182 | 820277 | hsa-mir-662 | + | 0 | 0 | 0 | 0 | 0 |
| chr13 | 92003144 | 92003226 | hsa-mir-19a | + | 894 | 19 | 10 | 710 | 281 |
| chr18 | 13611112 | 13611199 | hsa-mir-4526 | + | 2 | 0 | 0 | 0 | 0 |
| chr4 | 111781737 | 111781803 | hsa-mir-297 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 20179056 | 20179133 | hsa-mir-3135a | + | 2 | 1 | 1 | 0 | 0 |
| chr8 | 81153623 | 81153708 | hsa-mir-5708 | + | 1 | 1 | 0 | 0 | 0 |
| chr20 | 59139619 | 59139683 | hsa-mir-548ag-2 | + | 0 | 0 | 0 | 0 | 0 |
| chrX | 28513671 | 28513780 | hsa-mir-6134 | - | 1 | 0 | 0 | 0 | 0 |
| chr1 | 68649200 | 68649293 | hsa-mir-1262 | - | 3 | 2 | 1 | 0 | 1 |
| chr22 | 22007269 | 22007347 | hsa-mir-301b | + | 6 | 0 | 0 | 6 | 3 |
| chr6 | 170639848 | 170639932 | hsa-mir-4644 | + | 1 | 0 | 0 | 0 | 1 |
| chr16 | 2320713 | 2320773 | hsa-mir-3677 | + | 0 | 0 | 0 | 0 | 0 |
| chr5 | 174178736 | 174178790 | hsa-mir-4634 | + | 1 | 1 | 4 | 0 | 0 |
| chr14 | 100575755 | 100575851 | hsa-mir-151b | - | 74 | 8 | 10 | 20 | 18 |
| chr4 | 67142541 | 67142646 | hsa-mir-1269a | + | 1 | 0 | 0 | 0 | 0 |
| chr2 | 177465707 | 177465780 | hsa-mir-1246 | - | 8 | 3 | 2 | 2 | 3 |
| chr1 | 172113674 | 172113784 | hsa-mir-199a-2 | - | 5245 | 995 | 1547 | 3210 | 3759 |
| chrX | 76225828 | 76225926 | hsa-mir-325 | - | 2 | 0 | 0 | 0 | 0 |
| chr9 | 4850296 | 4850375 | hsa-mir-101-2 | + | 2938 | 121 | 89 | 5344 | 1719 |
| chr1 | 143163749 | 143163822 | hsa-mir-3118-2 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 176998498 | 176998581 | hsa-mir-488 | - | 124 | 0 | 0 | 5 | 14 |
| chr13 | 102619991 | 102620096 | hsa-mir-2681 | - | 1 | 0 | 0 | 0 | 0 |
| chr2 | 220158832 | 220158922 | hsa-mir-153-1 | - | 42 | 8 | 6 | 7 | 14 |
| chr2 | 75317938 | 75318041 | hsa-mir-5000 | + | 0 | 0 | 0 | 0 | 0 |
| chr9 | 135821093 | 135821158 | hsa-mir-548aw | + | 9 | 1 | 4 | 1 | 4 |
| chr2 | 219144847 | 219144911 | hsa-mir-6513 | - | 2 | 0 | 0 | 2 | 0 |
| chr17 | 7991373 | 7991465 | hsa-mir-4314 | + | 0 | 0 | 0 | 0 | 0 |
| chr4 | 79741905 | 79741975 | hsa-mir-5096 | + | 15 | 4 | 5 | 1 | 1 |
| chr19 | 54172410 | 54172508 | hsa-mir-512-2 | + | 5 | 0 | 0 | 1 | 1 |
| chr3 | 103242876 | 103242960 | hsa-mir-548ab | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 116866056 | 116866123 | hsa-mir-4472-2 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 10436172 | 10436246 | hsa-mir-885 | - | 1 | 0 | 2 | 0 | 0 |
| chr14 | 101341369 | 101341441 | hsa-mir-665 | + | 79 | 31 | 14 | 10 | 10 |
| chr4 | 9557788 | 9557937 | hsa-mir-548i-2 | - | 0 | 0 | 0 | 0 | 0 |
| chr8 | 27559189 | 27559284 | hsa-mir-3622b | - | 3 | 4 | 0 | 1 | 0 |
| chr14 | 101533060 | 101533138 | hsa-mir-656 | + | 5 | 0 | 0 | 3 | 4 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 163 | 623 | 3176 | 1354 | 607 | 1258 | 68 | 468 | 697 | 0.99990033 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01042004 | 0.06157765 |
| 79 | 37 | 43 | 18 | 39 | 264 | 27 | 9 | 22 | 0.50175708 | 0.74822015 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 1 | 0 | 10 | 3 | 5 | 39 | 0 | 6 | 0 | 0.10884583 | 0.2966208 |
| 7 | 28 | 9 | 4 | 7 | 27 | 12 | 8 | 8 | 0.00044233 | 0.00717452 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 212 | 493 | 4546 | 3996 | 1042 | 1301 | 1655 | 1316 | 1008 | 0.29332039 | 0.53082447 |
| 0 | 0 | 13 | 1 | 5 | 8 | 0 | 0 | 1 | 0.13800532 | 0.35046582 |
| 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0.71923558 | 0.90275157 |
| 0 | 0 | 3 | 1 | 2 | 6 | 0 | 2 | 3 | 0.74917106 | 0.92101749 |
| 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0.30539054 | 0.53593028 |
| 0 | 0 | 5 | 2 | 0 | 4 | 0 | 0 | 0 | 0.03730011 | 0.14730381 |
| 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0.71657056 | 0.90275157 |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.00051856 | 0.00805502 |
| 4 | 1 | 22 | 39 | 28 | 86 | 28 | 65 | 24 | 0.02601406 | 0.11329487 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 0 | 0 | 2 | 0 | 3 | 8 | 0 | 0 | 1 | 0.03110358 | 0.12998785 |
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0.11892936 | 0.3139037 |
| 9 | 33 | 168 | 76 | 128 | 218 | 21 | 37 | 78 | 0.74552729 | 0.91787508 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 1 | 6 | 12 | 20 | 29 | 33 | 6 | 17 | 2 | 0.52141922 | 0.7628928 |
| 5211 | 3057 | 15646 | 13443 | 16092 | 15039 | 2636 | 3874 | 6687 | 0.90093138 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.09486019 | 0.2762803 |
| 1427 | 3118 | 14927 | 4417 | 3320 | 4302 | 13428 | 1369 | 2124 | 0.95833044 | 1 |
| 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 13 | 15 | 12 | 13 | 0 | 0 | 68 | 0 | 0 | 0.00380961 | 0.03087437 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.34426564 | 0.58337378 |
| 9 | 23 | 27 | 9 | 8 | 55 | 2 | 8 | 14 | 0.10875373 | 0.2966208 |
| 0 | 0 | 2 | 2 | 1 | 1 | 0 | 1 | 0 | 0.27846304 | 0.51493562 |
| 2 | 4 | 8 | 1 | 4 | 2 | 6 | 3 | 3 | 0.13500522 | 0.34598184 |
| 1 | 0 | 2 | 2 | 0 | 1 | 0 | 0 | 0 | 0.17082632 | 0.40718703 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 4 | 1 | 8 | 10 | 6 | 17 | 2 | 15 | 3 | 0.52497458 | 0.76330157 |
| 2 | 1 | 4 | 0 | 2 | 0 | 1 | 2 | 0 | 0.1022153 | 0.28437213 |
| 0 | 0 | 4 | 0 | 5 | 6 | 0 | 2 | 7 | 0.01274968 | 0.07201637 |
| 0 | 0 | 4 | 0 | 0 | 2 | 3 | 1 | 1 | 0.0447423 | 0.16951147 |
| 0 | 0 | 1 | 4 | 1 | 4 | 1 | 0 | 2 | 0.70238781 | 0.90275157 |
| 35 | 11 | 29 | 19 | 35 | 18 | 1 | 4 | 7 | 0.01451578 | 0.0751595 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0.71923557 | 0.90275157 |
| 0 | 1 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0.07850385 | 0.26037575 |
| 9 | 3 | 16 | 9 | 14 | 10 | 3 | 0 | 7 | 0.72026382 | 0.90275157 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr14 | 105144030 | 105144086 | hsa-mir-4710 | - | 1 | 1 | 0 | 0 | 0 |
| chr7 | 29720349 | 29720444 | hsa-mir-550a-3 | - | 3 | 4 | 4 | 1 | 5 |
| chr22 | 20020661 | 20020743 | hsa-mir-185 | + | 1269 | 310 | 553 | 598 | 190 |
| chr15 | 89155055 | 89155165 | hsa-mir-7-2 | + | 30 | 1 | 7 | 11 | 11 |
| chr17 | 78393217 | 78393293 | hsa-mir-4730 | + | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101520642 | 101520718 | hsa-mir-382 | + | 285 | 61 | 28 | 25 | 138 |
| chr3 | 168269641 | 168269737 | hsa-mir-551b | + | 51 | 1 | 4 | 21 | 5 |
| chrX | 118780700 | 118780811 | hsa-mir-766 | - | 19 | 4 | 5 | 3 | 2 |
| chr11 | 9111858 | 9111926 | hsa-mir-5691 | - | 0 | 0 | 0 | 0 | 0 |
| chr3 | 50210758 | 50210852 | hsa-mir-566 | + | 0 | 0 | 0 | 0 | 0 |
| chr10 | 91352503 | 91352584 | hsa-mir-107 | - | 499 | 28 | 32 | 403 | 197 |
| chr14 | 101514994 | 101515085 | hsa-mir-544a | + | 1 | 0 | 0 | 0 | 0 |
| chr19 | 54264386 | 54264476 | hsa-mir-516a-2 | + | 4 | 0 | 0 | 0 | 0 |
| chr9 | 131154896 | 131154993 | hsa-mir-219-2 | - | 4 | 2 | 0 | 2 | 0 |
| chr11 | 118514717 | 118514797 | hsa-mir-6716 | + | 2 | 0 | 1 | 0 | 1 |
| chr2 | 189842817 | 189842887 | hsa-mir-1245a | + | 0 | 0 | 0 | 0 | 0 |
| chr4 | 10080234 | 10080316 | hsa-mir-3138 | - | 1 | 0 | 0 | 0 | 0 |
| chr12 | 109230683 | 109230782 | hsa-mir-619 | - | 6 | 3 | 3 | 1 | 1 |
| chr11 | 111383662 | 111383746 | hsa-mir-34b | + | 60 | 2 | 0 | 10 | 7 |
| chr8 | 27559193 | 27559276 | hsa-mir-3622a | + | 3 | 4 | 0 | 1 | 0 |
| chrX | 16645134 | 16645208 | hsa-mir-548am | - | 5 | 1 | 2 | 6 | 1 |
| chr14 | 64561741 | 64561843 | hsa-mir-548h-1 | - | 0 | 0 | 0 | 0 | 0 |
| chr17 | 46709851 | 46709921 | hsa-mir-196a-1 | - | 32 | 3 | 19 | 13 | 26 |
| chr14 | 101510534 | 101510620 | hsa-mir-1185-2 | + | 23 | 0 | 0 | 2 | 12 |
| chr17 | 36875943 | 36876001 | hsa-mir-4726 | + | 1 | 1 | 1 | 0 | 1 |
| chr1 | 19223564 | 19223642 | hsa-mir-1290 | - | 1 | 0 | 0 | 0 | 0 |
| chr9 | 6007825 | 6007904 | hsa-mir-4665 | + | 1 | 0 | 0 | 0 | 0 |
| chr11 | 43602943 | 43603033 | hsa-mir-129-2 | + | 18 | 11 | 6 | 11 | 22 |
| chr11 | 59362549 | 59362631 | hsa-mir-3162 | - | 1 | 0 | 0 | 0 | 0 |
| chr1 | 236016299 | 236016360 | hsa-mir-1537 | - | 0 | 0 | 0 | 0 | 0 |
| chr7 | 99691615 | 99691697 | hsa-mir-106b | - | 1192 | 99 | 138 | 1500 | 916 |
| chrX | 145082570 | 145082649 | hsa-mir-891b | - | 1 | 0 | 0 | 0 | 0 |
| chr11 | 57408670 | 57408759 | hsa-mir-130a | + | 2074 | 66 | 32 | 1000 | 2229 |
| chr2 | 69330813 | 69330887 | hsa-mir-3126 | + | 0 | 0 | 0 | 0 | 0 |
| chr3 | 31203195 | 31203279 | hsa-mir-466 | - | 2 | 0 | 0 | 0 | 0 |
| chr11 | 126858353 | 126858438 | hsa-mir-3167 | - | 1 | 0 | 0 | 0 | 0 |
| chr5 | 15935290 | 15935369 | hsa-mir-887 | + | 113 | 72 | 67 | 5 | 28 |
| chr17 | 73780601 | 73780688 | hsa-mir-4738 | - | 1 | 0 | 1 | 0 | 0 |
| chr21 | 17962556 | 17962645 | hsa-mir-125b-2 | + | 18755 | 8865 | 13473 | 4824 | 5389 |
| chr19 | 20510080 | 20510163 | hsa-mir-1270-1 | - | 4 | 1 | 2 | 1 | 1 |
| chr5 | 132763287 | 132763398 | hsa-mir-1289-2 | - | 0 | 0 | 0 | 0 | 0 |
| chr12 | 95228173 | 95228289 | hsa-mir-492 | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 117102645 | 117102733 | hsa-mir-548ac | - | 0 | 0 | 0 | 0 | 0 |
| chr19 | 46178185 | 46178282 | hsa-mir-642a | + | 9 | 1 | 3 | 3 | 2 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0207146 | 0.09750507 |
| 0 | 0 | 6 | 4 | 4 | 7 | 2 | 3 | 3 | 0.47222851 | 0.71331762 |
| 126 | 245 | 1158 | 1476 | 2091 | 2224 | 305 | 557 | 329 | 0.76403572 | 0.92898156 |
| 6 | 8 | 126 | 108 | 161 | 756 | 293 | 146 | 93 | 0.00745674 | 0.04859919 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 0.68634275 | 0.90275157 |
| 124 | 55 | 66 | 86 | 131 | 91 | 23 | 18 | 64 | 0.02193024 | 0.10019112 |
| 2 | 28 | 4 | 19 | 7 | 7 | 4 | 5 | 11 | 0.00510918 | 0.0368377 |
| 5 | 4 | 13 | 23 | 26 | 39 | 10 | 11 | 13 | 0.95916624 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 3 | 1 | 2 | 3 | 2 | 2 | 0 | 0.03124172 | 0.12998785 |
| 111 | 197 | 2237 | 1058 | 1123 | 1811 | 2991 | 1048 | 912 | 0.10062959 | 0.28400472 |
| 0 | 5 | 8 | 6 | 1 | 1 | 2 | 5 | 5 | 0.57986475 | 0.81581943 |
| 1 | 2 | 3 | 1 | 10 | 2 | 0 | 460 | 0 | 6.63E-05 | 0.00199388 |
| 1 | 0 | 2 | 7 | 5 | 4 | 11 | 4 | 2 | 0.66847996 | 0.90275157 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0070372 | 0.04684765 |
| 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 0.28679032 | 0.52204801 |
| 1 | 0 | 4 | 6 | 1 | 9 | 0 | 2 | 1 | 0.16203541 | 0.39430026 |
| 3 | 1 | 7 | 10 | 3 | 5 | 1 | 7 | 0 | 0.49889436 | 0.74514349 |
| 5 | 10 | 79 | 30 | 12 | 8 | 4 | 10 | 86 | 0.64026532 | 0.8741989 |
| 0 | 1 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0.07850385 | 0.26037575 |
| 2 | 1 | 12 | 8 | 6 | 17 | 8 | 10 | 6 | 0.72904282 | 0.90716677 |
| 0 | 0 | 3 | 0 | 1 | 0 | 0 | 3 | 1 | 0.1041783 | 0.28811327 |
| 17 | 38 | 1517 | 313 | 1420 | 2223 | 30 | 126 | 487 | 0.00073706 | 0.00981346 |
| 8 | 7 | 12 | 15 | 9 | 9 | 5 | 4 | 3 | 0.08617664 | 0.2762803 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01357253 | 0.07269882 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 10 | 14 | 6 | 1 | 2 | 3 | 4 | 3 | 1 | 0.00020077 | 0.00425273 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 2 | 1 | 1 | 6 | 1 | 1 | 4 | 0.02240375 | 0.10168384 |
| 417 | 1647 | 8148 | 5044 | 4292 | 9853 | 2071 | 8317 | 5126 | 0.114897 | 0.30549845 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 1279 | 1877 | 2293 | 1491 | 1284 | 1473 | 486 | 2450 | 962 | 0.11054548 | 0.30003945 |
| 0 | 0 | 1 | 4 | 2 | 1 | 0 | 2 | 0 | 0.07083911 | 0.24635094 |
| 0 | 0 | 2 | 2 | 2 | 16 | 4 | 1 | 0 | 0.11966696 | 0.3139037 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 8 | 27 | 50 | 11 | 29 | 16 | 60 | 13 | 12 | 0.02695382 | 0.11523376 |
| 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0.50916975 | 0.75276704 |
| 5625 | 3188 | 10370 | 6814 | 4708 | 5290 | 222 | 1666 | 2122 | 0.00014469 | 0.00329431 |
| 2 | 1 | 1 | 1 | 7 | 16 | 1 | 2 | 1 | 0.70882537 | 0.90275157 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 2 | 1 | 0 | 2 | 0 | 2 | 0 | 0.27846304 | 0.51493562 |
| 0 | 0 | 7 | 1 | 3 | 7 | 2 | 2 | 6 | 0.00787455 | 0.05061433 |
| 1 | 1 | 3 | 11 | 3 | 19 | 0 | 1 | 0 | 0.38442023 | 0.63412329 |

(continued)

**Table S4** (continued)

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromosome | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr2 | 11907569 | 11907651 | hsa-mir-548s | + | 0 | 0 | 0 | 0 | 0 |
| chr1 | 95211415 | 95211456 | hsa-mir-378g | - | 26 | 16 | 37 | 4 | 6 |
| chr6 | 155174493 | 155174570 | hsa-mir-1273c | + | 7 | 5 | 1 | 0 | 1 |
| chr1 | 32552549 | 32552608 | hsa-mir-5585 | + | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101530831 | 101530915 | hsa-mir-541 | + | 1 | 0 | 0 | 0 | 0 |
| chr22 | 28316513 | 28316599 | hsa-mir-3199-2 | + | 5 | 0 | 8 | 1 | 1 |
| chr6 | 119390211 | 119390308 | hsa-mir-548b | - | 17 | 0 | 1 | 2 | 12 |
| chr8 | 41128566 | 41128662 | hsa-mir-548ao | - | 0 | 0 | 0 | 0 | 0 |
| chr10 | 35368525 | 35368608 | hsa-mir-3611 | - | 20 | 14 | 12 | 0 | 2 |
| chr6 | 135560297 | 135560394 | hsa-mir-548a-2 | + | 0 | 0 | 0 | 0 | 0 |
| chr22 | 49937040 | 49937114 | hsa-mir-3667 | - | 1 | 1 | 1 | 0 | 0 |
| chrX | 149396238 | 149396318 | hsa-mir-2114 | + | 2 | 0 | 0 | 2 | 0 |
| chrX | 133303838 | 133303907 | hsa-mir-20b | - | 19 | 1 | 8 | 146 | 28 |
| chr14 | 102026623 | 102026759 | hsa-mir-1247 | - | 37 | 30 | 68 | 2 | 3 |
| chr2 | 176032360 | 176032437 | hsa-mir-933 | - | 1 | 0 | 0 | 0 | 0 |
| chrX | 146340277 | 146340368 | hsa-mir-509-2 | - | 0 | 0 | 0 | 0 | 0 |
| chr7 | 34980371 | 34980446 | hsa-mir-548n | - | 1 | 0 | 0 | 1 | 4 |
| chr17 | 13446845 | 13446963 | hsa-mir-548h-3 | - | 4 | 0 | 0 | 3 | 3 |
| chr19 | 54200786 | 54200871 | hsa-mir-525 | + | 1 | 0 | 0 | 0 | 0 |
| chr10 | 6194158 | 6194240 | hsa-mir-3155a | + | 1 | 0 | 2 | 0 | 0 |
| chr8 | 100548863 | 100548958 | hsa-mir-599 | - | 0 | 0 | 0 | 0 | 0 |
| chr14 | 101506405 | 101506485 | hsa-mir-376a-2 | + | 511 | 240 | 36 | 109 | 224 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 0 | 0 | 4 | 2 | 0 | 10 | 1 | 2 | 3 | 0.01319859 | 0.07269882 |
| 3 | 9 | 5 | 2 | 3 | 8 | 3 | 8 | 5 | 0.00260431 | 0.02360966 |
| 0 | 0 | 4 | 5 | 2 | 11 | 2 | 1 | 0 | 0.44479723 | 0.70097152 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 4 | 4 | 1 | 3 | 4 | 5 | 0 | 4 | 0.20394699 | 0.46247833 |
| 3 | 5 | 10 | 20 | 4 | 12 | 14 | 17 | 8 | 0.55251643 | 0.79073 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 5 | 0 | 0.29112969 | 0.52788496 |
| 3 | 5 | 9 | 14 | 26 | 21 | 11 | 17 | 6 | 0.56872303 | 0.80609294 |
| 0 | 0 | 2 | 3 | 0 | 7 | 2 | 1 | 1 | 0.02176641 | 0.10019112 |
| 0 | 0 | 2 | 1 | 1 | 0 | 1 | 0 | 0 | 0.77199449 | 0.93372218 |
| 2 | 0 | 60 | 10 | 3 | 12 | 13 | 5 | 64 | 0.00376858 | 0.03080977 |
| 13 | 22 | 88 | 58 | 29 | 523 | 88 | 124 | 881 | 0.13512595 | 0.34598184 |
| 3 | 4 | 10 | 31 | 20 | 5 | 0 | 0 | 1 | 0.01054979 | 0.06183902 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 2 | 4 | 3 | 4 | 0 | 3 | 3 | 0.89357109 | 1 |
| 3 | 0 | 9 | 8 | 2 | 19 | 6 | 7 | 3 | 0.70277707 | 0.90275157 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27554357 | 0.51361322 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.02467257 | 0.1100062 |
| 0 | 0 | 0 | 0 | 0 | 86 | 2 | 0 | 0 | 0.00071406 | 0.00964504 |
| 568 | 236 | 148 | 104 | 144 | 96 | 23 | 23 | 65 | 0.00066728 | 0.00958327 |

**Table S5** List of differentially expressed miRNAs

| miRNA Information | | | | | Breast Normal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chromo-some | start | end | miRNA | strand | SRR 191548 | SRR 191549 | SRR 191554 | SRR 191578 | SRR 191581 |
| chr11 | 2155363 | 2155439 | hsa-mir-483 | - | 201 | 95 | 53 | 4 | 12 |
| chr5 | 149112387 | 149112453 | hsa-mir-378a | + | 8867 | 4096 | 7011 | 1053 | 2691 |
| chr14 | 101506026 | 101506092 | hsa-mir-376c | + | 1564 | 1580 | 112 | 248 | 900 |
| chr8 | 41517958 | 41518026 | hsa-mir-486 | - | 800 | 158 | 1223 | 2070 | 166 |
| chr17 | 27188388 | 27188456 | hsa-mir-451b | + | 13127 | 781 | 1856 | 163887 | 5476 |
| chr4 | 5925001 | 5925055 | hsa-mir-378d-1 | - | 166 | 130 | 161 | 29 | 41 |
| chr22 | 42319225 | 42319301 | hsa-mir-378i | - | 263 | 196 | 157 | 28 | 104 |
| chr21 | 17911408 | 17911489 | hsa-mir-99a | + | 13414 | 6547 | 3467 | 3951 | 7922 |
| chr8 | 22102474 | 22102556 | hsa-mir-320a | - | 10074 | 6821 | 8313 | 1362 | 3652 |
| chr18 | 47652868 | 47652933 | hsa-mir-4320 | - | 29 | 18 | 63 | 11 | 6 |
| chr10 | 132760850 | 132760931 | hsa-mir-378c | - | 794 | 462 | 544 | 83 | 292 |
| chr20 | 60639857 | 60639941 | hsa-mir-3195 | + | 158 | 35 | 414 | 6 | 6 |
| chr14 | 101349315 | 101349412 | hsa-mir-127 | + | 2918 | 978 | 645 | 113 | 728 |
| chr11 | 121970464 | 121970552 | hsa-mir-125b-1 | - | 18569 | 8831 | 13470 | 4792 | 5261 |
| chr17 | 27188386 | 27188458 | hsa-mir-451a | - | 13127 | 781 | 1856 | 163887 | 5476 |
| chr5 | 85916313 | 85916392 | hsa-mir-3607 | + | 100 | 166 | 42 | 3 | 41 |
| chr17 | 27188550 | 27188636 | hsa-mir-144 | - | 1133 | 2 | 47 | 10519 | 156 |
| chr8 | 94928249 | 94928347 | hsa-mir-378d-2 | - | 213 | 162 | 200 | 30 | 47 |
| chr9 | 73424890 | 73425000 | hsa-mir-204 | - | 52 | 22 | 10 | 16 | 14 |
| chr1 | 24255559 | 24255637 | hsa-mir-378f | + | 88 | 68 | 76 | 12 | 39 |
| chr11 | 43602943 | 43603033 | hsa-mir-129-2 | + | 18 | 11 | 6 | 11 | 22 |
| chr21 | 17962556 | 17962645 | hsa-mir-125b-2 | + | 18755 | 8865 | 13473 | 4824 | 5389 |
| chr14 | 101506405 | 101506485 | hsa-mir-376a-2 | + | 511 | 240 | 36 | 109 | 224 |
| chr5 | 179442302 | 179442397 | hsa-mir-340 | - | 5 | 1 | 0 | 59 | 1 |
| chr1 | 154166140 | 154166219 | hsa-mir-190b | - | 41 | 1 | 20 | 9 | 7 |
| chr15 | 79502129 | 79502213 | hsa-mir-184 | + | 2 | 1 | 3 | 0 | 2 |
| chr5 | 54466359 | 54466450 | hsa-mir-449a | - | 2 | 0 | 0 | 0 | 0 |
| chr15 | 55665137 | 55665232 | hsa-mir-628 | - | 0 | 0 | 0 | 0 | 0 |
| chr2 | 219866366 | 219866430 | hsa-mir-375 | - | 308 | 77 | 72 | 13 | 81 |
| chr7 | 129414531 | 129414609 | hsa-mir-96 | - | 164 | 7 | 2 | 36 | 24 |
| chr17 | 46709851 | 46709921 | hsa-mir-196a-1 | - | 32 | 3 | 19 | 13 | 26 |

| | | Breast IDC | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR 191582 | SRR 191617 | SRR 191602 | SRR 191603 | SRR 191604 | SRR 191605 | SRR 191615 | SRR 191622 | SRR 191631 | p-value | adj-p-value |
| 16 | 15 | 35 | 5 | 6 | 6 | 1 | 2 | 5 | 2.69E-05 | 0.00108376 |
| 1024 | 2209 | 752 | 796 | 1065 | 1969 | 720 | 943 | 571 | 1.67E-06 | 0.00012973 |
| 1891 | 854 | 365 | 314 | 349 | 253 | 89 | 117 | 390 | 5.32E-05 | 0.00179823 |
| 82 | 197 | 253 | 90 | 78 | 52 | 5 | 13 | 20 | 2.48E-07 | 2.57E-05 |
| 3409 | 8109 | 6793 | 2032 | 2083 | 1434 | 2664 | 566 | 806 | 7.01E-18 | 3.27E-15 |
| 23 | 36 | 13 | 15 | 15 | 36 | 15 | 7 | 5 | 0.00014083 | 0.00329431 |
| 33 | 67 | 23 | 18 | 30 | 86 | 24 | 20 | 21 | 0.00068373 | 0.00958327 |
| 6958 | 5595 | 6273 | 4268 | 3615 | 4186 | 610 | 1604 | 3831 | 0.0002366 | 0.00469175 |
| 2864 | 1160 | 2095 | 1746 | 3159 | 2426 | 821 | 842 | 694 | 1.24E-05 | 0.00063984 |
| 4 | 0 | 11 | 5 | 4 | 7 | 0 | 0 | 0 | 0.00028508 | 0.00494969 |
| 94 | 201 | 73 | 52 | 90 | 179 | 119 | 80 | 65 | 0.00029239 | 0.00495475 |
| 20 | 0 | 11 | 20 | 7 | 11 | 0 | 2 | 1 | 2.01E-06 | 0.00014435 |
| 773 | 543 | 472 | 553 | 689 | 483 | 54 | 88 | 354 | 0.00069921 | 0.00958327 |
| 5540 | 3123 | 10308 | 6759 | 4623 | 5243 | 216 | 1595 | 2113 | 0.00014492 | 0.00329431 |
| 3409 | 8109 | 6793 | 2032 | 2083 | 1434 | 2664 | 566 | 806 | 7.01E-18 | 3.27E-15 |
| 14 | 3 | 1 | 3 | 1 | 5 | 8 | 1 | 2 | 3.58E-07 | 3.33E-05 |
| 116 | 451 | 789 | 256 | 168 | 163 | 108 | 80 | 115 | 7.18E-08 | 8.37E-06 |
| 24 | 40 | 20 | 19 | 22 | 43 | 15 | 12 | 3 | 0.00015176 | 0.00336753 |
| 42 | 30 | 24 | 11 | 2 | 2 | 1 | 1 | 3 | 0.00021629 | 0.00447959 |
| 7 | 28 | 9 | 4 | 7 | 27 | 12 | 8 | 8 | 0.00044233 | 0.00717452 |
| 10 | 14 | 6 | 1 | 2 | 3 | 4 | 3 | 1 | 0.00020077 | 0.00425273 |
| 5625 | 3188 | 10370 | 6814 | 4708 | 5290 | 222 | 1666 | 2122 | 0.00014469 | 0.00329431 |
| 568 | 236 | 148 | 104 | 144 | 96 | 23 | 23 | 65 | 0.00066728 | 0.00958327 |
| 5 | 28 | 954 | 731 | 131 | 916 | 357 | 728 | 503 | 0.00028393 | 0.00494969 |
| 5 | 10 | 1316 | 128 | 23 | 1242 | 862 | 3 | 255 | 0.00049045 | 0.00774742 |
| 1 | 3 | 138 | 385 | 213 | 23795 | 10 | 56 | 13 | 3.07E-12 | 9.53E-10 |
| 0 | 2 | 15 | 245 | 92 | 40 | 4 | 2 | 246 | 7.25E-06 | 0.00045065 |
| 0 | 0 | 23 | 14 | 15 | 36 | 13 | 10 | 17 | 5.75E-05 | 0.00184834 |
| 68 | 117 | 11449 | 443 | 312 | 9400 | 2208 | 374 | 2460 | 0.00025614 | 0.00484292 |
| 26 | 52 | 2172 | 1517 | 1632 | 3334 | 384 | 713 | 938 | 0.00069314 | 0.00958327 |
| 17 | 38 | 1517 | 313 | 1420 | 2223 | 30 | 126 | 487 | 0.00073706 | 0.00981346 |

## References


1. Sharp PA. RNAi and double-strand RNA. Genes Dev. 1999;13(2):139–41.
2. Golden DE, Gerbasi VR, Sontheimer EJ. An inside job for siRNAs. Mol Cell. 2008;31(3): 309–12.
3. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. Nature. 2009;457(7228):413–20.
4. Chapman EJ, Carrington JC. Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet. 2007;8(11):884–96.
5. Nilsen TW. Endo-siRNAs: yet another layer of complexity in RNA silencing. Nat Struct Mol Biol. 2008;15(6):546–8.
6. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94–108.
7. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009;10(2):126–39.
8. Ambros V, Chen X. The regulation of genes and genomes by small RNAs. Development. 2007;134(9):1635–41.
9. Brodersen P, Voinnet O. Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol. 2009;10(2):141–8.
10. Hussain MU. Micro-RNAs (miRNAs): genomic organisation, biogenesis and mode of action. Cell Tissue Res. 2012;349(2):405–13.
11. Kusenda B, et al. MicroRNA biogenesis, functionality and cancer relevance. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2006;150(2):205–15.
12. Faller M, Guo F. MicroRNA biogenesis: there's more than one way to skin a cat. Biochim Biophys Acta. 2008;1779(11):663–7.
13. Gregory RI, Chendrimada TP, Shiekhattar R. MicroRNA biogenesis: isolation and characterization of the microprocessor complex. Methods Mol Biol. 2006;342:33–47.
14. Winter J, et al. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol. 2009;11(3):228–34.
15. Lim LP, et al. Vertebrate microRNA genes. Science. 2003;299(5612):1540.
16. Lynam-Lennon N, Maher SG, Reynolds JV. The roles of microRNA in cancer and apoptosis. Biol Rev Camb Philos Soc. 2009;84(1):55–71.
17. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843–54.
18. Lodish HF, et al. Micromanagement of the immune system by microRNAs. Nat Rev Immunol. 2008;8(2):120–30.
19. Berardi E, et al. microRNAs in ES cell differentiation. Am J Physiol Heart Circ Physiol. 2012;303(8):H931–9.
20. Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. Cell. 2012;149(3):515–24.
21. Osman A. MicroRNAs in health and disease—basic science and clinical applications. Clin Lab. 2012;58(5–6):393–402.
22. Garzon R, Calin GA, Croce CM. MicroRNAs in cancer. Annu Rev Med. 2009;60:167–79.
23. Farazi TA, et al. miRNAs in human cancer. J Pathol. 2011;223(2):102–15.
24. Nygaard S, et al. Identification and analysis of miRNAs in human breast cancer and teratoma samples using deep sequencing. BMC Med Genomics. 2009;2:35.
25. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6(11):857–66.
26. Catto JW, et al. MicroRNA in prostate, bladder, and kidney cancer: a systematic review. Eur Urol. 2011;59(5):671–81.

27. Bartels CL, Tsongalis GJ. MicroRNAs: novel biomarkers for human cancer. Clin Chem. 2009;55(4):623–31.
28. Cortez MA, et al. MicroRNAs in body fluids—the mix of hormones and biomarkers. Nat Rev Clin Oncol. 2011;8(8):467–77.
29. Iorio MV, Croce CM. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol Med. 2012;4(3):143–59.
30. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133–41.
31. Friedlander MR, et al. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40(1):37–52.
32. Pantano L, Estivill X, Marti E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res. 2010;38(5):e34.
33. Hackenberg M, et al. miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res. 2009;37(Web Server issue):W68–76.
34. Zhu E, et al. mirTools: microRNA profiling and discovery based on high-throughput sequencing. Nucleic Acids Res. 2010;38(Web Server issue):W392–7.
35. Zhang Y, et al. CPSS: a computational platform for the analysis of small RNA deep sequencing data. Bioinformatics. 2012;28(14):1925–7.
36. Ronen R, et al. miRNAkey: a software for microRNA deep sequencing analysis. Bioinformatics. 2010;26(20):2615–6.
37. Maragkakis M, et al. DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res. 2009;37(Web Server issue):W273–6.
38. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15–20.
39. Lewis BP, et al. Prediction of mammalian microRNA targets. Cell. 2003;115(7):787–98.
40. Rehmsmeier M, et al. Fast and effective prediction of microRNA/target duplexes. RNA. 2004;10(10):1507–17.
41. Kertesz M, et al. The role of site accessibility in microRNA target recognition. Nat Genet. 2007;39(10):1278–84.
42. Robins H, Li Y, Padgett RW. Incorporating structure to predict microRNA targets. Proc Natl Acad Sci U S A. 2005;102(11):4006–9.
43. Long D, et al. Potent effect of target structure on microRNA function. Nat Struct Mol Biol. 2007;14(4):287–94.
44. Muckstein U, et al. Thermodynamics of RNA-RNA binding. Bioinformatics. 2006;22(10):1177–82.
45. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215–33.
46. Alon S, et al. Systematic identification of edited microRNAs in the human brain. Genome Res. 2012;22(8):1533–40.
47. Bass BL. RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem. 2002;71:817–46.
48. Ekdahl Y, et al. A-to-I editing of microRNAs in the mammalian brain increases during development. Genome Res. 2012;22(8):1477–87.
49. Reinhart BJ, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403(6772):901–6.
50. Ambros V, Lee RC. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. Methods Mol Biol. 2004;265:131–58.
51. Xu G, et al. Cloning and identification of microRNAs in bovine alveolar macrophages. Mol Cell Biochem. 2009;332(1–2):9–16.
52. Long JE, Chen HX. Identification and characteristics of cattle microRNAs by homology searching and small RNA cloning. Biochem Genet. 2009;47(5–6):329–43.
53. He X, et al. Cloning and identification of novel microRNAs from rat hippocampus. Acta Biochim Biophys Sin (Shanghai). 2007;39(9):708–14.

54. Pfeffer S, et al. Identification of virus-encoded microRNAs. Science. 2004;304(5671): 734–6.
55. Lagos-Quintana M, et al. Identification of tissue-specific microRNAs from mouse. Curr Biol. 2002;12(9):735–9.
56. Bentwich I, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet. 2005;37(7):766–70.
57. Morin RD, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008;18(4):610–21.
58. Glazov EA, et al. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res. 2008;18(6):957–64.
59. Creighton CJ, Reid JG, Gunaratne PH. Expression profiling of microRNAs by deep sequencing. Brief Bioinform. 2009;10(5):490–7.
60. Babiarz JE, et al. Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, dicer-dependent small RNAs. Genes Dev. 2008;22(20): 2773–85.
61. Lim LP, et al. The microRNAs of Caenorhabditis elegans. Genes Dev. 2003;17(8): 991–1008.
62. Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Res. 2003;31(13): 3429–31.
63. Lai EC, et al. Computational identification of Drosophila microRNA genes. Genome Biol. 2003;4(7):R42.
64. Hertel J, Stadler PF. Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data. Bioinformatics. 2006;22(14):e197–202.
65. Terai G, et al. miRRim: a novel system to find conserved miRNAs with high sensitivity and specificity. RNA. 2007;13(12):2081–90.
66. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005;15(8):1034–50.
67. Berezikov E, et al. Phylogenetic shadowing and computational identification of human microRNA genes. Cell. 2005;120(1):21–4.
68. Boffelli D, et al. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science. 2003;299(5611):1391–4.
69. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11(5):473–83.
70. Langmead B, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
71. Dohm JC, et al. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008;36(16):e105.
72. Linsen SE, et al. Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods. 2009;6(7):474–6.
73. Friedlander MR, et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol. 2008;26(4):407–15.
74. Cloonan N, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. 2011;12(12):R126.
75. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nat Rev Cancer. 2010;10(6):389–402.
76. Fernandez-Valverde SL, Taft RJ, Mattick JS. Dynamic isomiR regulation in Drosophila development. RNA. 2010;16(10):1881–8.
77. Zhou H, et al. Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res. 2012;40(13):5864–75.
78. Li SC, et al. miRNA arm selection and isomiR distribution in gastric cancer. BMC Genomics. 2012;13 Suppl 1:S13.
79. Li R, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25(15):1966–7.

80. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39(Database issue):D152–7.
81. Griffiths-Jones S, et al. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2005;33(Database issue):D121–4.
82. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;Chapter 4:Unit 4 10.
83. Hackenberg M, Matthiesen R. Annotation-Modules: a tool for finding significant combinations of multisource annotations for gene lists. Bioinformatics. 2008;24(11):1386–93.
84. Pasaniuc B, Zaitlen N, Halperin E. Accurate estimation of expression levels of homologous genes in RNA-seq experiments. J Comput Biol. 2011;18(3):459–68.
85. Moxon S, et al. A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics. 2008;24(19):2252–3.
86. Thomson DW, Bracken CP, Goodall GJ. Experimental strategies for microRNA target identification. Nucleic Acids Res. 2011;39(16):6845–53.
87. Krek A, et al. Combinatorial microRNA target predictions. Nat Genet. 2005;37(5):495–500.
88. Marin RM, Vanicek J. Efficient use of accessibility in microRNA target prediction. Nucleic Acids Res. 2011;39(1):19–29.
89. Marin RM, Vanicek J. Optimal use of conservation and accessibility filters in microRNA target prediction. PLoS One. 2012;7(2):e32208.
90. John B, et al. Human MicroRNA targets. PLoS Biol. 2004;2(11):e363.
91. Betel D, et al. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008;36(Database issue):D149–53.
92. Maragkakis M, et al. Accurate microRNA target prediction correlates with protein repression levels. BMC Bioinformatics. 2009;10:295.
93. Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 2006;34(Web Server issue):W451–4.
94. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406–15.
95. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 1981;9(1):133–48.
96. Miranda KC, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell. 2006;126(6):1203–17.
97. Wiese KC, Hendriks A. Comparison of P-RnaPredict and mfold—algorithms for RNA secondary structure prediction. Bioinformatics. 2006;22(8):934–42.
98. Reczko M, et al. Accurate microRNA target prediction using detailed binding site accessibility and machine learning on proteomics data. Front Genet. 2011;2:103.
99. Blow M, et al. A survey of RNA editing in human brain. Genome Res. 2004;14(12):2379–87.
100. Sommer B, et al. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell. 1991;67(1):11–9.
101. Paz-Yaacov N, et al. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc Natl Acad Sci U S A. 2010;107(27):12174–9.
102. Farajollahi S, Maas S. Molecular diversity through RNA editing: a balancing act. Trends Genet. 2010;26(5):221–30.
103. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem. 2010;79:321–49.
104. Warf MB, et al. Effects of ADARs on small RNA processing pathways in C. elegans. Genome Res. 2012;22(8):1488–98.
105. Lai EC. microRNAs: runts of the genome assert themselves. Curr Biol. 2003;13(23):R925–36.
106. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281–97.
107. Lee Y, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003;425(6956):415–9.

108. Bernstein E, et al. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001;409(6818):363–6.
109. Luciano DJ, et al. RNA editing of a miRNA precursor. RNA. 2004;10(8):1174–7.
110. Sharma PM, et al. RNA editing in the Wilms' tumor susceptibility gene, WT1. Genes Dev. 1994;8(6):720–31.
111. Kawahara Y, et al. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science. 2007;315(5815):1137–40.
112. Kawahara Y, et al. Frequency and fate of microRNA editing in human brain. Nucleic Acids Res. 2008;36(16):5270–80.
113. Habig JW, Dale T, Bass BL. miRNA editing—we should have inosine this coming. Mol Cell. 2007;25(6):792–3.
114. Chiang HR, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010;24(10):992–1009.
115. Vesely C, et al. Adenosine deaminases that act on RNA induce reproducible changes in abundance and sequence of embryonic miRNAs. Genome Res. 2012;22(8):1468–76.
116. Burroughs AM, et al. A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. Genome Res. 2010;20(10):1398–410.
117. de Hoon MJ, et al. Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. Genome Res. 2010;20(2):257–64.
118. Griffiths-Jones S. The microRNA registry. Nucleic Acids Res. 2004;32(Database issue):D109–11.
119. Barrett T, et al. NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res. 2011;39(Database issue):D1005–10.
120. Wang X. miRDB: a microRNA target prediction and functional annotation database with a wiki interface. RNA. 2008;14(6):1012–7.
121. Wang X, El Naqa IM. Prediction of both conserved and nonconserved microRNA targets in animals. Bioinformatics. 2008;24(3):325–32.
122. Piriyapongsa J, et al. microPIR: an integrated database of microRNA target sites within human promoter sequences. PLoS One. 2012;7(3):e33888.
123. Younger ST, Corey DR. Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. Nucleic Acids Res. 2011;39(13):5682–91.
124. Hsu SD, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011;39(Database issue):D163–9.
125. Vergoulis T, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res. 2012;40(Database issue):D222–9.
126. Xiao F, et al. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 2009;37(Database issue):D105–10.
127. Jiang Q, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009;37(Database issue):D98–104.
128. Kaya KD, et al. mESAdb: microRNA expression and sequence analysis database. Nucleic Acids Res. 2011;39(Database issue):D170–80.
129. Kanehisa M, et al. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40(Database issue):D109–14.
130. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
131. Yu W, et al. A navigator for human genome epidemiology. Nat Genet. 2008;40(2):124–5.
132. Nicoloso MS, et al. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res. 2010;70(7):2789–98.
133. Bruno AE, et al. miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3′UTRs of human genes. BMC Genomics. 2012;13:44.
134. Lekprasert P, Mayhew M, Ohler U. Assessing the utility of thermodynamic features for microRNA target prediction under relaxed seed and no conservation requirements. PLoS One. 2011;6(6):e20622.

135. Dweep H, et al. miRWalk—database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J Biomed Inform. 2011;44(5):839–47.
136. Yang JH, et al. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res. 2011;39(Database issue):D202–9.
137. Hiard S, et al. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Res. 2010;38(Database issue):D640–51.
138. Olejniczak M, et al. RNAimmuno: a database of the nonspecific immunological effects of RNA interference and microRNA reagents. RNA. 2012;18(5):930–5.
139. Gennarino VA, et al. HOCTAR database: a unique resource for microRNA target prediction. Gene. 2011;480(1–2):51–8.
140. Cho S, et al. miRGator v2.0: an integrated system for functional investigation of microRNAs. Nucleic Acids Res. 2011;39(Database issue):D158–62.
141. Lu TP, et al. miRSystem: an integrated system for characterizing enriched functions and pathways of MicroRNA targets. PLoS One. 2012;7(8):e42390.
142. Plaisier CL, Bare JC, Baliga NS. miRvestigator: web application to identify miRNAs responsible for co-regulated gene expression patterns discovered through transcriptome profiling. Nucleic Acids Res. 2011;39(Web Server issue):W125–31.
143. Ruepp A, Kowarsch A, Theis F. PhenomiR: microRNAs in human diseases and biological processes. Methods Mol Biol. 2012;822:249–60.
144. Ziebarth JD, et al. PolymiRTS database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. Nucleic Acids Res. 2012;40(Database issue): D216–21.
145. Szczesniak MW, et al. miRNEST database: an integrative approach in microRNA search and annotation. Nucleic Acids Res. 2012;40(Database issue):D198–204.

# Chapter 10
# Beyond miRNAs: Role of Other Noncoding RNAs in Cancer

**Roxana S. Redis and George A. Calin**

**Abstract** Noncoding RNAs are by definition RNA molecules with a short open reading frame that are not translated into proteins. With hundreds of publications yearly, these new players have emerged as master regulators of multiple biological processes. One specific subset of ncRNAs, namely, microRNAs, received prevalent attention over the past decade, yet it forms only the tip of the iceberg. The noncoding RNA world is constantly expanding, whether it is by exploring novel classes of ncRNAs or by gaining further insight into their function. Noncoding RNAs are split in two main classes according to their size: short ncRNAs and long ncRNAs, each class encompassing various subclasses. In this chapter we focus on the most important subclasses of ncRNAs (e.g., piRNAs, snoRNAs, lincRNAs, T-UCRs), highlighting the illustrative examples. Furthermore, we discuss their potential involvement in therapy and diagnostics and what the future might reserve for these key regulators.

**Keywords** microRNAs • Long noncoding RNAs • Cancer • Diagnosis • Therapy

R.S. Redis
Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, 1881 East Road Unit 1850, Houston, TX 77054, USA

Department of Molecular Science, University of Medicine and Pharmacy I. Hatieganu, Cluj-Napoca, Romania

G.A. Calin (✉)
Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, 1881 East Road Unit 1850, Houston, TX 77054, USA
e-mail: gcalin@mdanderson.org

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_10,

## 10.1 Introduction

It was only a few decades ago, that research was driven by the central dogma: DNA → RNA → Protein and it was commonly assumed that genes are synonymous of proteins, the key regulators and effectors of biological processes. RNA was merely the intermediate in this equation. But with the latest disclosure that 97–98 % of the transcriptional output of the human genome is non-protein-coding RNA (ncRNA) [1] the perspective gradually changed, unleashing a noncoding RNA revolution. Moreover, ncRNAs seem to constitute the majority of the transcriptional output in other animals as well, such as Drosophila, with an intron–exon ratio of approximately 1:1 [2].

By definition, "noncoding RNAs" are RNA molecules with a short open reading frame (sORF < 100 codons), that are not translated into proteins. However, there is growing evidence that sORFs present in some ncRNAs are being translated into small functional peptides [3, 4], and it has recently been argued that a RNA molecule can possess both coding and noncoding activities [5].

One specific subset of ncRNAs, namely microRNAs received ample attention from the scientific world. With hundreds of publications yearly, they are the best characterized ncRNAs, yet they represent only the tip of the iceberg [6]. The noncoding RNA world is constantly expanding, whether it is by exploring novel classes of ncRNAs or by gaining further insight into their function. While identifying novel noncoding RNAs has become more accessible, mostly due to the high-throughput technologies available (e.g., tilling arrays, cDNA sequencing, RNA-seq), exploring their functions is still challenging. Undoubtedly, they fulfill a wide range of physiological and pathological functions, as they dominate the genomic output of higher organisms and most of them are evolutionarily conserved. This chapter focuses mainly on the roles of ncRNAs in cancer (initiation, progression, and metastasis).

## 10.2 Small Noncoding RNAs vs. Long Noncoding RNAs

Noncoding RNAs can be categorized into two main classes based on their transcript size:

1. Small noncoding RNAs.
2. Large/long noncoding RNAs.

with the arbitrary cutoff of 200 nucleotides (Table 10.1). Depending on the type of ncRNA, transcription can occur by any of the three polymerases (RNA Pol I, II, or III).

### *10.2.1 Small ncRNAs*

The *small ncRNAs* (18–200 nt) class is reasonably perceived as the miRNA class and not much is known about the other members of this class. They are highly

**Table 10.1** Types of noncoding RNAs

| Type | Subclass | Symbol | Examples | References |
|---|---|---|---|---|
| *Small ncRNAs* (17–200 nt) | MicroRNAs | miRNAs | miR-15/16, miR-21 | [122, 123] |
| | Tiny transcription initiation RNAs | tiRNAs | Associated with the CAP1 gene | [45] |
| | PIWI-RNAs | piRNAs | piR-651, piR-832 | [19, 20] |
| | Transcription start-site associated RNAs | TSSa-RNAs | Associated with RNF12 and CCDC52 genes | [42] |
| | Small nucleolar RNAs | snoRNAs | U50, snoRD114-1 | [35, 39] |
| | Promoter-associated short RNAs | PASRs | | [43] |
| *Long ncRNAs* (more than 200 nt) | Long intergenic RNAs | lincRNAs | MALAT-1, HOTAIR | [66, 79] |
| | Long intronic RNAs | lncRNAs | COLDAIR | [86] |
| | Transcribed ultraconserved regions | T-UCRs | UC.338, UC.73 | [95, 101] |
| | Pseudogenes | | PTENP1 | [104] |
| | Telomere-associated noncoding RNAs | TERRAs | | [107, 108] |

conserved, generally named according to their genomic localization or function (e.g., transcription initiation RNAs—tiRNAs, centromere repeat associated small interacting RNAs—crasiRNAs, or promoter associated small RNAs—PASRs) and have a DICER-dependent (e.g., siRNAs, snoRNA) or DICER-independent (e.g., piRNA) mechanisms of biogenesis, though vaguely described. Small ncRNAs are involved in transcriptional and posttranscriptional gene silencing by specific base pairing with their targets.

This chapter discusses all representatives of the sncRNA class, with the exception of miRNAs, which are in detailed presented elsewhere in this book.

#### 10.2.1.1 Piwi-Interacting RNAs (piRNAs)

PiRNAs are typically 24–32 nt long RNAs, DICER-independent and bind the PIWI subfamily of Argonaute proteins that are involved in maintaining genome stability in germ line cells [7, 8]. They were first described as Piwi-dependent small RNAs mapping to the Drosophila flamenco locus [9], and initially designated as repeat-associated small interfering RNAs (rasiRNAs) [10], as they were thought to derive only from transposons and other repeated sequence elements [8]. It is now clear that piRNAs can be also derived from complex DNA sequence elements and also that rasiRNAs are a subclass of piRNAs [7].

Their exact mechanism of biogenesis is still poorly understood, but a new mechanism was described by Brennecke et al., similar to secondary siRNA generation, named the ping-pong model. According to this mechanism, the antisense piRNAs associate with PIWI/AUB complex while sense piRNAs associate with AGO3

protein. The PIWI and AUB proteins bind to maternally deposited piRNAs (primary piRNA), generating a complex that subsequently binds to the transcripts produced by retrotransposons and cleaves them forming a sense piRNA (secondary piRNA) capable of binding to AGO3. Finally, piRNA-AGO3 complex binds to the retrotransposon transcript, creating another set of antisense piRNAs [11, 12].

The complex formed by piRNAs and PIWI proteins suppresses transposable element expression and mobilization by two distinct mechanisms, both described in *Drosophila melanogaster*:

1. Cleavage of transposable element transcripts by PIWI proteins, mediated by piRNAs through base-pair recognition [13].
2. Heterochromatin mediated gene silencing [14].

The PIWI subfamily, as well as piRNAs, has been associated with germ cell development, stem cell self-renewal, and retrotransposon silencing.

In cancer, piRNAs are understudied and were until recently thought to be absent in cancer cells. However, the growing body of evidence showing the human PIWI orthologs, termed HIWI, to be modified in a variety of human cancers [15–18], suggests that in a tumorigenic state components of the piRNA-PIWI pathway of a cell are altered.

One example of piRNA aberrantly overexpressed in numerous cancers (e.g., gastric, colon, lung, and breast), compared to normal tissue, is piR-651 [19], that was shown to induce G2/M arrest and therby promote the development of cancer. On the contrary, expression of piR-832 in gastric cancer was considerably lower than in normal tissue and artificially increased levels of piR-832 inhibited the growth of gastric cancer cells [20]. Intriguingly, the peripheral blood levels of both piR-651 and piR-832 were significantly lower in patients with gastric cancer than those of controls, implying piRNAs could be valuable biomarkers in diagnosing gastric cancer [21].

Similarly, PIWI proteins have also been implicated in cancer development; PIWIL1 overexpression was linked to cell cycle arrest [18] and PIWIL2 overexpression to anti-apoptotic signaling and cell proliferation [22]. Additionally, ectopic expression of the genes encoding PIWI and other components of the piRNA machinery was shown to induce brain tumor growth [23].

#### 10.2.1.2 Small Nucleolar RNAs (snoRNAs)

SnoRNAs range in size from 60 to 300 nt and represent one of the abundant groups of small ncRNAs described in eukaryotes. They are found predominantly in the nucleus and are components of small nucleolar ribonucleoprotein complexes (snoRNPs), which are responsible for sequence-specific 2′-O-methylation [24] and pseudouridylation [25] of ribosomal RNA (rRNA). The sequences of snoRNAs are responsible for guiding the assembled snoRNPs to a specific target. In vertebrates, most of the snoRNAs are located within introns of protein-coding genes and transcribed by RNA Pol II, but they can be also processed from introns of lncRNAs [26, 27].

There are two major subclasses of snoRNAs, based on the presence of short consensus sequence motifs: (a) Box C/D (RUGAUGA/CUGA) and (b) Box H/ACA (ANANNA/ACA) [28, 29]. Within both subclasses, there are two functionally distinct groups that are likely not to be involved in pseudouridylation and methylation. The box H/ACA snoRNA, snR30 and the box C/D snoRNA, U3 and U14, are required for cleavage of the pre-ribosomal RNAs at the early processing site [30].

SnoRNAs are modulators of maturation and stabilization of rRNAs, these post-transcriptional modifications being essential for the production of precise and efficient ribosomes [31]. Interestingly, it was shown that snoRNAs are processed to produce small RNAs with miRNA-like functions (sno-miRNAs) [32]. Thus, the sno-miRNAs could have dual functions, as snoRNA and as precursor miRNA [33].

Several studies have indicated that alterations of snoRNAs play critical roles in cancer development and progression. The first report linking snoRNAs to cancer was presented by Chang et al., who proved that h5sn2, a box H/ACA snoRNA, was significantly downregulated in human meningiomas compared with normal brain tissues [34]. Consequently, Dong et al. identified snoRNA U50, in the common region of deletion at chromosome 6q14–15, in prostate cancer. Of the 11 genes located in this region, only U50 is mutated in prostate cancer cells [35]. Moreover, chromosome 6q14–15 is at a breakpoint of chromosomal translocation t(3;6)(q27;q15) for human B-cell lymphoma [36]. Other studies showed that growth arrest specific 5 (GAS5), a gene that hosts ten intronic snoRNAs and also encodes an lncRNA, controls mammalian apoptosis and cell growth [37]. GAS5 transcript levels were found to be substantially decreased in breast cancer samples relative to adjacent unaffected normal breast tissue.

Recently, Wang Z. group revealed that an H/ACA box snoRNA-derived miRNA50, miR-605 is potentially involved in stressed-induced stabilization of the p53. As p53 transcriptionally activates MDM2, one of its negative regulators, in addition to miR-605, the sno-miRNA might offset the MDM2 negative feedback loop, thus generating a positive feedback loop to enable rapid accumulation of p53 [38].

Furthermore, the finding that snoRD112-114 located at the DLK1-DIO3 locus are ectopically expressed in acute promyelotic leukemia (APL), suggests there is a relationship between chromosomal translocation and expression of snoRNA loci. Additionally, the in vitro experiments showed that the snoRD114-1 variant promotes cell growth through G0/G1 to S phase transition mediated by the Rb/p16 pathway [39].

#### 10.2.1.3 Telomere Specific Small RNAs (Tel-sRNAs)

Tel-sRNAs are a recently described class of small noncoding RNAs, approximately 24 nt long and 2′-O-methylated at the 3′ end. Their processing is DICER-independent and they are evolutionary conserved, similarly to piRNAs. It was reported that tel-sRNAs are upregulated in cells that carry null mutation of histone H3K4 methyltransferase MLL and downregulated in cells that carry null mutations of histone

H3K9 methyltransferase SUV39H, suggesting that they are subjected to epigenetic regulation, supporting the concept that tel-sRNAs are heterochromatin associated pi-like small RNAs [40]. Moreover, telomerase TERT activity was shown to be inhibited in vitro by an 18-mer RNA oligo of $(UUAGGG)_3$ through RNA duplex formation in the template region of the telomerase RNA component [41].

#### 10.2.1.4 RNAs Associated with Transcription Start Site/Promoters

Various classes of small ncRNAs that are associated with transcription start site of the genes, have been described such as transcription start site-associated RNAs (TSSs-RNAs) [42], promoter-associated small RNAs (PASRs) [43], promoter upstream transcripts (PROMPTs) [44] and transcription initiation RNAs (tiRNAs) [45]. Their size ranges from 16 to 200 nt and they are transcribed both sense and antisense around the promoter, yet they are not associated with any known protein-coding genes. These promoter-associated RNAs are assumed to be involved in transcription regulation by targeting epigenetic silencing complexes, even though their functions are not well described. Therefore, it is supposed that their altered expression might be linked to different diseases, including cancer. Watanabe et al. strengthen this assumption by demonstrating they have the potential to form double-stranded RNAs and to be processed further into endogenous siRNAs [46].

### *10.2.2 Long/Large ncRNAs*

Long noncoding RNAs (lncRNAs) form a heterogeneous class of noncoding transcripts, that can be transcribed both sense and antisense, from intronic or intergenic regions by RNA Pol II and display epigenetic marks consistent with a transcribed gene (H3K4me3 at the gene promoter, H3K36me3 throughout the gene body). They were originally discovered by large-scale sequencing of full-length cDNA libraries in mouse [47]. The molecular mechanisms by which this class of noncoding RNAs controls gene expression at various levels including chromatin modification, transcriptional and posttranscriptional processing, are not yet entirely understood [48, 49]. This gene regulation can be achieved by targeting either genomically local (*cis-regulation*) or genomically distant (*trans*-regulation) genes (Fig. 10.1). Four possible mechanisms of lncRNAs action have been characterized recently by Wang and Kang [50] and according to the authors these molecules can act as signals, decoys, guides or scaffolds. In addition, a new type of long ncRNAs as gene enhancers, named eRNAs, has also been associated with transcriptional regulation [51].

This class of ncRNAs forms the largest part of the mammalian noncoding transcriptome [52] and for most part is expressed in a disease-, tissue-, or developmental-stage manner converting them into attractive therapeutic targets [53–55].

The dispute around the proper term for this class of noncoding RNAs is an ongoing matter, whether they are referred to as long or large mRNA like-RNAs, they

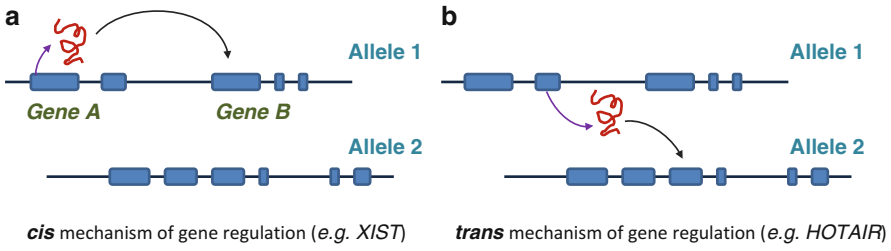


**Fig. 10.1** Long ncRNAs—*cis* vs. *trans* mechanism of gene regulation. LncRNAs can regulate gene expression via two distinct mechanisms: (**a**) *cis*-regulation, when they act upon genes located on the same allele; or (**b**) *trans*-regulation, when they act upon genes located on different alleles

have all the above listed characteristic and should not be wrongly described as lincRNAs (long or large intergenic RNAs) which are a subclass of lncRNAs. T-UCRs and antisense RNAs fall as well under the lncRNA umbrella term, and in this chapter we use the general term "long noncoding RNA" when defining the entire class.

Initial reports describing long noncoding RNAs preceded the discovery of miRNAs , XIST and H19 as being the first characterized lncRNAs. H19 is involved in imprinting, while XIST plays a critical function in X-chromosome inactivation [56, 57]. Yet with the discovery of the first miRNA, lin-14 [58] and their announced association with cancer [59] the focus of the ncRNA research diverted to the class of small ncRNAs. However, the emerging evidence of the past decade suggests that lncRNAs play an essential role in tumor biology with a greater complexity of their functions.

#### 10.2.2.1 Long Intergenic ncRNAs (lincRNAs) and Long Intronic ncRNAs

As implied by their names, the main difference between these two subclasses of long noncoding RNAs is their genomic locus: lincRNAs are transcribed from genomic regions between protein-coding genes, while lncRNAs from the introns of protein-coding genes. On the other hand, both subclasses have been shown to be evolutionary conserved [60–63]. With nearly 3,000 identified transcripts by histone mark signatures [60, 64], lincRNAs are much better described in the literature than long intronic ncRNAs.

After surveying several studies, Koziol and Rinn concluded that lincRNAs are required for the proper establishment of chromatin domains, possibly by steering epigenetic modifying complexes to their specific destinations [65]. This notion is supported by Khalil et al. study showing that lincRNAs bind to PRC2 and multiple other chromatin modifying complexes and are required for proper PRC2 mediated gene repression at gene loci that are normally repressed by PRC2 [64].

One of the first lincRNAs associated with cancer was *MALAT-1* (Metastasis-Associated Lung Adenocarcinoma Transcript 1), validated as a prognostic marker

for metastasis and patient survival in non-small-cell lung cancer (NSCLC) [66]. This highly conserved [67, 68] noncoding RNA is widely expressed in normal human tissues [66, 69] and upregulated in a variety of tumors, such as breast, prostate, colon, liver or uterus [70–73].

The 8 kb long MALAT-1 transcript can be subjected to posttranscriptional processing to yield a tRNA-like small cytoplasmic RNA molecule of 61 nucleotides, named mascRNA [74]. Although the function of the mascRNA is unknown so far, the generation of small RNAs by posttranscriptional processing of lncRNAs, emerges as a potential central theme of the ncRNA biology.

MALAT-1 is retained in the nucleus and particularly localizes to nuclear speckles [69]. Nuclear speckles are considered to be involved in the assembly, modification and/or storage of pre-mRNA processing machinery [75]. It has recently been reported that MALAT-1 is guiding alternative splicing of pre-mRNA by modulating the levels of serine/arginine rich splicing factors proteins [68]. Moreover, depletion of MALAT-1 affects the processing of a subset of pre-mRNAs relevant for cancer biology [76].

Furthermore, a recent study has implicated this lincRNA in the regulation of cell motility in lung cancer cells [77] and is thought to be achieved by transcriptional or posttranscriptional regulation of motility related genes. Additionally, RNA interference-mediated silencing of MALAT-1 reduced cell proliferation and invasion of cervical cancer cells in vitro [78].

A second lincRNA involved in cancer metastasis is *HOTAIR* (HOX Antisense Intergenic RNA), located in the mammalian HOXC locus at chromosome 12q13.13 [79]. It was found to be highly upregulated in both primary and metastatic breast tumors and its expression level in primary tumors positively correlates with metastases and poor outcome [80]. The same study showed that HOTAIR overexpression induces epigenetic changes (histone methylation) via PRC2 (polycomb repressive complex 2), consequently increasing cancer invasiveness and metastases. The HOTAIR RNA appears to be functioning as a scaffold, binding PRC2 complex, responsible for H3K27 methylation in the 5′ region of the RNA and LSD1, a histone lysine demethylase that mediates enzymatic demethylation of H3K4ME2 in the 3′ region of the RNA molecule [81, 82]. The result is the transcriptional silencing of the HOXD locus, which is responsible for remodeling of gene expression pattern of epithelial breast cells in order to resemble embryonic fibroblast (Fig. 10.2).

Further studies have described altered levels of HOTAIR in solid cancer—low levels in colon cancer [83], and high levels in hepatocellular carcinoma; the latter were correlated with both liver metastases and poor survival [84]. Furthermore, HOTAIR may be a prognostic marker not only for HCC recurrence [84], but also for lymph node metastases in HCC [85]. Collectively, these studies suggest that HOTAIR reprograms chromatin state to promote cancer metastasis by a mechanism not entirely elucidated.

Similar to HOTAIR, Heo et al. described a long intronic noncoding RNA termed *COLDAIR* (Cold Assisted Intronic noncoding RNA), which is required for the vernalization-mediated epigenetic-repression of FLC by PRC2 [86].

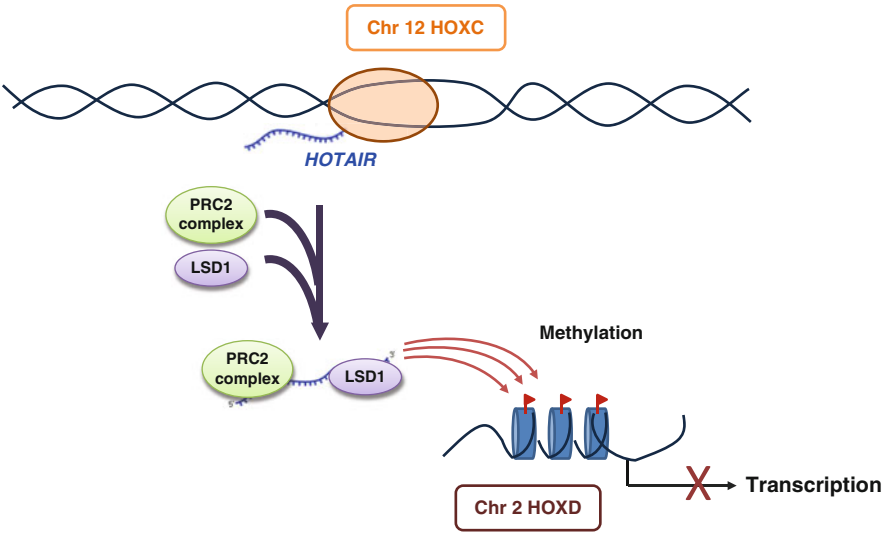


**Fig. 10.2** Functional mechanism of HOTAIR. HOTAIR, ncRNA is transcribed from the HOXC locus and by binding the PRC2 and LSD1 complex, mediates repression of transcription at the HOXD locus, a known PRC2 target

The first long noncoding RNA introduced as a potential tumor suppressor is *MEG3* (Maternally Expressed Gene 3). It is expressed in many normal human tissues, with the highest levels in the brain and pituitary gland [87, 88], while in cancer cells the expression is lost implying a potential role in suppression of cell growth. A possible mechanism of silencing is hypermethylation of the MEG3 regulatory region, which leads to loss of MEG3 expression [88].

MEG3 is a paternally imprinted, single copy gene comprised of 10 exons [87]. At least 12 isoforms have been detected by alternative splicing, with common exons 1–3 and 8–10, but varying in the combination of exons 4–7 [89]. Distinctly, the last intron of MEG3 encodes for miR-770, similar to the presence of miR-675 in H19 [90]. Investigations into the functions of MEG3 have linked this ncRNA to the notorious tumor suppressor, p53. It was documented that MEG3 can stimulate both p53-dependent and p53-independent pathways, contingent upon its secondary structure [89, 91].

*HULC* (Highly Upregulated in Liver Cancer) is a lincRNA discovered in a microarray-based study of gene expression in hepatocellular carcinoma in 2007 [92]. This evolutionary conserved ncRNA, transcribed from chromosome 6q24.3 is a promising biomarker candidate. In addition to liver cancer, colorectal carcinomas with metastasis to the liver, as well as HCC cell lines producing hepatitis B virus present high levels of HULC [93]. This lncRNA appears to be part of an intricate auto-regulatory network, which when disturbed leads to increased expression of HULC [94]. It functions as a "molecular decoy" or "miRNA sponge" binding miR-372, which results in derepression of PRKACB, a kinase targeting cAMP response element binding protein (CREB) and subsequent induction of HULC.

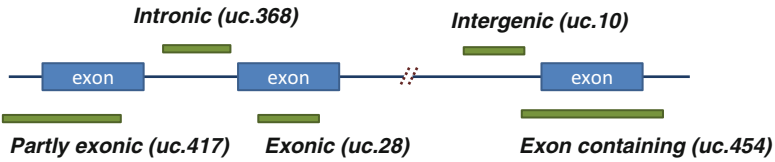


**Fig. 10.3** Genomic locations of T-UCRs

#### 10.2.2.2 Transcribed-Ultraconserved Region (T-UCRs)

Transcribed-ultraconserved regions are evolutionary conserved sequences located both in intragenic and intergenic regions of the human genome [95, 96]. They were first described in 2004 by Bejerano et al., who reported the presence of 481 segments in the human genome, that were larger than 200 nt and harbored 100 % identity with mouse and rat [97]. These ultraconserved regions are transcribed into products ranging from 200 to 779 nt in length and are classified into five categories according to location: intergenic (38.7 %), intronic (42.6 %), exonic (4.2 %), partly exonic (5 %), or exonic-containing (5.6 %) [98] (Fig. 10.3).

Their high degree of conservation combined with their tissue-specific expression, suggests that UCRs are likely to be of important to both ontogeny and phylogeny and have functionality in basic biology. Calin et al. have proven that differentially expressed UCRs in cancer systems could alter the functional characteristics of malignant cells. Moreover, UCRs reside in genomic regions associated with specific types of cancer and their specific patterns of expression in solid tumors and hematological malignancies can differentiate between phenotypically distinct groups [95].

Their functions are not well understood; it was proposed that untranscribed UCRs may function as enhancers [99], while T-UCRs may act as antisense inhibitors for protein-coding genes or other noncoding RNAs, such as miRNAs. The opposite is plausible as well, miRNAs targeting T-UCRs, especially because of the significant antisense complementarity of T-UCRs with miRNAs and the negative correlation between expression of specific T-UCRs and predicted interactor miRNAs [95, 100]. Furthermore, there is also evidence to suggest that epigenetic modifications can regulate T-UCRs [100].

One example of T-UCR relevant in colon cancer is UC.73A, which is one of the most upregulated T-UCRs, with an oncogenic effect on cell proliferation [95]. Similar, UC.338 was found to be significantly upregulated in HCC cell lines and to be part of a larger transcriptional unit, TUC338, which is involved in cell growth [101].

#### 10.2.2.3 Pseudogene RNAs

Pseudogenes are gene copies that do not code for proteins and are typically identified through annotation of disabled, decayed or incomplete protein-coding

sequences. Initially thought to have no function, it has recently been reported that some pseudogenes harbor the potential to regulate their protein-coding genes [102, 103]. For example, pseudogene pair: PTEN and PTENP1 (Phosphatase and Tensin homolog Pseudogene 1) share 3′UTR regions of their RNA transcripts, which both bind the same miRNAs. By binding the miRNAs, the PTENP1 lncRNA diminishes the effects of translational repression on PTEN, in this way allowing expression of the tumor suppressor. Specific mutations in the miRNA binding sites in PTENP1 cancer cells revoke PTENP1 function as a "miRNA decoy," leading to re-repression of PTEN and tumor growth [104].

Moreover, MYLKP1 (Myosin Light chain Kinase Pseudogene 1) is partially duplicated from the MYLK gene, which encodes non-muscle and smooth muscle myosin light chain kinases (smMLCK) isoforms implicated in the regulation of cell contractility and cytokinesis. The pair displays negatively correlated transcriptional patterns in normal and cancer cells with MYLKP1 strongly expressed in cancer cells, whereas smMLCK is highly expressed in non-neoplastic cells [105].

#### 10.2.2.4 Telomere-Associated Noncoding RNAs

The replication potential of normal cells is limited due to the appearance of either senescence or crisis, with the latter finally ending in cell death. On the contrary, cancer cells show nearly unlimited replicative potential. Telomeres, the chromosome ends, are crucial for the replication limit, as well as for chromosome stability [106].

It was believed until recently that telomeres are transcriptionally silent, when several groups independently reported that subtelomeric and telomeric regions, even though devoid of genes, can be transcribed into telomeric UUAGGG-repeat containing ncRNAs, termed TERRA [107–109]. TERRA molecules are conserved among eukaryotes and are believed to function as negative regulators of telomeres by directly inhibiting the telomerase enzyme [110]. HnRNP A1, a protein interaction partner of TERRA, acts together with POT1 (protection of telomerase 1) to remove RPA (replication protein A) from the telomeric ssDNA after DNA replication in order to promote telomere capping and preserve genomic stability [111]. For maintaining telomerase-mediated telomere lengthening, reduction of TERRA transcription is required. Thus, telomerase-positive cancer cells with high levels of subtelomeric methylation present low levels of TERRA compared with matched ALT-positive cancer cells or normal cells (Ng et al., Nucleic Acid Res, 2009). Moreover, Sampl et al. showed that the expression of TERRA in patients with glioblastoma multiform negatively correlates with the tumor grade, proposing a diagnostic value for TERRA molecules [112].

Recently, the GENCODE group [113] within the ENCODE (Encyclopedia of DNA elements) framework has produced a complete, high-quality human lncRNA annotation to date. Furthermore, an extensive list of long noncoding RNAs associated with cancer is provided in an overview article from Spizzo et al. [5].

## 10.3 Noncoding RNAs in Therapy and Diagnostics

The evidence linking small and long noncoding RNAs to pathological conditions and, in particular, cancer development and progression is increasing and it is advancing the idea that ncRNAs may become valuable therapeutic targets and useful biomarkers for diagnostic purposes. Until now, most of the work in this area has been focused on miRNAs and two main therapeutic approaches for targeting miRNAs have been developed: (a) inhibiting miRNA function; and (b) restoring miRNA function. Both approaches can be transposed to developing therapies for other noncoding RNAs if our understanding of their mechanism of action expands. It has been argued that mediating transcriptional gene silencing (TGS) pathways, especially those of tumor suppressors and oncogenes, could be of high therapeutic benefit [114], but successful inhibition of lncRNAs has proven to be more difficult than with miRNAs, mainly because of their complex secondary structures. Several positive results in knocking down lincRNAs have been reported, though [80, 115]. New bioinformatics methods to identify inhibitors for lncRNAs are being developed, one such example being SELEX (Systematic evolution of ligands by exponential enrichment) [115]. This method has already been used for determining RNA sequences that bind to pri-miRNAs [116]. An alternative way of targeting lncRNAs would be with small molecule compounds, which have already been tested for other uses in clinical trials to determine toxicity, pharmacokinetics and pharmacodynamics, some of them being already FDA approved for use in humans. This concept is currently investigated for miRNAs and we believe it could be an answer for other ncRNA-targeting therapies.

Long ncRNA may be valuable biomarkers for diagnosis, prognosis and prediction of response to therapy. Like circulating miRNAs, detectable in a wide variety of body fluids, lncRNAs have proven utility as fluid-based markers of specific cancers. Supporting this concept is the transformation of the prostate specific lncRNA, DD3, into a highly specific, nucleic acid amplification-based marker of prostate cancer, which presents greater specificity then serum prostate specific antigen (PSA) [117, 118]. HULC is a different example of lncRNA detected in the blood of hepatocarcinoma patients. Hence, it is reasonable to assume that a plethora of lncRNAs will be identified in various body fluids in the near future.

Additionally, tissue specificity and altered expression levels of lncRNAs in tumors compared to normal tissue, features shared among these noncoding RNAs, recommend these biomolecules for future clinical research.

## 10.4 NcRNAs: Where Do We Go from Here?

The noncoding RNA revolution is constantly expanding, although much effort is still required to determine and comprehend the full extent of their potential. To advance towards the final aim of translating all this knowledge into clinical practice,

two challenges need to be overcome: identifying all functional ncRNAs encoded in the human genome and understanding their mechanism of function.

Various genomic, epigenomic, and bioinformatics approaches have been taken in the attempt to find novel noncoding RNA elements in the human genome. Methods such as microarray-based profiling or second-generation sequencing (e.g., RNA-seq) contribute considerably to the assembling of a more detailed picture of the human ncRNA transcriptome. Additionally, several algorithms have been developed to identify potentially functional ncRNAs, such as RNAz, FOLDALING, RNAfold, or QRNA.

The biggest challenge though is exploring the functions and mechanism of action of the novel molecules. The classical RNA-interference method had already been implemented in in vitro studies, and it has gathered valuable mechanistic information. A step further is animal models, which render important details about function of ncRNAs in vivo. Moreover, the general assumption is that ncRNAs function in ribonucleoprotein complexes and as such, identifying their protein interaction partners will tell us more about their mechanism of action. In this sense, methods using complementary oligonucleotides to pull down ncRNAs associated with chromatin have been introduced: ChIRP (Chromatin isolation of RNA purification) and CHART (capture hybridization analysis of RNA targets) [119, 120]. Supplementary approaches are RIP-seq (RNA immunoprecipitation-sequencing) and PAR-CLiP (Phosphoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation) [121].

However, the key question remains: how will we efficiently translate the molecular signatures established in the laboratory to the clinical practice?

**Acknowledgements** Dr Calin is The Alan M. Gewirtz Leukemia & Lymphoma Society Scholar. He is supported also as a Fellow at The University of Texas MD Anderson Research Trust, as a University of Texas System Regents Research Scholar and by the CLL Global Research Foundation. Work in Dr. Calin's laboratory is supported in part by the NIH/NCI (CA135444), a Department of Defense Breast Cancer Idea Award, Developmental Research Awards in Breast Cancer, Ovarian Cancer, Brain Cancer, Prostate Cancer, Multiple Myeloma, and Leukemia SPOREs, a SINF MDACC_DKFZ grant in CLL, a SINF grant in colon cancer, the Laura and John Arnold Foundation, the RGK Foundation, and the Estate of C. G. Johnson, Jr.

## References

1. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. 2001;2(11):986–91.
2. Deutsch M, Long M. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res. 1999;27(15):3219–28.
3. Hanada K et al. A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. Genome Res. 2007;17(5):632–40.
4. Kondo T et al. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science. 2010;329(5989):336–9.

5. Spizzo R et al. Long non-coding RNAs and cancer: a new frontier of translational research? Oncogene. 2012;31(43):4577–87.
6. Ruvkun G. Molecular biology. Glimpses of a tiny RNA world. Science. 2001;294(5543): 797–9.
7. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science. 2007;318(5851):761–4.
8. Brennecke J et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089–103.
9. Sarot E et al. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics. 2004;166(3):1313–21.
10. Aravin AA et al. The small RNA profile during Drosophila melanogaster development. Dev Cell. 2003;5(2):337–50.
11. Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. Development. 2008;135(1):3–9.
12. Thomson T, Lin H. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu Rev Cell Dev Biol. 2009;25:355–76.
13. Gunawardane LS et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315(5818):1587–90.
14. Pal-Bhadra M et al. Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science. 2004;303(5658):669–72.
15. Grochola LF et al. The stem cell-associated Hiwi gene in human adenocarcinoma of the pancreas: expression and risk of tumour-related death. Br J Cancer. 2008;99(7):1083–8.
16. He W et al. Expression of HIWI in human esophageal squamous cell carcinoma is significantly associated with poorer prognosis. BMC Cancer. 2009;9:426.
17. Jiang J et al. Expression of HIWI in human hepatocellular carcinoma. Cell Biochem Biophys. 2011;61(1):53–8.
18. Liu X et al. Expression of hiwi gene in human gastric cancer was associated with proliferation of cancer cells. Int J Cancer. 2006;118(8):1922–9.
19. Cheng J et al. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. Clin Chim Acta. 2011;412(17–18):1621–5.
20. Cheng J et al. piR-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells. Cancer Lett. 2012;315(1):12–7.
21. Cui L et al. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. Clin Biochem. 2011;44(13):1050–7.
22. Lee TI et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006;125(2):301–13.
23. Janic A et al. Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. Science. 2010;330(6012):1824–7.
24. Kiss-Laszlo Z et al. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell. 1996;85(7):1077–88.
25. Ni J, Tien AL, Fournier MJ. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell. 1997;89(4):565–73.
26. Bortolin ML, Kiss T. Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. RNA. 1998;4(4):445–54.
27. Smith CM, Steitz JA. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. Mol Cell Biol. 1998;18(12):6897–909.
28. Terns MP, Terns RM. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. Gene Expr. 2002;10(1–2):17–39.
29. Weinstein LB, Steitz JA. Guided tours: from precursor snoRNA to functional snoRNP. Curr Opin Cell Biol. 1999;11(3):378–84.
30. Mannoor K, Liao J, Jiang F. Small nucleolar RNAs in cancer. Biochim Biophys Acta. 2012;1826(1):121–8.

31. Decatur WA, Fournier MJ. rRNA modifications and ribosome function. Trends Biochem Sci. 2002;27(7):344–51.
32. Ender C et al. A human snoRNA with microRNA-like functions. Mol Cell. 2008;32(4): 519–28.
33. Scott MS, Ono M. From snoRNA to miRNA: dual function regulatory non-coding RNAs. Biochimie. 2011;93(11):1987–92.
34. Chang LS et al. Differential expression of human 5S snoRNA genes. Biochem Biophys Res Commun. 2002;299(2):196–200.
35. Dong XY et al. SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. Hum Mol Genet. 2008;17(7):1031–42.
36. Tanaka R et al. Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma. Genes Cells. 2000;5(4):277–87.
37. Mourtada-Maarabouni M et al. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene. 2009;28(2):195–208.
38. Xiao J et al. miR-605 joins p53 network to form a p53:miR-605:Mdm2 positive feedback loop in response to stress. EMBO J. 2011;30(24):5021.
39. Valleron W et al. Specific small nucleolar RNA expression profiles in acute leukemia. Leukemia. 2012;26(9):2052–60.
40. Cao F et al. Dicer independent small RNAs associate with telomeric heterochromatin. RNA. 2009;15(7):1274–81.
41. Horard B, Gilson E. Telomeric RNA enters the game. Nat Cell Biol. 2008;10(2):113–5.
42. Seila AC et al. Divergent transcription from active promoters. Science. 2008;322(5909): 1849–51.
43. Kapranov P et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316(5830):1484–8.
44. Preker P et al. RNA exosome depletion reveals transcription upstream of active human promoters. Science. 2008;322(5909):1851–4.
45. Taft RJ et al. Tiny RNAs associated with transcription start sites in animals. Nat Genet. 2009;41(5):572–8.
46. Watanabe T et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. 2008;453(7194):539–43.
47. Okazaki Y et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature. 2002;420(6915):563–73.
48. Whitehead J, Pandey GK, Kanduri C. Regulation of the mammalian epigenome by long noncoding RNAs. Biochim Biophys Acta. 2009;1790(9):936–47.
49. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 2009;23(13):1494–504.
50. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011; 43(6):904–14.
51. Wang D et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011;474(7351):390–4.
52. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10(3):155–9.
53. Amaral PP et al. Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA. 2009;15(11):2013–27.
54. Fu X et al. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. DNA Cell Biol. 2006;25(3):135–41.
55. Ravasi T et al. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res. 2006;16(1):11–9.
56. Brannan CI et al. The product of the H19 gene may function as an RNA. Mol Cell Biol. 1990; 10(1):28–36.
57. Brown CJ et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature. 1991;349(6304):38–44.

58. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843–54.
59. Calin GA et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24): 15524–9.
60. Guttman M et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–7.
61. Guttman M et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010;28(5): 503–10.
62. Louro R et al. Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. Genomics. 2008;92(1):18–25.
63. Rearick D et al. Critical association of ncRNA with introns. Nucleic Acids Res. 2011; 39(6):2357–66.
64. Khalil AM et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009;106(28): 11667–72.
65. Koziol MJ, Rinn JL. RNA traffic control of chromatin complexes. Curr Opin Genet Dev. 2010;20(2):142–8.
66. Ji P et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003;22(39):8031–41.
67. Gutschner T, Baas M, Diederichs S. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. Genome Res. 2011;21(11): 1944–54.
68. Tripathi V et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010;39(6):925–38.
69. Hutchinson JN et al. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics. 2007;8:39.
70. Guffanti A et al. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics. 2009;10:163.
71. Lin R et al. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene. 2007;26(6):851–8.
72. Luo JH et al. Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas. Hepatology. 2006;44(4):1012–24.
73. Yamada K et al. Phenotypic characterization of endometrial stromal sarcoma of the uterus. Cancer Sci. 2006;97(2):106–12.
74. Wilusz JE, Freier SM, Spector DL. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell. 2008;135(5):919–32.
75. Zhao R, Bodnar MS, Spector DL. Nuclear neighborhoods and gene expression. Curr Opin Genet Dev. 2009;19(2):172–9.
76. Lin R et al. Control of RNA processing by a large non-coding RNA over-expressed in carcinomas. FEBS Lett. 2011;585(4):671–6.
77. Tano K et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. FEBS Lett. 2010;584(22):4575–80.
78. Guo F et al. Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. Acta Biochim Biophys Sin (Shanghai). 2010;42(3):224–9.
79. Rinn JL et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129(7):1311–23.
80. Gupta RA et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6.
81. Hayami S et al. Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. Int J Cancer. 2011;128(3):574–86.

82. Tsai MC et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689–93.
83. Kogo R et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011;71(20):6320–6.
84. Yang Z et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011;18(5):1243–50.
85. Geng YJ et al. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res. 2011;39(6):2119–28.
86. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science. 2011;331(6013):76–9.
87. Miyoshi N et al. Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. Genes Cells. 2000;5(3):211–20.
88. Zhang X et al. A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. J Clin Endocrinol Metab. 2003;88(11):5119–26.
89. Zhang X et al. Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. Endocrinology. 2010;151(3):939–47.
90. Hagan JP et al. At least ten genes define the imprinted Dlk1-Dio3 cluster on mouse chromosome 12qF1. PLoS One. 2009;4(2):e4352.
91. Zhou Y et al. Activation of p53 by MEG3 non-coding RNA. J Biol Chem. 2007;282(34): 24731–42.
92. Panzitt K et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology. 2007;132(1):330–42.
93. Matouk IJ et al. Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis. Eur J Gastroenterol Hepatol. 2009;21(6):688–92.
94. Wang J et al. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 2010;38(16):5366–83.
95. Calin GA et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell. 2007;12(3):215–29.
96. Katzman S et al. Human genome ultraconserved elements are ultraselected. Science. 2007; 317(5840):915.
97. Bejerano G et al. Ultraconserved elements in the human genome. Science. 2004;304(5675): 1321–5.
98. Mestdagh P et al. An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. Oncogene. 2010;29(24):3583–92.
99. Nobrega MA et al. Scanning human gene deserts for long-range enhancers. Science. 2003;302(5644):413.
100. Lujambio A et al. CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. Oncogene. 2010;29(48):6390–401.
101. Braconi C et al. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. Proc Natl Acad Sci U S A. 2011; 108(2):786–91.
102. Harrison PM et al. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res. 2005;33(8):2374–83.
103. Pink RC et al. Pseudogenes: pseudo-functional or key regulators in health and disease? RNA. 2011;17(5):792–8.
104. Poliseno L et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033–8.
105. Han YJ et al. A transcribed pseudogene of MYLK promotes cell proliferation. FASEB J. 2011;25(7):2305–12.

106. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet. 2005;6(8):611–22.
107. Azzalin CM et al. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science. 2007;318(5851):798–801.
108. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol. 2008;10(2):228–36.
109. Schoeftner S, Blasco MA. A "higher order" of telomere regulation: telomere heterochromatin and telomeric RNAs. EMBO J. 2009;28(16):2323–36.
110. Redon S, Reichenbach P, Lingner J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. Nucleic Acids Res. 2010;38(17):5797–806.
111. Flynn RL et al. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. Nature. 2011;471(7339):532–6.
112. Sampl S et al. Expression of telomeres in astrocytoma WHO grade 2 to 4: TERRA level correlates with telomere length, telomerase activity, and advanced clinical grade. Transl Oncol. 2012;5(1):56–65.
113. Harrow J et al. GENCODE: producing a reference annotation for ENCODE. Genome Biol. 2006;7 Suppl 1:S4 1–9.
114. Morris KV. RNA-directed transcriptional gene silencing and activation in human cells. Oligonucleotides. 2009;19(4):299–306.
115. Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res. 2011;71(1):3–7.
116. Lunse CE et al. An aptamer targeting the apical-loop domain modulates pri-miRNA processing. Angew Chem Int Ed Engl. 2010;49(27):4674–7.
117. Hessels D et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003;44(1):8–15. discussion 15-6.
118. Tinzl M et al. DD3PCA3 RNA analysis in urine–a new perspective for detecting prostate cancer. Eur Urol. 2004;46(2):182–6. discussion 187.
119. Chu C et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44(4):667–78.
120. Simon MD et al. The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A. 2011;108(51):20497–502.
121. Hafner M et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010;141(1):129–41.
122. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215–33.
123. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet. 2004;5(7):522–31.

# Chapter 11
# Translational Implications for Noncoding RNA in Cancer

**Duaa Dakhallah, Melissa Piper, and S. Patrick Nana-Sinkam**

**Abstract** MicroRNAs (miRNAs) are involved in human carcinogenesis. Several studies have shown that restoring a normal miRNome (defined as the full spectrum of miRNA expression in a given genome) in cancer cells can exert an antitumoral effect. Therefore, miRNA-based anticancer treatments can be envisioned. This chapter discusses the translational aspects of miRNA research with particular focus on the different delivery methods to over-express/silence specific miRNAs as well as on the clinical trials currently ongoing and using miRNA-based treatments.

**Keywords** microRNA • Translational • Delivery • Profiling • Trials

## 11.1 Introduction

Recently, microRNA (miRNA) research has suggested that these small molecules may harbor a remarkable therapeutic potential in many diseases. These small, 18–25 nucleotide in length, molecules are endogenously processed in the cell through a

D. Dakhallah, Ph.D. • M. Piper, Ph.D.
Davis Heart and Lung Institute, The Ohio State University, Columbus, OH, USA

Wexner Medical Center at the Ohio State University, Columbus, OH, USA

Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Medical Oncology, The Ohio State University Columbus, Columbus 43210, OH, USA

S.P. Nana-Sinkam, MD, FCCP. (✉)
Davis Heart and Lung Institute, The Ohio State University, Columbus, OH, USA

James Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA

Wexner Medical Center at the Ohio State University, Columbus, OH, USA

Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Medical Oncology, The Ohio State University Columbus, Columbus 43210, OH, USA
e-mail: Patrick.Nana-Sinkam@osumc.edu

M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8_11,

highly organized and orchestrated set of events that integrates polymerases, ribonucleases, trafficking binding proteins, directing proteins to the ‘UTRs, and modifying/stabilizing proteins. All of these factors contribute to the eventual regulation of endogenous mRNA transcripts. A multitude of studies have demonstrated global dys-regulation of miRNAs across both solid and hematological malignancies. In addition, studies focused on select miRNA: target relationships have shown that miRNAs alter processes fundamental to tumor initiation and progression. While the exact mechanisms of biogenesis, trafficking, and mRNA–miRNA silencing remains the subject of intense investigation, it is increasingly apparent that their patterns of expression have the potential to inform clinical decision making. The manner by which endogenous “fine-tuning” of the genome can be leveraged for therapeutic benefit represents an emerging area of investigation.

Given their ability to alter several cellular, molecular and biological processes, miRNA modulation in a biological system represents an ideal avenue for tuning homeostasis in the bioenvironmental condition. However, the path from laboratory based miRNA investigation to in vivo manipulation and eventual human application is full of several challenges. The ability for miRNAs to simultaneously regulate multiple biological pathways while attractive also contributes to their complexity as potential therapeutics. Nevertheless, we have observed that miRNAs can inform clinical decision making beyond therapeutic delivery. miRNA signatures may function as diagnostic, prognostic and therapeutic biomarkers and are actively being investigated in the setting of several ongoing clinical trials. In addition, while still early, select, well-studied miRNAs such as miR-122 have reached clinical trial as directed therapies. Thus, there are clear translational applications for miRNAs to the study of cancer. In this chapter, we review the translational implications of miRNA biology in cancer with a focus on defining miRNA signatures for clinical use, addressing the challenges and opportunities in both human and animal miRNA delivery and lastly, discuss the state of miRNA in clinical trials. We should also recognize that there are several other members of the ncRNA family including lncRNA, Ultraconserved regions (UCRs), and small nucleolar RNAs (snoRNAs) that contribute to fundamental biology. Given that most investigations of these members remain at the in vitro and early in vivo stages, they are not the focus of the review.

## 11.2 Global miRNA Profiling and Clinical Decision Making

One of the earliest observations in miRNA biology was that miRNA are globally altered in malignancy [1]. Such initial observations have lead to a series of both global and single functional miRNA studies all focused on elucidating the role for these molecules in tumor biology. Investigators have noted that select miRNAs are either up-regulated or down-regulated in solid and hematological malignancies and tend to be localized to fragile regions of the human genome [2]. Thus, miRNAs have been described as either functioning as disease specific tumor suppressors,

oncogenes or in select cases, both [3]. Similarly to the early days of transcriptome profiling, several studies have applied miRNA signatures for the purposes of identifying a biomarker for diagnosis, prognosis, or therapeutic response [4]. For example, miRNA profiling studies have identified signatures in lung cancer, breast cancer, colorectal cancer, lymphoma, and acute leukemia to name a few and also suggest that malignancies may be classified based on miRNA signature [5–11]. Importantly, global alterations in miRNA processing have also been associated with clinical outcomes. For example, Karube et al. demonstrated that reduced expression of the key processing endonuclease Dicer was associated with poor outcome and poorly differentiated lung tumors [12]. A more recent study by Khoshnow et al. showed that loss of Dicer expression correlated with higher histological grade and recurrence [13]. Faggad et al. showed that reduced Dicer within colorectal cancer correlated with shorter survival [14].

However, miRNA studies are susceptible to pitfalls similar to those of other platforms including reproducibility, determination of ideal assay for measurement and need for optimal bioinformatics analysis [15]. Despite these concerns, investigators have identified several miRNAs including miR-155, let-7, miR-21, and miR-34 for example as miRNAs fundamental to tumor development. These miRNAs are now being integrated into clinical trials as biomarkers for clinical decision making.

### *11.2.1 Noninvasive miRNA as Biomarkers in Cancer*

A multitude of studies have demonstrated that miRNAs are detectable by less invasive means including circulation, sputum and urine. However, the circulation remains the best studied noninvasive compartment for miRNA detection [16, 17]. Distinct circulating miRNA signatures have been identified in cancers including lung (miR-21, miR-210, and miR-486-5p) and breast (miR-195 and let-7a) [18–20]. However, these signatures have yet to reach clinical application. A recent elegant study conducted by Boeri et al. utilized a circulating miRNA signature as a predictive biomarker in lung cancer screening. The investigators were able to identify a miRNA pattern that could predict the development of lung cancer among a group of individuals at high risk for the development of lung cancer [21].

miRNA localization in the circulation must also be considered when utilizing them as potential biomarkers. We now know that patterns for miRNA expression may vary between serum, peripheral blood mononuclear cells, and whole blood and [22] each of these compartments may carry differing yet important information. This is an important consideration during the development of any circulating miRNA biomarker for clinical application. Recently, exosomes have emerged as important carriers of miRNA as well as a means for intercellular communication [16, 23–28]. In addition to miRNA, these membrane bound particles (30–100 nm) may harbor lipids, mRNA or proteins [29]. miRNA signatures within exosomes have been identified in renal cell, ovarian, hepatocellular carcinoma and glioblastoma [26, 27, 30, 31].

Studies examining global miRNA expression are prone to biases similar to those observed in mRNA analysis. These biases may be driven by the starting materials (blood, primary tumor, fresh, frozen, paraffin embedded), platform for analysis, and strategy for validation [15, 32]. Several platforms are currently utilized including hybridization, PCR, and most recently sequenced based and HITS-CLIP analyses [33–35]. However, investigators have yet to reach consensus on the most suitable detection method. What is increasingly clear is that in order for the profiling of miRNAs to reach clinical application, it will have to be conducted and integrated within the context of a "biological system" that incorporates other components of the human genome coupled with robust informatics.

## 11.3 Target Polymorphisms and miRNA Function

The loss of gene function or tumor suppressive properties can be reflective of increased miRNA expression or mutations in seed sequences generating a new miRNA binding site within the mRNA 3′ UTR. Furthermore, a mutation while rare or a single nucleotide polymorphisms (SNP) in the 3′ UTR of genes do occur to alter miRNA seed sequences and may inhibit or enhance miRNA–mRNA interactions. Recent studies identified mutations in the 3′ UTR for several candidate cancer genes in ovarian breast and colon cancers [36–38]. While often no changes in gene expression are apparent for several of the mutated genes, investigators have yet to rule out transcriptional repression through the analysis of protein expression. Furthermore, in the case of ovarian cancers investigators identified somatic mutations in 6 % of the ovarian tumors in 5 miRNA genes. Notably, 4 of the miRNA mutations occurred in the hairpin loop, while one was present in the mature miRNA. Since proteins bind to the hairpin loop to regulate miRNA stability, it is likely that these mutations alter the miRNA expression. However, the functional consequences of such mutations are still unknown in ovarian cancer.

A study by Knotorovich and colleagues investigated changes in miRNA binding sites in genes associated with breast cancer. This study further expanded the concept of altered miRNA binding sites in a set of genes known to be associated with breast cancer patients positive for BRCA1 and BRCA2 gene mutations [39]. Using a computational approach they identified, approximately 17,500 miRNA binding sites in 66 breast cancer related genes that had ~1,000 known single nucleotide polymorphisms (SNPs). Of these SNPs, they reported 4 SNPs that were statistically significant in the risk of developing breast cancer. Of particular note, a SNP in ATF1 gene alters the binding of miR-320 and patients harboring this SNP have a twofold increase in developing not only breast cancer but ovarian cancer. Interestingly, a decrease in miR-320 expression has been noted in patients with breast cancer [40]. Thus, while miR-320 is decreased in breast cancer patients, using miR-320 as treatment may not be beneficial for all patients due to the presence of SNPs in the miRNA seed sequence in genes associated with breast cancer. Thus, other miRNAs that target these specific genes should be considered.

## 11.4 Chemotherapeutics and miRNA

miRNAs may also be used to guide decision making in the use of chemotherapeutic agents. Given that both solid and hematological tumors are heterogeneous, it is unlikely that any single line of agent will be uniformly successful. Given the propensity for miRNAs to target multiple pathways, patterns of their expression may be utilized to guide the choice of traditional agent. In addition, miRNA manipulation may also be used to augment the effects of chemotherapeutic agents. For example, global profiling studies may distinguish "responders" versus "nonresponders." Ferracin et al. determined that a miRNA signature could distinguish CLL patients who either responded or did not respond to fludarabine [41]. In vitro, miRNAs can be used to sensitize cancer cells to an agent. Manipulation of miR-1 which is usually down-regulated in several solid malignancies, sensitized lung cancer cells to the commonly used doxorubicin [42]. *miR-155* knockdown rendered breast cancer cells sensitive to chemotherapy through regulation of FOXO3a [43]. *miR-21* has been implicated in several solid tumors. In particular miR-21 is over-expressed in non small cell lung cancer (NSCLC). A recent investigation determined that elevated miR-21 expression correlated with platinum based chemoresistance in NSCLC patients and that in vitro reduction could sensitize lung cancer cells [44]. A separate study by Garofalo et al. demonstrated that miRNA including miR-30b, miR-30c, miR-221, and miR-222 contribute to the acquired resistance to tyrosine kinase inhibitors that is often observed in NSCLC [45].

## 11.5 Therapeutic Delivery of miRNA

### *11.5.1 General Approach to miRNA Manipulation*

Fundamentally, miRNA manipulation is based on either selective silencing or augmentation of miRNA expression followed by careful examination for target expression and phenotypic effect. Thus two approaches exist for developing miRNA-based therapeutics: miRNA antagonists and miRNA mimics [46]. One would use miRNA mimetics or antagonists (antagomirs or anti-miRs) to replete or knockdown miRNA expression, respectively. While each modality uses similar strategies, there are some unique approaches to the molecular composition and delivery. It has been postulated that the therapeutic application of targeting miRNAs should solely be based and their expression in disease tissue and not the gene target expression [47]. While this logic is simplistic in nature, knowing a set of genes targeted by a specific miRNA may enable one to use a combinatorial miRNA therapeutic approach. As such, this philosophy has been suggested to enhance the therapeutic effect in malignancies by minimizing resistance to therapy [48]. miRNA antagonists are generated to inhibit endogenously expressed miRNA that are over-expressed in disease tissues. Examples of miRNA inhibitors are anti-miRNAs, locked nucleic acids, and

antagomiRs. This approach is comparable to other inhibitory therapeutic methods that are focused on a single gene inhibition and follows machinery similar to that of short interfering RNAs (siRNA). These chemically engineered antagonists are 20 (nucleotides) nt in length designed to complement target mature miRNA with high affinity and specificity causing silencing of expression and activity [49]. This often results in de-repression of the target mRNA. Since the binding between miRNA and antagomiR has irreversible complementary, it fails to be processed by RISC and undergoes degradation. One potential concern is that miRNA silencers may act as exogenous miRNAs and bind to other mRNAs causing unplanned side effects.

### *11.5.2 Design of miRNA Mimetics and Antagomirs*

The use of either miRNA mimetics or antagomirs involves short RNA molecules. Much of our understanding in the design and delivery of these molecules uses the short interfering (si) RNA technology. However, each of these molecules has specific characteristics and requirements to ensure their stability and function in vivo. To replete miRNA expression, short single stranded (ss) RNA molecules are used. These molecules can contain the same sequence and be perfectly complementary to the mature miRNA. While these mimics are not recognized as foreign to the body, they are 100–1,000-fold less potent than double stranded (ds) mimics [47]. Notably, in this context, mismatches are acceptable in the complementary strand. Improved stability, activity and half-life of the miRNA mimetics can be improved by using sugar and phosphate modifications including 2′-O-methyl, 2′F, 2′$NH_2$, 2′H, phosphorothioates and locked nucleic acids (LNA) [47]. If using a ds mimic, these modifications can be incorporated in both strands. Notably, terminal modification to the antisense strand of the mimic such as inverted bases, biotin, and alkyl groups do not alter the miRNA function. However, a tenfold increase in effectiveness has been reported in the miRNA mimic when a 2′-modified nucleotide is incorporated in the sense miRNA strand and this modification is added to the flanking nucleotides in the antisense strand.

Antisense oligonucleotides that are either ss or ds can be effective in knocking down miRNA expression. These molecules act as a sink to bind the mature miRNA and can be partially or completely complementary to the miRNA. As with mimics, modifications of the antagomir increases hybridization with its miRNA target, stability, and cellular uptake. Modifications of 2′-O-methyl, 2′-O-methoxy ethyl, and LNA are advantageous and as well increase length. Furthermore, these modifications lead to optimal systemic delivery not requiring an adjuvant to increase cellular uptake. However, unlike the miRNA mimics, the miRNA antagonists can be perceived as foreign by the immune system much like siRNA molecules.

There are three types of modified anti-miRNA oligonucleotides (AMOs): 2-OH residues of the ribose by 2′-O-methyl modified oligonucleotides (2′-OMe), 2′-O-methoxyethyl (2′-MOE), and locked nucleic acid (LNA) [50, 51]. Recently, many studies have shown that in vivo targeting of miR-122 by 2′-MOE ASO intraperitoneally resulted in degradation of the miRNA and thus modulation of many

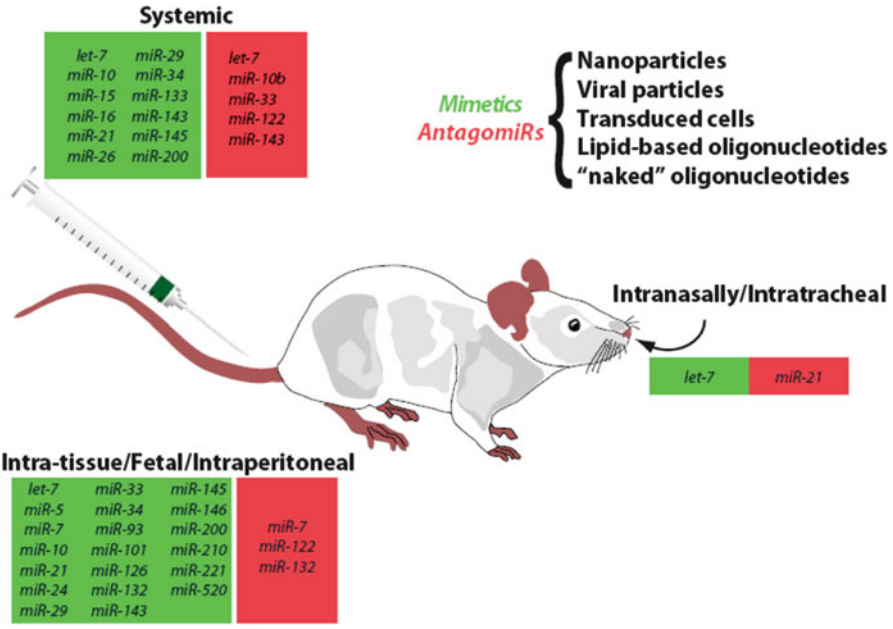


**Fig. 11.1**

genes involved in lowering cholesterol levels and fatty acid in plasma of normal mice [52, 53]. In a diet-induced obesity mouse model, miR-122 inhibition increased fatty-acid oxidation, which led to significant improvement in liver steatosis [54]. Alternatively, miRNA replacement therapy may involve the use of miRNA mimics with a goal of restoring miRNAs whose loss of function is believed to contribute to biological effect. Replacing miRNAs leads to reactivation of biological pathways and restoration of the normal cellular homeostasis [47]. For example, let-7 miRNA is known as a tumor suppressor in several solid malignancies through negative regulation of multiple oncogenes including RAS, and cell cycle promoters [52, 53]. Intratumoral injection of let-7b reduces tumor growth in a K-ras-dependent mouse model of non small cell lung cancer (NSCLC).

One obstacle to in vivo miRNA repletion is the lack of prolonged stability. Adjuvant chemical modification to the backbone of the therapeutic miRNAs has been shown to enhance and optimize miRNA activity, prevent their degradation in biofluids and allow for target tissue specificity. These modifications facilitate specific targeting for miRNA delivery. Cholesterol-conjugated single-stranded RNA (ssRNA) has been used to modulate miRNA expression to be directly delivered in vivo with stable, high efficiency and specificity [54, 55]. Several strategies are now available to enhance in vivo stability. These include 2′F,2′O-Me and 2′H substitutions of the RNA backbone to increase its stability while not affecting its efficiency [56, 57].

Several approaches to in vivo miRNA delivery have been investigated (Fig. 11.1). The most common effective dissemination route for most miRNA therapeutic

delivery is systemic. In addition, there are multiple potential carriers that can enhance miRNA delivery to target diseased tissue. These should be carefully selected as inappropriate selection of a delivery vector can negatively affect gene activity (silencing or re-expression) leading to further off-target effects. Carrier systems are particularly advantageous because both pharmacokinetics and physical size can be altered by differents type of delivery systems resulting in variability in half-life [49, 58].

### *11.5.3 Vector Systems for miRNA Delivery*

Several systems exist for miRNA delivery in which miRNA genes can either be over-expressed or silenced. Vector based delivery represents one modality. Modified lentiviral, retroviral, or adenoviral vectors and adenoviral associated viral vectors represent the most efficient approach in gene delivery [59]. Currently, many viral gene therapy clinical trials rely on adenoviruses and retroviruses. Additional viral vectors have been used such as lentiviruses, pox viruses, adeno-associated viruses (AAV), alphaviruses, and herpes viruses. These viral vectors differ in gene delivery efficiency, permeability, stability, duration of genomic changes, and vital parameters such as long-term expression, efficacy, toxicity, targeting specificity, ability to regulate gene expression and tolerability. For example, viral vector therapeutic treatment requires re-administration in growing cell populations to prevent any side effects of viral infection.

Currently, more that 2,500 gene therapy based trials are in progress in cancer and cardiovascular disease. However, trials involving miRNA delivery are limited to those targeting miR-122. Lentiviral vectors are another very promising method to deliver a short-hairpin RNA (shRNA) construct in many mammalian systems and in treating cancer. For example, lentiviral-shRNA delivery of let-7 in mice resulted in a downregulation of activated Ras oncogene and suppressed tumor growth [53, 60]. In addition, introducing smad3-targeted shRNA using a lentiviral vector resulted in regeneration of satellite cells to repair-injured muscles and old tissues [61]. Reintroduction of the let-7 tumor suppressor miRNA by adenoviral or lentiviral delivery provided proof-of-concept for let-7 novel miRNA replacement therapy to treat lung cancer [62]. A similar approach was effectively established for liver and metastatic prostate cancer [63, 64]. An attractive aspect to viral delivery lies in their ability to infect and transducer more than one cell, both dividing and nondividing cells and different cell types as well, yielding a prolonged expression of the therapeutic genes [46].

However, there are also disadvantages. A lack of specificity of viral infection can cause possible tRM leading to DNA mutations. Insertional mutagenesis or abnormal gene expression can be triggered by viral infections, causing toxicity and immunogenicity. In addition, viral vectors are still potentially immunogenic even if they lack the pathogenic effect.

### *11.5.4 Nanoparticle Delivery*

Nanoparticles (NPs) are engineered small particles with <100 nm dimension harboring robust kinetic stability and rigid morphology [65]. There are several NP properties that make them particularly suited for medical application. These properties include having a relatively large functional surface that allows for superior binding, absorption and carrying of other compounds. Engineered NP composition also varies; NPs may be composed of lipids, phospholipids, lactic acid, dextran or have chemical characteristics similar to many polymers, silica carbon and metals. For drug delivery, a biodegradable NP is essential for the transport and release of therapeutic drug in efficient specific way.

Lipid Nanoparticles (LNPs) are safe and effectively used in drug delivery due to the biocompatibility of their lipid matrix. The most important advantages of LNPs are their ability to be used in topical drug delivery, the ability to modulate drug release and solubility and stability [66, 67]. LNPs can be delivered by several routes including intra muscular, subcutaneous and intravenous administration due to their small particle size, which enable LNPs to target particular organs. In addition, oral aqueous dispersion, tablets, pellets, capsules or powders spray-drying or freeze-drying are all viable options for delivery [67]. Recently, LNPs have been utilized for ocular delivery to enhance corneal bio-adhesion and drug permeation [67, 68]. There are, however, several potential obstacles to using nanoparticles including toxicity, optimizing delivery, uptake and specific targeting which directly affects the therapeutic index [66]. Another method of in vivo delivery of miRNA—based therapy using nanoparticles therapeutic delivery is facilitated by the use of natural products such as atelocollagen, a cationic polymer that electrostatically interacts with miRNAs. This approach leads to the formation of smaller nanoparticles that can increase specificity by enhancing its permeability and retention [69].

Recently, proof of concept for in vivo miRNA delivery was demonstrated using a murine model of NSCLC and lymphoma [70, 71]. Trang et al., showed that using a neutral lipid emulsion as a carrier to deliver miRNA (let-7) by either direct intratumoral injection or by systemic injections resulted in reduced tumor growth and repression of many oncogenes. In addition, this method of therapeutic delivery, caused less accumulation of miRNAs in the liver or kidney suggesting high specificity for the lung [70, 72].

## 11.6 Human Delivery

### *11.6.1 Preclinical Considerations for miRNA Therapy in Humans*

There are several factors to consider prior to the design and initiation of miRNA therapies in human. While results in animal models are encouraging, translating these findings to the clinic remains a challenge. While targeting a specific miRNA

may be optimal, there are conditions where expression of the enzymes Drosha and DICER (important in miRNA processing) are decreased. Notably, in ovarian, breast, and lung cancers, poor prognosis is associated with decreased expression of these enzymes [12, 13, 73]. It should be noted that in many malignancies, there appears to be an overall decrease in miRNA expression. Alternatively, it is possible that a master key miRNA regulator may exist that is critical in maintaining the expression of Drosha and Dicer. While in cancers, one would want to replete these miRNAs as treatment, it may be beneficial to use several miRNAs as a combinatorial therapy. Therefore, these observations support the concept that combinatorial therapies may be more beneficial [48]. As many studies in mice have shown, miRNAs are stable for long periods when injected intravenously. However, in humans, there are several limitations to the systemic administration of unmodified naked miRNAs. These limitations include renal clearance and rapid degradation secondary to proteases in the serum and other body fluids [47, 74, 75]. Additional issues encountered in miRNA therapy include failure to cross the capillary endothelium for large particles, uptake of the miRNAs by tissue macrophages, inability to enter the extracellular matrix, inefficient cellular endocytosis, and the inability for release by the endosome within the cell. Therefore, numerous studies have investigated a variety of mechanisms, modifications and adjuvants to overcome these obstacles and enhance miRNA uptake and biological efficacy. Wu and colleagues have reported that utilization of a cationic lipid increases the stability and permeability of the miRNAs in the lung when injected systemically compared to transfection reagents [78]. In that particular study, as expected, the liver compared to the lung preferentially took up the lipoplex-miRNA complexes. Furthermore, when miRNAs were complexed with a transfection reagent as opposed to the lipoplexes, they were primarily localized to the kidneys.

It is also important to note, that a single miRNA can target upwards of 100 genes [1]. Thus, systemic administration of miRNAs may not only be problematic for stability but may elicit off target effects. Just as modifications to the miRNA mimic or antagonist can increase their stability, additional optimizations can overcome the challenges of tissue specificity. Several strategies have been developed such as viral vectors to target specific tissues. Intratumor administration appears to be beneficial for breast and prostate cancers among others. Notably, inhalation and intranasal administrations may efficiently deliver miRNAs and antagomirs to lung cancers. Similar to lipoplexes, Trang et al. have reported that complexing miRNA mimics with a cationic lipid solution enhances the uptake of the let-7b miRNA injected directly into tumors. As a result, repleting let-7b using an intratumor delivery route decreased tumor size in mice. Notably, administration of miR-34 in mice with NSCLC effectively reduces tumor growth [70]. Interestingly, fluorescent naked miRNAs delivered intranasally are efficiently and quickly taken up in murine lung tissue suggesting another potential method to deliver miRNAs to the lung [77].

As previously mentioned, systemic delivery of miRNA mimics or antagomirs may lead to unwanted effects. However, systemic treatment with these molecules is still a desirable route of administration due to the ease of delivery. Thus, many studies have and currently are exploring tissue specific targeting using viral particles or

lipid-based expression vectors. In fact, the use of adeno-associated viruses has proven successful to target lung cancer in murine models [76]. Others have reported that intranasal delivery of miRNAs encapsulated in viral particles can reduce tumor burden in murine lung cancer models. However, efficient uptake of the expression vector has some limitations due to their size.

## 11.7 miRNAs in Clinical Trials

On ClinicalTrials.gov website, currently there are 114 listed clinical trials involving miRNAs (Table 11.1). Many of these trials involve using miRNAs as biomarkers for a wide variety of diseases. Additional clinical trials are investigating the expression of miRNA processing machinery such as Dicer and Drosha since, as previously mentioned, these molecules are decreased in a wide variety of cancers.

Currently, Santaris/Mirna Therapeutics Inc. is at the forefront using an antagomir for miR-122 to treat hepatitis [59, 77, 78]. This company recently completed a Phase I trial to examine the safety, tolerated dose, Pharmacokinetics and Pharmacodynamics of the miR-122 antagomir in healthy men. More recently, the company has initiated a Phase II study using the LNA-modified miR-122 antagonist, marketed as Miravirsen, to treat patients infected with hepatitis C virus. Currently, Mirna Therapeutics has seven other miRNAs in various stages for therapeutics including, miR-34, let-7b, and miR-16.

While the vast majority of these trials investigate cancers miRNAs are being evaluated, other diseases including diabetes, asthma, inflammatory bowel disease, coronary heart disease, and sepsis. In addition, since miRNA production is conserved among invertebrates as well as viruses, it is no surprise that viruses produce miRNAs. In fact, the majority of the viral miRNAs do not target the viral genome but rather the host. These viral miRNAs can recognize mRNA targets that are critical for the viral replication and hijack the immune response to evade detection. It is possible that understanding this miRNA expression pattern may help to identify latent and active viral infections. In fact, there are several clinical trials listed at ClinicalTrials.gov to examine miRNA expression patterns in virally infected individuals.

## 11.8 Summary

miRNAs represent small noncoding molecules that have the capacity for regulating fundamental biology. Successful in vitro and animal modeling has led to the possibility of integrating miRNA biology into clinical diagnostics and therapeutics. The era of using a clinical miRNA signatures to guide traditional therapeutic decision making is upon us. While miRNAs as biomarkers are rapidly entering the clinics, using them as therapeutics remains more of a work in progress. The ongoing clinical trials using miR-122 in hepatitis are encouraging and it is anticipated that additional

**Table 11.1** Active clinical trials

| Study type | Condition/disease | Sample type | Objectives/patient cohort | Status |
|---|---|---|---|---|
| Therapeutic | Hepatitis C | N/A | Antagomir to miR-122 | Unknown |
| Specific mIRNA analysis | GVHD | Blood* | miR-155 and global miRNA Expression | Recruiting |
| | Colorectal carcinomas | Tissue | miR-21 expression in advanced disease | Recruiting |
| | NSCLC and esophagus squamous cell carcinoma | Unknown | miR-326 expression | Ongoing, not recruiting |
| | Prostate adenocarcinoma | Tissue | X-linked miRNAs | Upcoming |
| | Pediatric AML | Bone marrow | miRs-34a, −538e, −193e and −198 | Ongoing, not recruiting |
| | Hypertension and CMV | Blood | hcmv-miR-UL 112 | Unknown |
| | Type 2 diabetic | Neointima | miRNAs-21, −126, −143 and −145 | Recruiting |
| miRNA Profiling | Breast cancer | Tissue | Response to Bevacizumab therapy** | Upcoming |
| | | Blood/tissue | Treatment exposed breast tumors** | Recruiting |
| | | Blood | Response to tamoxifen or hormone therapy** | Upcoming |
| | | Tissue | Invasive lobular carcinoma in situ | Recruiting |
| | | Tissue | Unilateral breast cancer of epithelial origin | Recruiting |
| | | Tissue | HER2 Negative breast cancer** | Recruiting |
| | Hematological malignancies and disorders | Blood/CSF | Pediatric leukemia, lymphoma and CNS Tumors** | Ongoing, not recruiting |
| | | Blood | B-cell acute lymphoblastic leukemia (ALL) | Upcoming |
| | | Blood/bone marrow | AML** | Recruiting |
| | | Tissue | Treatment resistant diffuse large B-cell lymphoma | Recruiting |
| | | Blood/bone marrow | Comparison of pediatric AML and ALL | Ongoing, not recruiting |
| | | Blood/bone marrow | Relapse Multiple Myeloma | Ongoing, not recruiting |
| | | Bone marrow | Pediatric Relapsed TEL/AML 1-positive ALL | Recruiting |
| | | Bone marrow | Pediatric AML | Ongoing, not recruiting |
| | | Blood/bone marrow | Pediatric high-risk ALL | Recruiting |

(continued)

**Table 11.1** (continued)

| Study type | Condition/disease | Sample type | Objectives/patient cohort | Status |
|---|---|---|---|---|
| | | Blood/bone marrow | Acute myelogenous leukemia (AML) | Recruiting |
| | | Unknown | Pediatric ALL* | Unknown |
| | | Blood/tissue/bone | Down syndrome and AML or | Ongoing, not recruiting |
| | | Marrow | Myeloproliferative disease | |
| | | Unknown | Multiple myeloma: newly diagnosed** | Recruiting |
| | | Blood/bone marrow | Refractory AML** | Unknown |
| | | Blood | Chronic lymphocytic leukemia** | Recruiting |
| | | Blood | Relapsed or refractory T-cell lymphoma** | Recruiting |
| | | Blood | AML patients who failed standard-of-care treatment | Upcoming |
| | | Blood | Multiple myeloma: untreated previously diagnosed** | Ongoing, not recruiting |
| | | Blood/skin/bone marrow | Hematological Disorders | Recruiting |
| | | Blood | Umbilical cord NK cells | Unknown |
| | CNS Tumors | Tissue/blood | Pre-treatment miRNAs in Neuroblastoma | Recruiting |
| | | Tissue | Neurofibromatosis Type 1 | Recruiting |
| | | Blood/urine/CSF | Pediatric CNS tumors** | Recruiting |
| | | Tissue | Renal, Brain and CNS Rhabdoid Tumors | Upcoming |
| | | Blood/tissue | Gliomas | Recruiting |
| | | Blood/CSF | Pediatric CNS tumors | Recruiting |
| | | Tissue | Pediatric Neuroblastoma | Unknown |
| | | Tissue/blood | Pediatric Adrenocortical Tumors | Ongoing, not recruiting |
| | | Tissue | CNS tumors | Recruiting |
| | | Tissue | High and low risk neuroblastoma | Ongoing, not recruiting |
| | | Blood/tissue/bone marrow | Pediatric neuroblastoma, osteosarcoma, retinoblastoma, Ewing sarcoma and soft tissue sarcoma | Recruiting |
| | Other malignancies | Blood | Ovarian carcinoma** | Recruiting |
| | | Tissue | Pediatric ovarian and testicular tumors | Ongoing, not recruiting |
| | | Tissue | Endometrial cancer: stage I–III | Unknown |

(continued)

**Table 11.1** (continued)

| Study type | Condition/disease | Sample type | Objectives/patient cohort | Status |
|---|---|---|---|---|
| | | Tissue | Renal cell carcinoma | Unknown |
| | | Tissue/blood | Renal cell carcinoma** | Recruiting |
| | | Tissue | Hepatocellular carcinoma | Recruiting |
| | | Tissue/blood/bone | Chemoresistant and metastatic prostate cancer | Recruiting |
| | | Unknown | High risk prostate cancer | Recruiting |
| | | Blood | Hormone-resistant prostate cancer** | Recruiting |
| | | Tissue | Esophago-gastric Adenocarcinomas** | Recruiting |
| | | Tissue/blood | Metastatic colon cancer | Recruiting |
| | | Tissue/blood | Thyroid tumors benign and metastatic | Invitational |
| | | Tissue | Anaplastic thyroid carcinomas** | Recruiting |
| | Lung cancers and diseases | Tissue/BAL | Asthmatic patients | Recruiting |
| | | Blood | Pulmonary arterial hypertension | Recruiting |
| | | Blood | Acute asthma exacerbation | Recruiting |
| | | Blood | Premenstrual asthma | Recruiting |
| | | Unknown | Advanced NSCLC** | Recruiting |
| | | Blood | NSCLC: stage IIIB-IV** | Ongoing, not recruiting |
| | | Tissue | Chest cancers: lung, esophagus. Gl, mesothelioma, and breast cancer** | Recruiting |
| | Coronary heart disease | Unknown | Angina** | Recruiting |
| | Organ transplantation | Blood/urine | Kidney transplant recipients | Recruiting |
| | | Unknown | CMV-infected organ transplant patients | Ongoing, not recruiting |
| | | Tissue/blood/urine | Donor kidney | Recruiting |
| | | Tissue/blood | Liver allograft rejection | Ongoing, not recruiting |
| | ICU/sepsis | Blood | Pulmonary and abdominal sepsis | Upcoming |
| | | Blood | Circulating septic miRNAs | Recruiting |
| | | Blood | Trauma-associated sepsis | Upcoming |
| | | Blood | ICU acquired paresis | Recruiting |
| | | Blood/BAL/CSF/urine | SIRS and sepsis | Recruiting |
| | | Blood | Type 1 diabetes | Unknown |

(continued)

**Table 11.1** (continued)

| Study type | Condition/disease | Sample type | Objectives/patient cohort | Status |
|---|---|---|---|---|
| | Diabetes/obesity | Unknown | Type 2 diabetes | Recruiting |
| | | Blood | Type 2 diabetes** | Upcoming |
| | | Blood | Obesity | Recruiting |
| | | Blood | Multiple sclerosis | Invitational |
| | Multiple sclerosis | Blood/CSF/ urine/ saliva | Early multiple sclerosis and clinical isolated syndrome | Recruiting |
| | Other diseases | Tissue | Endometrium after ovarian hormone stimulation | Ongoing, not recruiting |
| | | Tissue/blood | Ocular diseases | Recruiting |
| | | Tissue | Psoriasis** | Recruiting |
| | | Blood/urine | Chronic kidney disease | Ongoing, not recruiting |
| | | Tissue | Lower urinary tract dysfunction | Recruiting |
| | | Tissue/blood | Inflammatory bowel disease | Unknown |
| | | Tissue | Irritable bowel syndrome | Ongoing, not recruiting |
| | | Blood | Sleep/aging | Ongoing, not recruiting |

*BAL* bronchoalveolar fluid, *CNS* cerebral nervous system, *CSF* cerebral spiral fluids, *NSCLC* non small cell lung cancer

miRNAs may reach such application. However, the complexities of miRNAs as true therapeutics in human cancer are not lost on investigators. Further preclinical testing will be required prior to making miRNA-based therapeutics a reality.

## References

1. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6(11):857–66.
2. Croce CM. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet. 2009;10(10):704–14.
3. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer. 2006;6(4):259–69.
4. Takamizawa J et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res. 2004;64(11):3753–6.
5. Yanaihara N et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9(3):189–98.
6. Liu CG et al. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc Natl Acad Sci U S A. 2004;101(26):9740–4.
7. Calin GA et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc Natl Acad Sci U S A. 2004;101(32):11755–60.

8. Lu J et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043): 834–8.
9. Iorio MV et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005;65(16):7065–70.
10. Calin GA et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med. 2005;353(17):1793–801.
11. Volinia S et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006;103(7):2257–61.
12. Karube Y et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci. 2005;96(2):111–5.
13. Khoshnaw SM et al. Loss of Dicer expression is associated with breast cancer progression and recurrence. Breast Cancer Res Treat. 2012;135(2):403–13.
14. Faggad A et al. Down-regulation of the microRNA processing enzyme Dicer is a prognostic factor in human colorectal cancer. Histopathology. 2012;61(4):552–61.
15. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. Nat Rev Genet. 2012;13(5):358–69.
16. Hunter MP et al. Detection of microRNA expression in human peripheral blood microvesicles. PLoS One. 2008;3(11):e3694.
17. Mitchell PS et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105(30):10513–8.
18. Heneghan HM et al. Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. Ann Surg. 2010;251(3):499–505.
19. Audimoolam VK et al. Predictors of graft and patient survival in subjects undergoing transplantation for hepatitis C virus infection. Liver Transpl. 2010;16(4):536–7.
20. Shen J et al. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. BMC Cancer. 2011;11:374.
21. Boeri M et al. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc Natl Acad Sci U S A. 2011;108(9): 3713–8.
22. Arroyo JD et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci U S A. 2011;108(12):5003–8.
23. Umezu T et al. Leukemia cell to endothelial cell communication via exosomal miRNAs. Oncogene. 2012;32(22):2747–55.
24. Palma J et al. MicroRNAs are exported from malignant cells in customized particles. Nucleic Acids Res. 2012;40(18):9125–38.
25. Fabbri M et al. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. Proc Natl Acad Sci U S A. 2012;109(31):E2110–6.
26. Kogure T et al. Intercellular nanovesicle-mediated microRNA transfer: a mechanism of environmental modulation of hepatocellular cancer cell growth. Hepatology. 2011;54(4):1237–48.
27. Grange C et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. Cancer Res. 2011;71(15):5346–56.
28. Meckes Jr DG et al. Human tumor virus utilizes exosomes for intercellular communication. Proc Natl Acad Sci U S A. 2010;107(47):20370–5.
29. Valadi H et al. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9(6):654–9.
30. Skog J et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol. 2008;10(12):1470–6.
31. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. Gynecol Oncol. 2008;110(1):13–21.
32. Pritchard CC et al. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila). 2012;5(3):492–7.
33. Jiang J et al. Real-time expression profiling of microRNA precursors in human cancer cell lines. Nucleic Acids Res. 2005;33(17):5394–403.

34. Chi SW, Hannon GJ, Darnell RB. An alternative mode of microRNA target recognition. Nat Struct Mol Biol. 2012;19(3):321–7.
35. Chi SW et al. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 2009;460(7254):479–86.
36. Chen X et al. Identification of ten serum microRNAs from a genome-wide serum microRNA expression profile as novel noninvasive biomarkers for nonsmall cell lung cancer diagnosis. Int J Cancer. 2012;130(7):1620–8.
37. Smits KM et al. A let-7 microRNA SNP in the KRAS 3′UTR is prognostic in early-stage colorectal cancer. Clin Cancer Res. 2011;17(24):7723–31.
38. Zhang L et al. Functional SNP in the microRNA-367 binding site in the 3′UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. Proc Natl Acad Sci U S A. 2011;108(33):13653–8.
39. Kontorovich T et al. Single nucleotide polymorphisms in miRNA binding sites and miRNA genes as breast/ovarian cancer risk modifiers in Jewish high-risk women. Int J Cancer. 2010;127(3):589–97.
40. Yan LX et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA. 2008;14(11):2348–60.
41. Ferracin M et al. MicroRNAs involvement in fludarabine refractory chronic lymphocytic leukemia. Mol Cancer. 2010;9:123.
42. Nasser MW et al. Down-regulation of micro-RNA-1 (miR-1) in lung cancer. Suppression of tumorigenic property of lung cancer cells and their sensitization to doxorubicin-induced apoptosis by miR-1. J Biol Chem. 2008;283(48):33394–405.
43. Kong W et al. MicroRNA-155 regulates cell survival, growth, and chemosensitivity by targeting FOXO3a in breast cancer. J Biol Chem. 2010;285(23):17869–79.
44. Gao W et al. MiRNA-21: a biomarker predictive for platinum-based adjuvant chemotherapy response in patients with non-small cell lung cancer. Cancer Biol Ther. 2012;13(5):330–40.
45. Garofalo M et al. EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers. Nat Med. 2012;18(1):74–82.
46. Aagaard L, Rossi JJ. RNAi therapeutics: principles, prospects and challenges. Adv Drug Deliv Rev. 2007;59(2–3):75–86.
47. Bader AG, Brown D, Winkler M. The promise of microRNA replacement therapy. Cancer Res. 2010;70(18):7027–30.
48. Kim M, Kasinski AL, Slack FJ. MicroRNA therapeutics in preclinical cancer models. Lancet Oncol. 2011;12(4):319–21.
49. Akhtar S, Benter I. Toxicogenomics of non-viral drug delivery systems for RNAi: potential impact on siRNA-mediated gene silencing activity and specificity. Adv Drug Deliv Rev. 2007;59(2–3):164–82.
50. Yang M, Mattes J. Discovery, biology and therapeutic potential of RNA interference, microRNA and antagomirs. Pharmacol Ther. 2008;117(1):94–104.
51. Bijsterbosch MK et al. Modulation of plasma protein binding and in vivo liver cell uptake of phosphorothioate oligodeoxynucleotides by cholesterol conjugation. Nucleic Acids Res. 2000;28(14):2717–25.
52. Johnson CD, Esquela-Kerscher A, Stefani G, Byrom M, Kelnar K, Ovcharenko D, et al. The let-7 microRNA represses cell proliferation pathways in human cells. Cancer Res. 2007;67(16):7713–22.
53. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. Cell. 2005;120(5):635–47.
54. van Solingen C, Seghers L, Bijkerk R, Duijs JM, Roeten MK, et al. Antagomirmediated mediated silencing of endothelial cell specific microRNA-126 impairs ischemia induced angiogenesis. J Cell Mol Med. 2009;13:1577–85.
55. Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, et al. Silencing ofmicroRNAs in vivo with “antagomirs”. Nature. 2005;438(7068):685–9.

56. Rettig GR, Behlke MA. Progress towards in vivo use of siRNAs. Mol Ther. 2012;20(3): 483–512.
57. Chiu YL, Rana TM. siRNA function in RNAi: a chemical modification analysis. RNA. 2003;9(9):1034–48.
58. Hollins AJ, Omidi Y, Benter IF, Akhtar S. Toxicogenomics of drug delivery systems: exploiting delivery system-induced changes in target gene expression to enhance siRNA activity. J Drug Target. 2007;15(1):83–8.
59. Elmen J et al. LNA-mediated microRNA silencing in non-human primates. Nature. 2008; 452(7189):896–9.
60. Brummelkamp TR, Bernards R, Agami R. Stable suppression of tumorigenicity by virus-mediated RNA interference. Cancer Cell. 2002;2:243–7.
61. Carlson ME, Hsu M, Conboy IM. Imbalance between pSmad3 and Notch induces CDK inhibitors in old muscle stem cells. Nature. 2008;454:528–32.
62. Trang P, Medina PP, Wiggins JF, Ruffino L, Kelnar K, Omotola M, et al. Regression of murine lung tumors by the let-7 microRNA. Oncogene. 2009;29:1580–7.
63. Takeshita F, Patrawala L, Osaki M, Takahashi RU, Yamamoto Y, Kosaka N, et al. Systemic delivery of synthetic MicroRNA-16 inhibits the growth of metastatic prostate tumors via downregulation of multiple cell-cycle genes. Mol Ther. 2010;18(1):181–7.
64. Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, et al. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. Cell. 2009;137:1005–17.
65. Su J et al. Silencing microRNA by interfering nanoparticles in mice. Nucleic Acids Res. 2011;39(6):e38.
66. Müller RH, Radtke M, Wissing SA. Solid lipid nanoparticles (SLN) and nanostructured lipid carriers (NLC) in cosetic and dermatological preparations. Adv Drug Deliv Rev. 2002;54 Suppl 1:S131–55.
67. Battaglia L, Gallarate M. Lipid nanoparticles: state of the art, new preparation methods and challenges in drug delivery. Expert Opin Drug Deliv. 2012;9(5):497–508.
68. Attama AA, Reichl S, Müller-Goymann CC. Diclofenac sodium delivery to the eye: in vitro evaluation of novel solid lipid nanoparticle formulation using human cornea construct. Int J Pharm. 2008;355(1–2):307–13.
69. Takeshita F, Minakuchi Y, Nagahara S, Honma K, et al. Efiicient delivery of small interfering RNA to bone-metastatic tumers-by using atelocollagen in vivo. Proc Natl Acad Sci U S A. 2005;102(34):12177–82.
70. Trang P et al. Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. Mol Ther. 2011;19(6):1116–22.
71. Babar IA et al. Nanoparticle-based therapy in an in vivo microRNA-155 (miR-155)-dependent mouse model of lymphoma. Proc Natl Acad Sci U S A. 2012;109(26):E1695–704.
72. Wiggins JF, Ruffino L, Kelnar K, Brown D, et al. Development of lung cancer therapeutic based on the tumor suppressor miRNA-34. Cancer Res. 2010;70(14):5923–30.
73. Merritt WM et al. Dicer, Drosha, and outcomes in patients with ovarian cancer. N Engl J Med. 2008;359(25):2641–50.
74. Bader AG. miR-34—a microRNA replacement therapy is headed to the clinic. Front Genet. 2012;3:120.
75. Wu Y et al. MicroRNA delivery by cationic lipoplexes for lung cancer therapy. Mol Pharm. 2011;8(4):1381–9.
76. Esquela-Kerscher A et al. The let-7 microRNA reduces tumor growth in mouse models of lung cancer. Cell Cycle. 2008;7(6):759–64.
77. Elmen J et al. Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. Nucleic Acids Res. 2008;36(4):1153–62.
78. Stenvang J et al. The utility of LNA in microRNA-based cancer diagnostics and therapeutics. Semin Cancer Biol. 2008;18(2):89–102.

# Index



M. Fabbri (ed.), *Non-coding RNAs and Cancer*, DOI 10.1007/978-1-4614-8444-8,