Mansi Arora · Deepak Kaul

Cancer RNome: Nature & Evolution

Springer

Cancer RNome: Nature & Evolution

Mansi Arora • Deepak Kaul

Cancer RNome: Nature & Evolution

Springer

Mansi Arora
Department of Experimental Medicine and Biotechnology
Post Graduate Institute of Medical Education and Research
Chandigarh, India

Deepak Kaul
Department of Experimental Medicine and Biotechnology
Post Graduate Institute of Medical Education and Research
Chandigarh, India

ISBN 978-981-13-1567-1 ISBN 978-981-13-1568-8 (eBook)
https://doi.org/10.1007/978-981-13-1568-8

Library of Congress Control Number: 2018953152

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

About the Authors

Dr. Mansi Arora is currently working as Assistant Professor in Biotechnology, Government P.G. College, Panchkula, Haryana. She completed her PhD from the Department of Experimental Medicine and Biotechnology, Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh, India. She has also served as a Demonstrator in the same department. In her relatively short scientific career to date, she has published several papers and has presented work at various leading international/national conferences and symposia.

Dr. Deepak Kaul is former Senior Professor and Head of the Department of Experimental Medicine and Biotechnology, Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh, India.

He completed his PhD at the prestigious All India Institute of Medical Sciences (AIIMS), New Delhi. He is also founder of the Molecular Biology Unit at PGIMER and has been associated with PGIMER for more than three decades. His main areas of research are molecular medicine, functional genomics, genomic immunity, and cancer biology. Dr. Kaul has published over 150 papers, which speaks the volume about his contributions in the field. He is a fellow/member of various respected international societies, such as the Royal College of Pathologists (London), Indian National Science Academy (India), World Academy of Art and Science, International College of Angiology, New York Academy of Sciences, and the Biochemical Society, London. He is on the editorial board of several reputed international journals, including the *Journal of Antivirals and Antiretrovirals*, *Word Journal of Cardiology*, *Leukemia Research*, and *Molecular and Cellular Biochemistry*. He is also an expert member of several Government of India task force committees.

Preface

During my formative years of understanding biology from organism to organs to tissues to cells to organelles to molecules, I got fascinated by the elegantly simple idea enunciated by Prof. Crick in the year 1958. This simple idea widely known as "Central Dogma" of molecular biology served us well over the years to seek and explore human genome in health and diseased state. Although it was recognized early on by Prof. Crick followed by others that RNA has both a genotype and a phenotype, the exact role of RNA in the complex system of "DNA-RNA-Protein" remains enveloped in the mystery especially when around 90% of the human genomic DNA is transcribed of which 2% is translated into protein and the remaining 98% is noncoding RNAs (ncRNAs). According to the "Central Dogma," it was postulated that only proteins are destined to impart genomic-dynamics through their ability to regulate chromatin dynamics within human cells. However, mounting evidence now exists to support the view that ncRNAs play crucial and critical role in a great variety of cellular processes, including transcriptional regulation, chromatin dynamics, RNA processing and modification, mRNA stability and translation, and even protein degradation and translocation. Hence, it is not unreasonable to assume that human genome dynamics would have been impossible to achieve and sustain without the existence of an "RNome" that holds alternate splice isoforms of all the protein coding genes as well as transcribed abundant ncRNAs (siRNAs, miRNAs, and long noncoding RNAs) with critical regulatory functions within its fold. Needless to mention here that the "RNome" has emerged in the recent years to govern all the dynamic aspects of human cellular genome. A phenomenon that forced my childish instinct to define "RNome" in following poetic fashion:

Patterns that arise from RNome
hold fold to fold the genome
Within the convoluted beads of nucleosome
to create a mysterious home
for life to arise from loam
in the form of a rare polychrome
created by waves of proteome
that rise to fall within cytosome
in tune with script of RNome
death ends in life under dome.

In organismic machine, the person with the disease, the impact of the disease on the person, the way personal characteristics modify disease presentation, and the family and the community the patient comes from are features that engages a clinician at present. However, these features disappear from the view, when the clinician's gaze passes right through the person with disease to focus on the molecular traits of the disease process. Human cancer compromises a group of diseases that involves cellular immortality, unbridled growth, dedifferentiation, and potential for metastases. In the year 1924, Prof. Otto Warburg postulated that the prime cause of cancer is the replacement of aerobic respiration with glycolysis within the normal body cells. Oncology research community, across the globe, has not only accepted this basic postulation of Prof. Warburg but also recognized the addiction of all cancer cells to DNA methylation responsible for silencing of genes involved in cellular death and growth arrest.

The coupling of next-generation sequencing (NGS) platforms with established chromatin technologies has presented us with previously unparalleled view of the human RNome and its implications in oncogenesis. However, appreciation of complexity and plasticity of the "RNome" has dramatically increased over the last few years especially keeping in view the crucial and critical role of miRNAs, within the engine room of cancer, affecting nodal points in cell cycle regulation, genome integrity, stress responses, apoptosis, and metastasis. Although gaps still exist in our understanding of how RNome-plasticity gives rise to "Cancer RNome" thereby ensuring human cells to be and remain cancerous despite the fact that an overwhelming amount of data strongly links deregulated ncRNA expression to the etiology of many cancers.

Chandigarh, India Deepak Kaul

Contents

RNome: Evolution and Nature

1

Abstract

In spite of identical genetic information, different types of the cells in the body perform their own specific functions. It is the differential expression of the same genome that governs this phenotypic diversity of cells and allows development and functioning of complex organisms. The regulation of gene expression has been extensively studied over the years. A key discovery in this regard was that despite being almost entirely transcribed, only 2% of the genome codes for protein. This revelation triggered a global research into the world of non-coding RNAs (ncRNAs). Advancements in the RNA sequencing technologies and methods of studying RNA–RNA/RNA–DNA and RNA–protein interactions are continuously adding to the pool of cellular RNAs. ncRNAs are highly diverse in terms of their structure and function and can be broadly divided on the basis of their (a) size (small or long) or (b) functions (regulatory or housekeeping). This chapter discusses biogenesis, mode of actions, and cellular functions of both coding and non-coding RNAs (small as well as long ncRNAs).

Keywords

Biogenesis · Circular RNAs · RNA function · Small non-coding RNAs · Long non-coding RNAs · RNome

The RNome represents a conceptual RNA-based home of life, believed to evolve in a period of time of primitive Earth's history roughly about four million years ago. In the last five decades or so, this RNome concept has emerged to empower RNA as a plausible precursor to the complex system of DNA-RNA-proteins on which the living state is based. The classical experiments with bacteriophage Qβ, conducted by Spiegelman, showed how viral RNA could evolve as a function of time in response to selection (Mills et al. 1967). This study revealed as to how in vitro

M. Arora, D. Kaul, *Cancer RNome: Nature & Evolution*,
https://doi.org/10.1007/978-981-13-1568-8_1

evolution can occur in the absence of basic unit of life – the cell (Joyce 2007). Subsequently several pioneering findings (Biebricher and Orgel 1973; Biebricher et al. 1985; Robertson and Joyce 1990; Tuerk and Gold 1990; Ellington and Szostak 1990; Beaudry and Joyce 1992) unambiguously demonstrated the evolutionary abilities of RNA and made it difficult to ignore the possibility that the life originated with the RNA molecule. The catalytic repertoire of RNA as well as existence of small non-coding regulatory RNAs added up a new dimension to the contribution of RNome to the evolution and regulation of various biological processes that sustain life in all its dimensions.

1.1 The Journey of RNA: From a Mere Adaptor to Central Regulator

By the early 1950s, DNA had been established to be the carrier of genetic information, and its structure had also been elucidated (Watson and Crick 1953). It was also known that the protein synthesis occurs in the cytoplasm in an RNA-rich milieu, with the link between rRNA and ribosomes as the platform of protein synthesis being given in the mid-1950s (Palade 1955). The researchers, at that time, were posed with an important question as to how the information flows from this linear sequence of nucleotides present in the nucleus to the proteins that are being synthesized in the cytoplasm. It was James Watson who proposed the existence of a coding RNA that carries the information from DNA to the protein synthetic machinery in the cytoplasm, which was subsequently proved experimentally in the year 1961 by Brenner et al. and thus came the concept of messenger RNA (mRNA) (Brenner et al. 1961; Eddy 2001). At around the same time, Francis Crick proposed that an "adaptor molecule" acts as an intermediary between the triplet genetic code and corresponding encoded amino acid. Interestingly, Crick also predicted that the adaptor would be an RNA molecule, which would in fact be evolutionarily preferred over protein as a small specific molecule that recognizes triplet codon simply by base pairing (Eddy 2001). Hoagland and coworkers observed the biochemical nature of these adaptors –termed the transfer RNAs (Hoagland et al. 1958) – recognized as the second genetic code, linking nucleic acid sequence to the amino acid code. Thus, RNA came to be known to exist in three forms, i.e., ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA), all of which were apparently involved in protein synthesis (Eddy 2001).

Based on these observations, Crick formulated the "central dogma" of molecular biology stating that genetic information flows unidirectionally from DNA to RNA to protein, with RNA considered to be a mere intermediary (Crick 1958). However, modifications in this original proposal, which was believed for nearly 40 years, started with the emergence of new knowledge. First was the discovery of reverse transcriptase – which permits the information in RNA to be "back transcribed" into DNA (Baltimore 1970; Temin and Mizutani 1970). The findings fetched a Nobel Prize to Baltimore, Dulbecco, and Temin in 1975. The second groundbreaking discovery was made by Thomas Cech and Sidney Altman who

demonstrated that certain RNAs can themselves function as enzymes, termed as the "ribozymes" that have a fundamental role in mRNA translation, processing of other RNAs such as tRNAs, and splicing of heterogeneous nuclear (hn)RNA (Kruger et al. 1982; Guerrier-Takada et al. 1983; Doudna and Cech 2002). This dual function of information storage and catalysis inspired the "RNA world hypothesis" by Walter Gilbert, which postulates that RNA preceded the DNA as hereditary material and was a key molecule that provided the precursors for the evolution of early life on Earth (Atkins and Gesteland n.d.). Another discovery challenging the role of RNA as a modest messenger was that of "splicing" and subsequently "regulated alternative splicing" that allows generation of multiple proteins from a single gene due to the alternative modes of processing of the same RNA transcript. With this discovery, the concept of "one gene one polypeptide" faded away (Sharp 1994, 2005; Brett et al. 2002).

The next breakthrough was the discovery of small regulatory RNAs (lin-4 and let-7) that regulate the developmental timings of *Caenorhabditis elegans (C. elegans)* (Lee et al. 1993; Reinhart et al. 2000). However, these small RNAs were considered interesting peculiarities until the discovery of "antisense RNA regulation" in plants and "RNA interference" in *C. elegans* (Fire et al. 1998; Waterhouse et al. 1998). RNAi, where a gene is silenced post-transcriptionally by the introduction of a dsRNA, quickly established itself as an important molecular biology tool that not only allows functional screening of genes in experimental settings but also has a tremendous therapeutic potential. The discovery led to an explosion of research on RNA. The coming years saw an unprecedented increase in the number of publications describing the diverse kinds of non-coding RNAs and their fundamental role in regulation of gene expression, especially in animals. These studies caused a major paradigm shift, which is still underway, in our understanding from the concept of primordial RNA world to a contemporary RNA world that includes coding RNAs, ribozymes, and diverse short and long non-coding RNAs that operate at different levels to not only perform housekeeping functions but also regulate the epigenetic pathways that govern development and differentiation in multicellular eukaryotes (Cech 2012). The discovery of CRISPR (Horvath and Barrangou 2010) attests that even the prokaryotic life forms dwell in this modern RNA world, opening up a vast field for exploration.

So, the question that comes to the mind is: What are the unique properties of RNA that makes it such a versatile biopolymer?

1.2 RNA: A Versatile Polymer

RNA is a ubiquitous and biochemically versatile molecule having myriad roles in the cell ranging from translating the genetic information to processing of RNA and regulation of gene expression as well as housekeeping functions (Caprara and Nilsen 2000). The multifunctionality of RNA stems from its unique physiochemical properties (Geisler and Coller 2013).

1. RNA recognizes specific sequences of both DNA and RNA by simple base-pairing interactions as compared to proteins which require much more times of genome sequence space to achieve the same feat (Filipovska and Rackham 2012).
2. Complementary regions in the single-stranded RNA can base pair with each other to form complex three-dimensional structures, which are not only central to its catalytic abilities but also provide intricate recognition surfaces for binding of molecular targets ranging from small molecules/ligands to metal ions to proteins that determine its diverse nongenetic functions (Spirin 2002). Any change in the base composition of the RNA is reflected in the change of shape as well as function, endowing it with enormous flexibility, highly increasing the repertoire of molecular targets that RNA can bind with high affinity and specificity. By virtue of its sequence-specific base pairing and assembly of RNP (ribonucleoprotein) complexes, RNA can recruit generic effector molecules to distinct loci (Geisler and Coller 2013).
3. In addition to the structural dynamicity, RNA exhibits tremendous plasticity in expression as it can be rapidly transcribed and degraded. Moreover, since a regulatory RNA gene does not need to be translated in order to function, it could quickly transition from a state of being transcriptionally inactive to that of fully functional (Geisler and Coller 2013).
4. As RNAs are relatively less conserved, immune to mutations, and evolutionarily malleable, they provide a molecular platform for the rapid evolution of diverse cellular activities (Geisler and Coller 2013).
5. The integration of retroviruses in the genome and the presence of numerous processed pseudogenes suggest the possibility of RNA-mediated changes in the genome to become heritable (Geisler and Coller 2013).

These defining properties of RNAs raise exciting possibilities as to what roles non-coding RNAs could have in the cell in both physiological and pathological states. In this chapter we focus on the biology of the coding mRNAs along with the ever-increasing number of small and long non-coding RNAs, detailing their biosynthetic pathways and their basic physiological functions in the cell.

1.3 The Expanding RNA World

The word "RNome" refers to the sum total of all the RNA species, coding as well as non-coding, that is present in a cell at a particular time point. Recent advances in large-scale sequencing of genomes, depth and quality of RNA sequencing, and whole transcriptome analysis have revealed the complex nature of mammalian transcriptome. Although approximately 90% of the genome is transcribed in eukaryotes, only 2% serves as blueprint for proteins in humans and other mammals. The rest of the transcripts do not code for the proteins and function at the RNA level

(Carninci et al. 2005; ENCODE Project Consortium et al. 2007; Harrow et al. 2012). The non-coding RNAs (ncRNAs), such as tRNAs, rRNAs, small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs), have long been known to be constitutively expressed, performing vital "housekeeping functions" in the gene expression pathways. But the advent of next-generation sequencing projects, such as FANTOM (Functional Annotation of Mammalian cDNA), has revealed the abundance of numerous ncRNAs across human genome, which exceed protein-coding genes, not just in numbers but also in complexity (Carninci et al. 2005). The proof that these non-coding RNAs are functional entities in the cell and not mere a transcriptional noise came from the studies that showed that (a) the percentage of non-coding RNA correlates with the developmental complexity of higher eukaryotes, (b) the conservation of the sequences coding for most of the ncRNAs, and (c) the ubiquitous differential transcription of the vast majority of the genome (Clark et al. 2011; Djebali et al. 2012). As the research uncovering the intricacies of the entire transcriptome of the cell continues, the catalogue of functional RNA species is being enriched at a rapid pace.

1.4 Classification of the Cellular RNome

As already mentioned the whole RNome of a eukaryotic cell can be divided into coding and non-coding RNAs. ncRNAs can be classified into either housekeeping or regulatory ncRNAs. Housekeeping ncRNAs are most often constitutively expressed and include tRNAs, rRNAs, snRNAs, and snoRNAs. Regulatory ncRNAs fall into several classes based on their length, biogenesis, polarity (sense or antisense), and putative functions but share a common functional theme of regulating gene expression (Clark et al. 2013). However, in this discussion, we stick to the basic classification of ncRNAs, i.e., based on their size. Long ncRNAs are typically >200 nt long and function without major prior processing. By contrast, small ncRNAs are processed from longer precursors by endogenous RNAses. microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs) are the most well-studied small ncRNAs that participate in RNAi. We discuss the origin, biogenesis, and functions of each of these categories in the subsequent sections.

1.5 Coding RNAs: The Messenger RNAs (mRNAs)

mRNAs form the blueprint for the proteins synthesis in the cells. In eukaryotes, mRNAs are transcribed from the antisense strand of the DNA by RNA polymerase II. The information on the mRNA is in the form of triplet codons of nucleotides that are decoded by tRNAs and translated into proteins with the help of ribosomal machinery.

1.5.1 Biosynthesis and Processing of mRNAs

The transcription of the coding genes by RNAPII gives rise to heterogeneous nuclear RNAs (hnRNAs) or pre-mRNAs that undergo characteristic modifications and splicing reactions to give rise to mature mRNAs. One of the key modifications is the capping of the 5′-end of the mRNA. The 5′-cap is mostly the 7-methylguanosine (m7G), which is connected to the first nucleotide of the transcript by a unique 5′ to 5′-triphosphate bridge. The cap is synthesized at an early stage of transcription by the action of three enzymes, namely, RNA triphosphatase, guanylyl transferase, and N7-methylase (Ghosh and Lima 2010). Depending upon the type of RNA and the organism, the 5′-cap can be further modified. The simple monomethylated cap is referred to as Cap 0 and is present in lower eukaryotes. Addition of methyl groups at the 2′-OH of first or first and second nucleotides of the transcript yields Cap 1 and Cap 2, respectively. These two are present in higher eukaryotes (Furuichi and Shatkin 2000). The cap participates in several aspects of RNA biology, ranging from protection by 5′-exonucleases, mRNA maturation, and export to the nucleus to initiation of translation (Moore 2005). 5′-cap is able to bind proteins due to its unique chemical structure which bears a positively charged nucleobase and a negatively charged triphosphate chain. The structure confers the m7G with the ability to form stacking interactions with aromatic proteins in the protein binding cavities (Quiocho et al. 2000; Ziemniak et al. 2013).

The 3′-poly (A) tailing of the mRNA transcript is coupled with transcription termination. The RNAPII transcribes past the 3′-end of the mRNA, and the cleavage site is recognized as target for endonucleolytic cleavage and polyadenylation. The cleavage site in most of the pre-mRNAs is flanked on the either side by two *cis*-acting signals. The sequence AAUAAA, also known as poly (A) sequence, is located ~11–30 base pairs upstream from the cleavage site. The second signal is the presence of downstream U-rich or GU-rich region. Both these sequences are recognized by CPSF (cleavage polyadenylation and specificity factor) and Cst F (cleavage stimulation factors), which then cleave the RNA transcript. Next, the enzyme poly (A) polymerase adds poly (A) tail of ~200 residues in a non-template-dependent manner. Similar to 5′-cap, the poly (A) tail is important for protection form 3′ → 5′ exonucleases, transport of mature mRNA to the nucleus, and initiation of translation (Krebs et al. 2014).

Most of the genes in humans are interrupted genes, harboring a number of introns. The removal of these introns is of fundamental importance to gene expression. The splicing of pre-mRNA occurs co-transcriptionally and is catalyzed by "spliceosome," which is a dynamic multi-subunit protein complex consisting of snRNAs (U1, U2, U4, U5, and U6) and a number of associated proteins (Wahl et al. 2009). The spliceosome orchestrates two transesterification reactions necessary for removal of introns and joining of adjacent exons. The spliceosome operates by step-by-step assembly, initiated by the recognition of 5′-splice site (5′-SS) by complementary base pairing with U1 snRNA. The branch site sequence (BSS) then binds to U2 snRNP to form a pre-splicing complex known as complex A. The U4/U6.U5 tri-snRNP, in which U4 and U6 snRNAs are extensively base

paired, binds next to form complex B1. A series of rearrangements follow that result in destabilization and release of U1 and U4 snRNPs. This results in the formation of complex B2, with the simultaneous activation of spliceosome, leading to the first nucleophilic attack of splicing step, in which the 2′-OH group of the bulged-out branch point adenosine attacks the 5′ SS. After the first step of splicing, complex B2 is converted into complex C, and the second step of splicing follows, resulting in the production of mature mRNA and the release of the excised intron lariat and the U2, U5, and U6 snRNPs, which are recycled for further rounds of pre-mRNA splicing (Wahl et al. 2009; Karijolich and Yu 2010).

The hnRNA may also undergo "alternative splicing," i.e., differential removal of introns from the pre-mRNA by combinatorial use of different splice sites. Alternative splicing is significant as it leads to the formation of multiple proteins from a single gene (Salton and Misteli 2016). Further, alternative splicing may also reduce translation of mRNAs by inclusion of an exon that harbors a premature stop codon, leading to sequestration and degradation of transcripts. Such exons are known as "poison exons" and are important in gene regulation (Ge and Porse 2014). The importance of alternative splicing is demonstrated by RNAseq-mediated deep mining of cellular transcriptome studies, which show that >90% the human pre-mRNAs are alternatively spliced. The phenomenon of alternative splicing is governed by many factors such as RNA secondary structure, strength of 5′- and 3′-splice sites, splicing enhancer and silencer sequence elements, exon/intron architecture, and transcription by RNA polymerase II. However, their relative contribution in determination of final splicing pattern has not yet been fully elucidated, and unraveling of a "splicing code" is a thrust area of current research (Lee and Rio 2015).

Finally, there are two pathways for mRNA turnover: 5′ → 3′ decay and 3′ → 5′ decay. The 5′ → 3′ pathways are initiated by the Dcp1/Dcp2 decapping complex, in which Dcp2 is a Nudix pyrophosphatase and Dcp1 is a regulatory subunit. It removes the cap from the mRNA by cleaving between α and β phosphates to produce m7GDp and 5′-monophosphate mRNA, which is subsequently degraded by exonucleases such as Xrn1. In the 3′→5′ pathway, decapping is preceded by the degradation of mRNA from the 3′-end by the exosome (Houseley and Tollervey 2009; Balagopal et al. 2012).

1.6 miRNAs

miRNAs are the best-known class of non-coding RNAs. Since their discovery in *C. elegans* over a decade ago, miRNAs have emerged as one of the most abundant and important classes of small non-coding RNAs, and their misregulation has been shown to be causally linked to various diseases in humans (Sayed and Abdellatif 2011). The growing number of miRNAs is curated at the miRBase database (http://www.mirbase.org/) (Kozomara and Griffiths-Jones 2014). The latest release (Release 22, March 2018) of the database contains 38,589 entries representing hairpin precursor miRNAs, expressing 48,885 mature miRNA products, in 271 species

including animals, plants, unicellular algae, and even viruses. Computational studies estimate that more than 60% of total genes can be regulated by miRNAs, despite the fact that miRNAs constitute only 1–3% of the human genome (Friedman et al. 2009; Ghildiyal and Zamore 2009). miRNAs are cell-endogenous, 20–24 nucleotide long, non-coding RNAs that are expressed in cell- and development stage-specific manner and influence diverse biological functions through RNA interference-mediated post-transcriptional repression of target genes (Davis and Hata 2009). miRNAs have been shown to regulate almost all the cellular processes investigated so far, in physiological as well as pathological conditions. While some miRNAs, such as let-7, are perfectly conserved across diverse species, others share identical 5′-end sequences (typically nucleotides 2–8 called the "seed region") comprising families of miRNAs, which may have redundant roles or regulate genes in a common pathway (Aalto and Pasquinelli 2012). miRNAs down-regulate the gene expression by binding completely or partially complementary regions in the 3′-UTRs of their target genes. Moreover, since a particular miRNA can target multiple genes (owing to their small size and imperfect base pairing), they provide a mechanism for simultaneous regulation of several genes involved in a particular physiological pathway. In addition to their endogenous physiological role as regulators of gene expression, cells can also passively and/or actively release miRNAs and thus function as paracrine molecules that regulate gene expression in other cells.

The nomenclature of miRNAs is varied. The miRNAs that were identified during early genetic studies were named after their phenotypes, e.g., *lin-4* and *lsy-6*. Interestingly lin-4 was identified even before the term microRNA was coined. Since 2002, the miRBase (http://www.mirbase.org/) maintains the record of all annotated miRNAs. Thereafter the miRNAs that were discovered were given numerical names that are simply sequential, say, for example, miR-125. The predicted stem-loop portion of the primary transcript is designated as mi**r**-125 and mature miRNAs being referred to as mi**R**-125. miRNAs sisters that are encoded from the same gene are indicated with lettered suffixes (e.g., *mir-125a* and *mir-125b*). If the same mature miRNA is generated from distinct genomic loci and precursor sequences, numeric suffixes are added at the end of mature miRNA name, for example, *mir-125b-1* and *mir-125b-2*. If two mature miRNAs are produced form each locus, one from the 5′-strand and one from the 3′-strand of the precursor, they are named as miR-125a-5p and miR-125a-3p. However, usually one arm is predominantly expressed and is more biologically active than the other arm. The former is known as "guide strand," and the latter is termed "passenger strand" and is designated as miR-125* (Wright and Bruford 2011; Ha and Kim 2014).

1.6.1 Genomic Organization of miRNAs

miRNA genes have been mapped on all chromosomes in humans except for the Y-chromosome (Ul Hussain 2012). Approximately 50% of known miRNAs are found in clusters that are transcribed as polycistronic primary transcripts. Based on their genomic location, miRNA genes can be categorized as intergenic, intronic, and exonic (Rodriguez et al. 2004).

(A) *Intergenic miRNAs*

Majority of the mammalian miRNA genes (70%) are located in defined transcriptional units having their own promoters and other regulatory elements. Intergenic miRNAs can be monocistronic or polycistronic, are mostly transcribed by RNA polymerase II (RNAP II), and bear a 7-methyl guanylate cap at the 5′-end and poly (A) tail at the 3′-end, similar to mRNAs. A subset of miRNAs have also been demonstrated to be transcribed by RNAP III (Rodriguez et al. 2004; Lee et al. 2004; Cai et al. 2004).

(B) *Intronic miRNAs*

Intronic miRNAs (single or in clusters) are present in the intronic regions of annotated genes, both protein coding and non-coding. Intronic miRNAs can be transcribed from the same promoter as their host genes and processed from the introns of the host gene transcripts (Ul Hussain 2012). Since these miRNAs genes share the same promoters as their host gene, they usually have similar expression profiles as those of the host gene (Baskerville and Bartel 2005). However, not all intronic miRNAs are co-expressed with their host genes. Nucleosome mapping studies and RNAP II chromatin immunoprecipitation studies indicate that 25–33% of intronic miRNAs are transcribed from independent promoters, though they reside within the introns (Corcoran et al. 2009). Furthermore, the whole intron of a protein-coding gene may act as the exact sequence of the pre-miRNA with splice sites on either side. Such introns are termed mirtrons, and, since these are processed by the splicing machinery, the first step involved in their biogenesis is not required for their maturation. Certain variant forms of mirtrons have also been identified such as 5′-tailed mirtrons (mirtrons having a sequence extension at the 5′-end) and 3′-tailed mirtrons (mirtrons having a sequence extension at the 3′-end) (Ruby et al. 2007).

(C) *Exonic miRNAs*

Exonic miRNAs are rare and often overlap an exon and an intron of a non-coding gene. These miRNAs are also proposed to be transcribed by their host gene promoter, and their maturation may exclude host gene function (Rodriguez et al. 2004). Sometimes miRNAs may also be embedded in other long non-coding RNAs, resulting in dual functionality of the primary transcript (Xie and Steitz 2014).

1.6.2 Biogenesis of miRNAs

The biogenesis of the miRNAs is the most well-characterized among all the classes of non-coding RNAs. Recently, many studies have also uncovered certain alternative mechanisms of miRNA biogenesis as well as pathways of recycling of miRNAs (García-López et al. 2013a). In this section, we start with describing the canonical pathway of miRNA biogenesis and discuss the other two pathways subsequently.

1.6.2.1 Canonical miRNA Biogenesis Pathway

Two simple observations led to the formulation of current model of miRNA maturation. Firstly, miRNAs are transcribed as long primary transcripts (pri-miRNAs) that are first trimmed into the hairpin intermediates (pre-miRNAs) and subsequently cleaved into mature miRNAs. And secondly, the catalytic activities for the first and the second processing are compartmentalized into the nucleus and the cytoplasm, respectively, thus necessitating the nuclear export of pre-miRNA into the cytoplasm (Fig. 1.1a).

As stated above, miRNA genes are majorly transcribed by RNAP II (with a few exceptions that are transcribed by RNA pol III), yielding primary transcripts denoted as primary miRNAs (pri-miRNAs) that are capped, polyadenylated, and usually several kilobases long containing one or more long hairpin structures in which mature miRNA sequences are embedded (Cai et al. 2004). The miRNAs that are located closely in the genome are transcribed together as cluster of miRNAs and then subsequently processed (García-López et al. 2013a). Since the structural features of these hairpins are unique to pri-miRNAs, they distinguish them from the various RNA stem-loop-like structures present in the nucleus. The pri-miRNA hairpin contains a long hairpin stem of ~30 bps with flanking 5′- and 3′-single-stranded ends (Han et al. 2006; Zeng and Cullen 2004) allowing it to be recognized and processed by a microprocessor complex consisting of a ribonuclease Drosha (RNase III enzyme), the RNA-binding protein DGCR8 (DiGeorge syndrome critical region gene 8 (also known as Pasha, partner of Drosha)) in invertebrates (Lee et al. 2003; Han et al. 2004), and other accessory factors. DGCR8 recognizes single-stranded to double-stranded RNA junctions at the 5′- and 3′-ends of pri-miRNA, whereas Drosha cleaves the pri-miRNA hairpin at a distance of approximately 11 bp from the site recognized by DGCR8, to release the 55–70 nt pre-miRNA hairpin (Xie and Steitz 2014). In case of intronic miRNAs, Drosha cleavage occurs co-transcriptionally before the splicing of the host RNA (Kim and Kim 2007; Morlando et al. 2008). The resulting RNA hairpin is called precursor miRNA or pre-miRNA and usually possesses a 5′ monophosphate and a 2-nt overhang at the 3′-end which is the hallmark of RNase III products.

Since further processing of pre-miRNA occurs in the cytoplasm, it has to be exported from the nucleus to the cytoplasm. The nuclear export machinery comprises of exportin-5 and Ran-GTP, a nuclear GTP-binding protein, and recognizes structural features of pre-miRNAs such as the 3′-overhang and the double-stranded stem (with a minimal length of 16 bp) (Yi et al. 2003; Zeng and Cullen 2004; Lund et al. 2004). After translocation to the cytoplasm, Ran-GTP is hydrolyzed to

Fig. 1.1 (continued) by exonucleases to give rise to pre-miRNAs. The pre-miRNAs are exported from the nucleus by XPO5 and merge with the canonical biogenesis pathway. (**c**) moRNAs are produced during canonical miRNA biogenesis of adjacent miRNAs by the action of the microprocessor complex. (**d**) Simtrons are a type of mirtrons that neither require spliceosome nor Dicer for their biogenesis. They are processed from the introns by Drosha and possibly an unknown binding partner (but do not require DGCR8). They are further processed by unidentified factors and enter the RISC complex with any of the four human Argonaute proteins

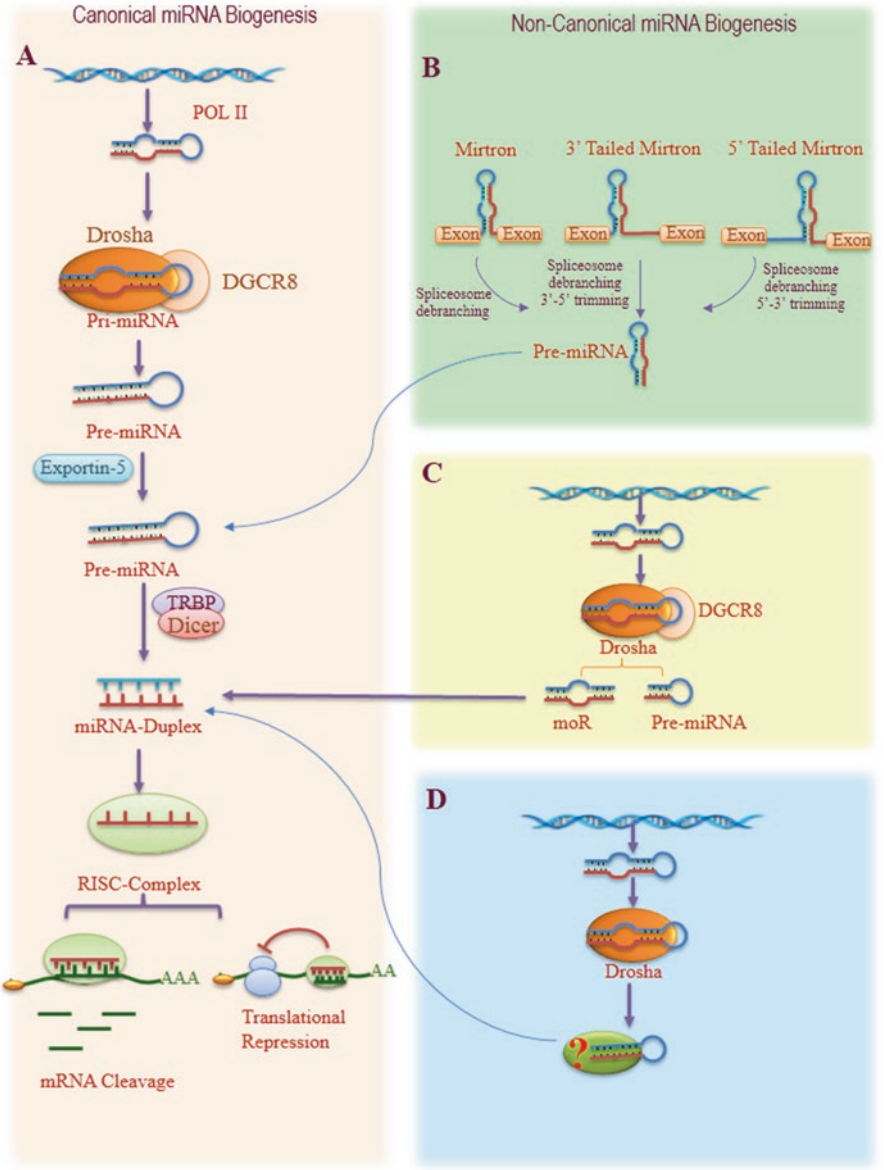

Fig. 1.1 Schematic overview of canonical and non-canonical pathways of miRNA biogenesis. (**a**) Canonical miRNAs are processed by microprocessor complex (Drosha and DGCR8) to form pre-miRNAs, which are exported from the nucleus by XPO5, and cleaved by Dicer to give rise to miRNA duplex. The miRNA/miRNA* duplexes are loaded into RISC complex, out of which guide strand is retained while passenger strand is degraded. MiRNA-RISC binds to complementary regions in the 3′-UTRs of target mRNAs and guides target mRNA degradation or translational repression depending on the sequence complementarity. (**b**) Mirtrons are excised from the host pre-mRNA by the spliceosome and linearized by debranching. Tailed mirtrons are further processed

Ran-GDP (by GTPases present in the cytoplasm), resulting in the disassembly of the complex and release of pre-miRNA. Depletion of exportin-5 from the cell results in reduced levels of mature miRNA without a nuclear accumulation of pre-miRNA, suggesting that exportin-5 is not only involved in the export of pre-miRNA but may also protect it from nucleolytic attack in the nucleus (Yi et al. 2003).

Following their export from the nucleus, pre-miRNAs are processed in the cytoplasm into mature miRNA duplexes by the cytoplasmic RNase III known as Dicer, through interaction with the protein TRBP, which recognizes the 3′-ends generated by Drosha and cleaves the pre-miRNA at approximately two helical turns (approximately 22 nt) away to produce a miRNA–miRNA duplex with 2-nt 3′-overhangs at both ends (Ul Hussain 2012; García-López et al. 2013a). Mirtrons, which are spliced out of the mRNA transcript, are known to bypass the Drosha processing, but their nuclear export and cleavage by Dicer follow the canonical pathway (Ruby et al. 2007) (Fig. 1.1a).

Dicer cleavage of pre-miRNA results in the formation of mature double-stranded miRNA whose one strand is called the guide strand, while the complementary strand is called the passenger strand. The guide strand is incorporated into an effector complex known as miRISC (miRNA-containing RNA-induced silencing complex), whereas the passenger strand is released and degraded (Kawamata and Tomari 2010). The relative thermodynamic stability of the two ends of the duplex determines which strand is to be selected, with the strand having relatively unstable base pairs at the 5′-end typically remaining in the effector complex (e.g., G:U pair versus G:C pair) (Khvorova et al. 2003; Schwarz et al. 2003). The miRISC is a ribonucleoprotein complex which consists of a miRNA strand, the Argonaute (AGO) protein family, the GW182 protein family (glycine-tryptophan [GW] repeat-containing protein of 182 kDa) and some other accessory proteins. Out of these, the proteins of the Argonaute family play a vital role in downstream functions of this silencing complex. Mammals have seven members in the Ago protein family that are categorized into two subfamilies: the AGO subfamily and the PIWI subfamily. Out of these, the AGO subfamily (AGO1-4) is involved in the miRNA and endogenous small interfering RNAs (endo-siRNAs) pathways, whereas PIWI proteins are associated with piRNAs (García-López et al. 2013a). The AGO proteins contain three evolutionarily conserved domains, namely, PAZ, MID, and PIWI, which interact with the 3′- and 5′-ends of the miRNA, respectively (Ender and Meister 2010). AGO2 has an enzymatically competent RNaseH-like PIWI domain, which endonucleolytically cleaves the perfectly complementary mRNA targets at the center of the miRNA–mRNA duplex. AGO proteins bridge the miRNA to the silencing effectors, whereas the GW182 proteins act downstream of the AGO proteins to effect miRNA-mediated repression (Fabian et al. 2010).

Mechanism of Action and Functions of miRNAs

The miRNAs carry out post-transcriptional gene silencing, where the miRNA functions as a guide in recognizing the target mRNA by base pairing, while the downstream effector functions are performed by AGO proteins. The complementarity

between the miRNA (assembled in the miRISC) and target mRNA determines whether the target mRNAs would be destroyed or their translation would be repressed (Fabian et al. 2010). Out of the 20–22 nt, only 2–8 nts from the 5′-end of the miRNA, known as the "seed" region, form a perfect match with the 3′-UTRs of the target mRNA (Ameres et al. 2007). The binding site of miRNAs is not restricted to 3′-UTR, with several reports validating the targeting of 5′-UTRs, promoters, and open reading frames of the genes (Lee 2013). The choice between translational repression and destruction is governed by the degree of mismatch between the miRNA and its target mRNA, with degradation being the outcome for perfectly complementary targets. miRNAs have been reported to repress the translation of target mRNA by blocking translation initiation and/or elongation or co-translational protein degradation. Alternatively, targets of miRNA may be sequestered away from the translation machinery and transferred to processing-bodies (p-bodies), where mRNAs are known to be degraded as a consequence of deadenylation by the deadenylase complex (CCR4-CAF1-NOT1) followed by decapping by the DCP1/DCP2 decapping enzyme complex and finally degradation by the major cytoplasmic 5′-to-3′ exonuclease XRN1 (Fabian et al. 2010).

Since miRNAs can inhibit the translation of imperfectly matched targets, each miRNA might target multiple genes, and each gene in turn might be regulated by several miRNAs (Meltzer 2005), and hence miRNA-mediated regulation has been implicated in almost all the biological processes studies so far (Filipowicz et al. 2008). Besides, miRNAs can also target other non-coding RNAs such as long non-coding RNAs, further complicating the transcriptomic networks (Zhou et al. 2010). miRNAs can also affect transcriptional gene silencing in the nucleus. In addition, miRNAs can be released in the exosomes and affect the gene expression of target mRNAs in the neighboring cells. Interestingly, certain miRNAs have also been found to increase rather than inhibit the expression of certain human genes. Such a phenomenon is known as RNA activation (RNAa) and is brought about by binding of miRNAs to the promoter regions and promoting activating chromatin modifications or by competing with the promoter-binding repressor (Li et al. 2006; Huang et al. 2010). It is important to note here that the hexanucleotide element at the 3′-end of miRNA can direct the translocation of the mature miRNAs to the nucleus in order to target the genomic DNA (Lee 2013).

Regulation of miRNA Expression

The canonical pathway of miRNA biogenesis can be regulated at three levels: (a) at the level of transcription of pri-miRNAs, (b) the editing mechanisms that disrupt the processing of miRNA precursors, and (c) through the regulation of miRNA processing enzymes such as DROSHA and DICER (García-López et al. 2013a).

RNAP II-mediated transcription is one of the major regulatory steps involved in the biosynthesis of miRNAs. A large-scale mapping of 175 human miRNA promoters through nucleosome positioning and chromatin immunoprecipitation-on-genomic DNA microarray chip (or ChIP-on-chip) analysis suggested that the promoter structure of miRNA genes, including the relative frequencies of CpG islands, TATA box, TFIIB recognition, initiator elements, and histone

modifications, is indistinguishable between the promoters of miRNA and mRNA (Ozsolak et al. 2008; Corcoran et al. 2009). Furthermore, DNA-binding factors that regulate miRNA transcription largely overlap with those that control protein-coding genes. For instance, c-Myc, a transcription factor, is known to activate the transcription of the miR-17-92 cluster (Tagawa et al. 2007). Autoregulation of miRNA expression is found when the transcription factor which regulates miRNA expression is targeted by the miRNA itself, allowing tight control of miRNA and transcription factor levels. An example of this type of regulation is the c-Myb-miR-15a loop in hematopoiesis (Zhao et al. 2009). In addition to the involvement of various canonical transcription factors, other epigenetic mechanisms, such as DNA methylation and histone modification, are also known to regulate the expression of miRNA genes. Besides the transient regulation mediated by acetylation and methylation, miRNA are also subjected to stable epigenetic control through genomic imprinting. Two well-characterized imprinted regions, H19 and the Dlk1-Gtl2 domains, have been found to contain miRNA clusters (Seitz et al. 2004; Cai and Cullen 2007).

The pre-, pri-, and even mature miRNAs can also be edited by a family of proteins known as adenosine deaminases acting on RNA (ADARs) that deaminate adenosines to inosines. Inosine is recognized as guanosine by the translational machinery as well as in direct base pairing. These modifications not only disrupt the recognition and processing of miRNA precursors by the processing machineries (Drosha and Dicer) but can also change the targets of mature miRNAs (upon editing of mature miRNAs), a phenomenon known as retargeting (Kawahara et al. 2007; Heale et al. 2009; García-López et al. 2013b).

The activity of DROSHA/DGCR8 and DICER and the transport of the miRNA complex from the nucleus to the cytosol are also targets of regulation (Suzuki et al. 2009; Davis-Dusenbery and Hata 2010). For example, the binding of ribonucleoprotein hnRNPA1 to the loop region of *pri-microRNA-18a* facilitates its processing by DROSHA/DGCR8 (Michlewski et al. 2010), while the protein lin-28 represses the DICER-mediated processing of let-7 miRNA (Chang et al. 2009).

1.6.2.2 Non-canonical Biogenesis of miRNAs

The increasing appreciation of the alternative mechanisms of biogenesis of miRNAs, in the past few years, has instigated a paradigm shift in our understanding of the ever-increasing complexity of RNA interference in the cell. Such non-canonical pathways of miRNA biogenesis often bypass the processing by Drosha/DGCR8 complex and utilize the molecular machineries participating in the biogenesis and processing of other classes of cellular RNAs (Xie and Steitz 2014). Notably, the functioning of Dicer is nearly indispensable for the production of both canonical and non-canonical miRNAs. In the absence of Dicer, a loss of almost all the functional miRNAs is observed (Abdelfattah et al. 2014). However, as with all the biological phenomena, there are certain exceptions where the intermediate precursor is cleaved by Ago2 instead of Dicer (Lee 2013). The presence of such multiple pathways of miRNA biogenesis opens a plethora of regulatory options for the differential expression of individual miRNAs (Winter et al. 2009).

(a) Mirtrons: miRNA biogenesis via splicing

The mirtron pathway was the first alternative miRNA biogenesis mechanism to be described from the studies carried out in *Drosophila melanogaster* and *C. elegans* (Ruby et al. 2007; Okamura et al. 2007). Since then they have also been found in the short introns of vertebrates, mammals, and even in rice genomes (Berezikov et al. 2007; Glazov et al. 2008; Zhu et al. 2008; Sibley et al. 2012). As mentioned earlier, mirtrons originate from the intronic regions of mRNAs and are processed into pre-miRNA-like hairpins by the spliceosome machinery, bypassing the Drosha/DGCR8 processing (Fig. 1.1b). Their independence of the Drosha/DGCR8 was demonstrated by their increased abundance in *Drosha* mutants (Martin et al. 2009) and *DGCR8* knockout cells (Babiarz et al. 2008). After the excision of intron in the form of a lariat, it is resolved by the action of debranching enzyme (Ldbr) in the nucleus to give rise to a pri-miRNA hairpin (suitable for Dicer cleavage) that is exported to the cytoplasm in exportin-5-dependent manner and merges with the canonical pathway. It is important to note that mirtron hairpins formed after splicing are ~10 bp shorter than canonical pri-miRNA hairpins and thus are able to bypass the Drosha cleavage step (Okamura et al. 2007). Although a typical mirtron has well-defined 5′-and 3′-ends that are generated by the excision within the splice donor and acceptor sites of the mRNA (and the mutation of these sites abolishes their production), there are certain exceptions where the RNA-generating hairpin resides toward one end of the intron (Ruby et al. 2007). Such mirtron has a single-stranded tail at the 3′-end (3′-tailed mirtrons) or 5′-end (5′-tailed mirtrons) that has to be processed by RNA exosome (which usually stops its action at the stem-loop secondary structure of pre-mirtron) before export into the cytoplasm and cleavage by Dicer (García-López et al. 2013a) (Fig. 1.1b).

(b) Simtrons

Simtrons were discovered by M.A. Havens et al. while studying the origin of predicted mirtrons (Havens et al. 2012). To their surprise, a subset of mirtrons was not processed by spliceosomes (i.e., mirtron processing pathway) or canonical miRNA processing pathway. Rather their biogenesis occurred by a novel pathway that involved Drosha, but did not require DGCR8 (Fig. 1.1d). They termed these mirtron variants as "simtrons," for splicing-independent mirtrons. In addition, not only were simtrons found to be transported to the cytoplasm in an exportin-5-independent manner; their processing was also independent of Dicer or Ago2. However, they were able to bind all four classes of Ago proteins and were capable of gene silencing (Havens et al. 2012).

(c) Other pathways of miRNA generation

miR-451 can be processed by an alternative pathway in which it is processed by Ago-2 itself, bypassing the Dicer cleavage (Yang et al. 2010). miRNAs can also be processed from snoRNAs in a Drosha-independent and Dicer-dependent pathway

similar to mirtrons. Further, tRNAs can also be sliced by dicer or RNAse Z to give rise to short tRNA-derived fragments which associate with argonaute proteins and carry out miRNA-like functions (Cole et al. 2009; Brameier et al. 2011). The biogenesis of both these types of miRNAs is discussed later in the chapter.

1.6.3 Bioavailability and miRNA Recycling

The bioavailability of the miRNAs depends on many factors such as expression of the miRNA biogenesis pathway factors, presence of miRNA reservoirs, as well as competing endogenous RNAs such as circular RNAs, lncRNAs, and pseudogenes. The global down-regulation of expression of genes coding for miRNA biogenesis factors such as DICER or DGCR8 stops the production of new mature miRNAs and is important in processes such as embryonic development or cell type differentiation (González-González et al. 2008). For example, during the early stages of embryo development after fertilization, the maternal program is suppressed to initiate the zygotic activation program. Studies in Dgcr8 -/- mutant mice suggest that miRNA activity is suppressed during these stages in mammals (González-González et al. 2008; García-López et al. 2013a). However, this ubiquitous down-regulation of miRNA biogenesis may not imply absence of specific miRNAs. In the early stages of embryo development (when the miRNA biogenesis was dramatically down-regulated), specific miRNAs such as *mmu-miR-292-3p* and *mmu-miR-292-5p* could be preserved as double-stranded molecules through the "protection" from the binding to mRNA targets, pseudogenes, duplex passenger strands, or other types of RNA-specific reservoirs (García-López and del Mazo 2012; García-López et al. 2013a). Mature mRNAs, lncRNAs, and circular RNAs, all have been shown to act as "miRNA reservoirs" and participate in miRNA recycling. In response to demand, these molecular reservoirs can directly provide many functional miRNAs (Gu et al. 2007; Memczak et al. 2013; García-López et al. 2013a).

Besides acting as molecular reservoirs, certain RNA species may act as natural decoys that compete for the common pool of miRNAs. Such RNA species are known as "competing endogenous RNAs (ceRNAs)." The ceRNA hypothesis states that any RNA transcript that harbors microRNA response elements (MREs) can sequester corresponding miRNAs and thereby regulate the expression of their target genes (Karreth and Pandolfi 2013). Many long non-coding RNAs and pseudogenes act as endogenous ceRNAs for different miRNAs. Phosphatase and tensin homolog pseudogene (PTENP1) was the first RNA for whom ceRNA activity was demonstrated in mammals. The overexpression of PTENP1 causes an increase in the PTEN expression resulting in suppression of cell growth (Poliseno et al. 2010).

1.6.4 microRNA-Offset RNAs

Recently, comprehensive small RNA sequencing led to the discovery of a novel class of small RNAs related to miRNA termed as miRNA-offset RNA (moRNAs;

moRs; MORs). They were first reported in the simple chordate ascidian *Ciona intestinalis* in 2009, followed by their identification in several viruses, mouse and human cells, including hESCs (Shi et al. 2009; Langenberger et al. 2009; Jurak et al. 2010; Umbach et al. 2010; Zhou et al. 2012; Asikainen et al. 2015; Zhao et al. 2016). moRNAs are ~20 nt long, usually included in miRNA hairpin precursor. They are present adjacent to or sometimes even overlapping the miRNA sequences, both 5p and 3p, thus suggesting that they may be arising due to Drosha/DGCR8-mediated cleavage of the pre-miRNA (Bortoluzzi et al. 2011; Asikainen et al. 2015). moRNAs can be derived from either arm of pre-miRNAs (independent of the arm from which mature miRNAs are derived), although those derived from the 5′-arm are more prevalent (Bortoluzzi et al. 2011) (Fig. 1.1c). moRNAs seem to be conserved across the species (Langenberger et al. 2009), and their expression is differentially regulated in different developmental stages, at least in *C. intestinalis* (Shi et al. 2009). However, their expression does not always correlate with that of the corresponding mature miRNA, sometimes being even greater than that of the mature miRNAs (Umbach et al. 2010).

While the microRNAs are well-established regulators of gene expression, the biological activity of moRNAs, which can be considered as coproducts of miRNAs, is less explored. Umbach and colleagues were the first to suggest that moRs could regulate endogenous target mRNAs. Using a luciferase-based assay, they demonstrated that moR-rR1-3-5p, a viral moRNA, has moderate inhibitory effect on the expression of an artificial mRNA (Umbach et al. 2010). An experimental evidence of the biological activity of moRNAs was recently provided by Jin Zhao et al. (Zhao et al. 2016). They demonstrated that moR-21 regulates the post-transcriptional expression of genes in an Argonaute-2-dependent manner and inhibits the proliferation of vascular smooth muscle cells (VSMCs). Further, miR-21 and moR-21 were not only found to regulate different genes in the same pathway but also antagonize each other in the regulation of certain genes (Zhao et al. 2016). Since a number of moRNAs were found to be located in the nucleus, it seems that moRNAs may be involved in regulation of transcription or splicing, similar to certain nuclear localized miRNAs (Taft et al. 2010). Such moRNAs may either be re-imported into the nucleus after their processing in the cytoplasm, or they may be processed in the nucleus itself by nuclear small RNA synthesis enzymes (Asikainen et al. 2015).

1.7 Short Interfering RNAs (siRNAs)

siRNAs can be both endogenous and exogenous in origin. Endogenous siRNAs or endo-siRNAs or esiRNAs are 20–23 (~21) nt long non-coding RNAs that were first discovered in worms. Since the biosynthesis of endo-siRNAs involves generation of a dsRNA precursor, endo-siRNAs were observed only in the organisms that possessed the RNA-dependent RNA polymerase (RdRP) activity such as plants, *C. elegans*, and fission yeast. In C. *elegans*, the dicer cleavage of long dsRNAs generates primary siRNAs which then associate with argonaute family of proteins, and

their interaction with target mRNA recruits an RdRP that produces secondary siRNAs, using target mRNA as a template (Ketting et al. 2001; Sijen et al. 2007).

The endo-siRNAs were previously thought to be unlikely to be present in the mammalian cells because the absence of an RdRP-dependent amplifying mechanism would prevent accumulation of biologically significant levels of endo-siRNAs. Secondly, the presence of dsRNA precursors would have signaled a viral infection and triggered an antiviral immune response within the cell. Third, dicer was thought to be a cytoplasmic enzyme, while the export of endo-siRNA precursors to the cytoplasm seemed unlikely (Piatek and Werner 2014).

However, elegant independent studies collectively reported the occurrence of endo-siRNA-mediated silencing pathway responsible for silencing of transposable elements and certain mRNAs, both in the gonadal and somatic cells of *Drosophila* and mouse oocytes. Interestingly, these cells do not mount an interferon response. These siRNA were seen to be originating from endogenous dsRNAs. We now know that endo-siRNAs can be derived from long dsRNA transcripts arising from repetitive elements, transposons, heterochromatic regions, or intergenic elements of the genome (Tam et al. 2008; Watanabe et al. 2008; Ghildiyal et al. 2008; Saito and Siomi 2010). The double-stranded (ds) intermediates can be generated in multiple ways. Intramolecular hairpins arise from "natural" sources such as inverted repeats, bidirectional transcription, or pseudogenes hybridized to mRNA. While the transcriptional read-through of the inverted repeats gives rise to hairpin RNAs upon transcription, the other two kinds of locus generate sense and antisense pairs that arise in *cis* (intermolecular) and *trans* (interchromosomal), respectively (Röther and Meister 2011; Aalto and Pasquinelli 2012) (Fig. 1.2).

The siRNA processing from these dsRNA intermediates is dependent on Dicer but is independent of Drosha. Since long dsRNAs in the cytoplasm are also the substrates for other enzymes such as cytoplasmic protein kinase R and ADAR (adenosine deaminase acting on RNA), dicing is likely to occur in the nucleus, followed by export of the diced RNA fragment to the cytoplasm (Wang and Carmichael 2004; Ender and Meister 2010). Just like the processing of miRNA precursors, cleavage of ds siRNA precursors by dicer results in the formation of small dsRNA (with a 3′ 2 nucleotide overhang) in which one strand is the guide strand and the other is passenger strand. The guide strand is then incorporated into argonaute protein containing RISC, with the siRNAs associating mostly with the Ago2 to direct cleavage of the target mRNAs that have regions complementary to the siRNA. Thus, miRNAs and endo-siRNAs are both similar in the context that they originate from dsRNA precursors that are processed by Dicer and associate with RISC to exert their regulatory functions, both at transcriptional and post-transcriptional level (Piatek and Werner 2014) (Fig. 1.2). However, both the species of small RNAs are distinct in the origin of their dsRNA precursor as well as requirements for mRNA target recognition. The endo-siRNAs exhibit high fidelity base pairing with their target mRNAs, not allowing more than three mismatches. As a consequence, endo-siRNAs elicit direct cleavage of target mRNAs as opposed to miRNAs that mostly repress translation. Both flies and mammals utilize this pathway for biogenesis, but an important difference is that while flies have a distinct

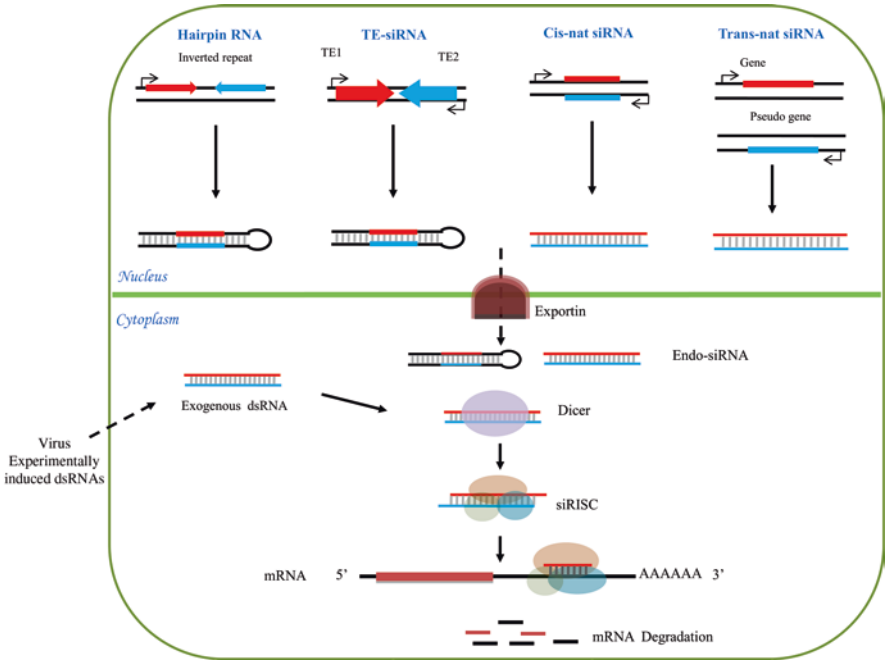

Fig. 1.2 Endo- and exo-siRNA biogenesis pathway. Endo-siRNA precursors are derived from transcriptional read-through of inverted repeat sequences, transposable elements, or sense–antisense pairs (both *cis* and *trans*). Dicer processes the precursor into ~21-nt endo-siRNAs, which are bound to AGO2 to form si-RISCs. Exo-siRNAs are derived from experimentally introduced dsRNAs or viral RNAs and are processed by Dicer to form si-RISCs. si-RISCs bind with perfect complementarity to the target mRNA and leading to its degradation

Ago and Dicer which are involved in the endo-siRNA biogenesis, mammals have only one form of Dicer, indicating that the miRNA and endo-siRNA branches of RNAi are potentially intertwined (Ender and Meister 2010).

Endo-siRNAs are a part of an important genome surveillance mechanism that keeps a check on mobile genetic elements. They also regulate the expression of specific protein coding mRNAs, at both transcriptional and post-transcriptional level. Moreover, endo-siRNAs may also play a "passive" role in mammals where degradation of long cellular dsRNAs would prevent the inappropriate activation of cellular defense mechanisms such as the dsRNA-activated protein kinase (PKR) (Nilsen 2008).

Exogenous siRNAs (exo-siRNAs), on the other hand, are the dicer cleavage products formed from the RNA intermediates that are introduced exogenously into the cytoplasm of the cell, e.g., viral RNA and transgenes, and merge with the canonical siRNA pathway (Fig. 1.2). Exo-siRNAs are also synthesized artificially and are more renowned for their use in experimental systems for knockdown of specific genes (Röther and Meister 2011).