

Peculiarities of non coding micrnas organization in *H. Sapiens* and some other species

Abstract

It is universally accepted today that micro (mi) non coding (nc) RNAs are endogenous ~22 nt RNAs that can play important regulatory roles in a vast number of cellular processes including cell growth, disease, embryogenesis, gene regulation, signal transduction, receptor activation, etc. by targeting mRNAs for cleavage or translational repression. Mi RNAs comprise one of the more abundant classes of gene regulatory molecules in multi cellular organisms and likely influence the output of many protein-coding genes. The shorter *Lin-4 RNA* is now recognized as the founding member of an abundant class of tiny regulatory RNAs called micro RNAs or mi RNAs. Now it has become clear that large number of regulatory small (rs) RNAs are compromised due to traditional approaches to identify mi RNAs mainly to hair-pin loop bred typical signs. The rsRNA-target interactions have been studied with the use of mi RNA recognition elements binding with Ago proteins (AGO-cross linking), microprocessor complex subunit DGCR8 involved in the initial step of mi RNA biogenesis cross-linking, ligation, and sequencing of hybrids (CLASH), proteome and expression data. The advent of next generation sequencing (NGS) technologies, powerful enough to detect most of the existing small (s) RNA sequences even with low abundance, gave rise to a sudden surge in the reporting of novel mi RNAs.

Keywords: micro(mi RNA) genes, PRI-mi RNA, pre-mi RNA, PRI- mi RNA promoters, CPG islands, 5'cage tags, SIRNAS, RSRNAS, RSRNAS targets, PRNA, NORC

Volume 2 Issue 5 - 2017

Kupriyanova NS

The Institute of Gene Biology RAS, Russia

Correspondence: Kupriyanova NS, The Institute of Gene Biology RAS, Vavilov str 5/13, Moscow, Russia, Email kupriyanova-45@mail.ru

Received: September 12, 2017 | **Published:** December 21, 2017

Introduction

In 1993 Victor Ambros and colleagues discovered that *Lin-4*, a gene known to control the of *C.Elegans* larval development produces a pair of small RNAs.^{1,2} For seven years after this discovery, there was no evidence for *Lin-4*-like RNAs beyond nematodes. This all changed upon the discovery that *Let-7*, another gene in the *C.Elegans* encoded a second ~22nt regulatory RNA. Homolog's of the *Let-7* gene were soon identified in human and fly genomes, and *Let-7* RNA itself was detected in human, Drosophila, and eleven other bilateral animals.³ One year later, a total of over one hundred additional genes for tiny non coding RNAs were detected. The RNA products of these genes resembled the *Lin-4* and *Let-7* RNAs. They were ~22nt endogenously expressed RNAs and potentially processed from one arm of a stem loop precursor. They were generally conserved in evolution-some quite broadly, others only in more closely related species. But unlike *Lin-4* and *Let-7* RNAs, many of them were not expressed in distinct stages of development but were more likely to be expressed in particular cell types. Intensified cloning efforts have revealed numerous additional mi RNA genes in mammals, fish, worms, and flies.³⁻⁷ The shorter *Lin-4* RNA is now recognized as the founding member of an abundant class of tiny regulatory RNAs called micro RNAs or mi RNAs.^{3,4} Mi RNA functions include control of cell proliferation, cell death, and fat metabolism in flies,^{8,9} neuronal patterning in nematodes,¹⁰ modulation of hematopoietic lineage differentiation in mammals,^{11,12} and control of leaf and flower development in plants. Since the report of the *Lin-4* RNA and its regulation of *Lin-14* the major topics discussed were mi RNA genomics, mi RNA biogenesis, and mi RNA regulatory mechanisms.¹³

Genomics: the mi RNA genes

About a quarter of the human mi RNA genes are in the introns

of pre-mRNAs. These are preferentially in the same orientation as the predicted mRNAs, suggesting that most of these mi RNAs are processed from the introns. This arrangement provides a convenient mechanism for the coordinated expression of a mi RNA and a protein. Such coordinate expression could be useful, to explain the conserved relationships between mi RNAs and host mRNAs. A striking example of this conservation involves *Mir-7*, found in the intron of *HNRNP K* in both insects and mammals.¹⁴ The majority of worm and human mi RNA genes are isolated and not clustered.¹⁵ Orthologs of *C.Elegans Lin-4* and *Let-7* are clustered in the fly and human genomes and are co expressed. A 693 bp genomic fragment rescues the *Lin-4* deficiency, implying that all the elements required for the regulation and initiation of transcription are located in this short segment.¹³ Some of the more interesting genomic locations of mi RNA genes include those in the Hox clusters. The *Mir-10* gene lies in the Antennapedia complex of insects and in the orthologous locations in two Hox clusters of mammals.¹⁰⁻¹⁶ The Hox mi RNAs are especially good candidates for having interesting functions in animal development. Nearly all of the cloned mi RNAs are conserved in closely related animals such as human and mouse, or *C.Elegans* & *C.Briggsae*.¹⁰⁻¹⁷ Many mi RNAs are conserved more broadly among the animal lineages.^{10,16} More than a third of the *C.Elegans* mi RNAs have easily recognized homologs among the human mi RNAs.¹⁰ When comparing distant lineages, considerable expansion or contraction of gene families is apparent, the most striking example being the *Let-7*, which has four identified members in *C.Elegans* and at least 15 in human, but only one in Drosophila.^{10,16} Micro RNAs and their associated proteins appear to be one of the more abundant rib nucleoprotein complexes in the cells. Nonetheless, mi RNAs whose expression is restricted could still be missed in cloning efforts. Thus, computational approaches have been developed to complement experimental approaches to mi RNA genes identification. Gene-finding approaches that do not depend on

homology or proximity to known genes have also been developed and applied to entire genomes. The two most sensitive computational scoring tools are MiRscan, which has been systematically applied to nematode, and vertebrate candidates^{10,17} and mirseeker, which has been systematically applied to insect candid at.¹⁸ Both MiRscan and mirseeker have identified a lot of genes that were subsequently verified experimentally. This might be the situation in humans-perhaps because the vertebrate genomes used in the analysis are more highly diverged. More recently identified mammalian mi RNA genes appear relatively less likely to be conserved in fish, particularly those genes cloned from embryonic stem cells and those genes cloned from embryonic stem cells and mammalian brain and the 14 mi RNA candidates residing in a large imprinted cluster.¹⁹ These recent data suggest that the more difficult-to-clone mammalian mi RNAs are less likely to be conserved in fish and thus less likely to have been identified computationally, which implies that a confident upper bound on the number of human genes is difficult to determine using analyses that extended to fish.

Mi RNA biogenesis

Now it is universally accepted that micro (mi) non coding (nc) RNAs are important regulators of a vast number of cellular processes including cell growth, disease, embryogenesis, gene regulation, signal transduction, receptor activation, etc. Approximately 50% of mammalian mi RNAs are expressed from introns of protein-coding genes; the primary transcript (pri-mi RNA) is therefore assumed to be a part of the host transcript. More less is known about the structure of pri-mi RNAs expressed from intergenic regions. The 5' end of the pri-mi RNA is predicted from transcription start sites, CpG islands and 5' CAGE tags mapped in the upstream flanking region surrounding the precursor mi RNA (pre-mi RNA). The 3' end of the pri-mi RNA is predicted based on the mapping of poly A signals. The predicted pri-mi RNAs are also analyzed for promoter and insulator-associated regulatory regions. The two candidate RNA polymerases for pri-mi RNA transcription are pol II and pol III. Mi RNAs processed from the interns of protein-coding host genes are undoubtedly transcribed by pol II. The pri-mi RNAs can be quite long, more than one 1kb, which is longer than typical pol III transcripts. These presumed pri-mi RNAs often have internal runs uridine residues, which would be expected to prematurely terminate pol III transcription. Now it has become clear that a large number of regulatory small (rs) RNAs are compromised due to traditional approaches to identify mi RNAs mainly to hair-pin loop bred typical signs. The rsRNA-target interactions have been studied with the use of AGO-cross linking, DGCR8 knockdown, CLASH, proteome and expression data. Several of the potential rs RNAs have emerged as a critical cancer biomarker controlling some important aspects of cell system. The advent of next generation sequencing (NGS) technologies, powerful enough to detect most of the existing sRNA sequence even with low abundance, gave rise to a sudden surge in the reporting of novel mi RNAs.²⁰ NGS led discoveries not only impacted the number but also the concept of mi RNAs.²¹ Now, si RNAs are reported to be able to form endogenously also and cause the same impact as typical hairpin loop derived mi RNAs.²² A wide range of endogenous regulatory small RNAs exist within the animal cell system, originating from antisense transcripts (nati-siRNA), being generated from degradation products (rasi-RNAs) or piwi RNAs.²³⁻²⁶ Besides NGS led discoveries not only impacted the number but also the concept of mi RNAs. Now, si RNAs are reported to be able to form endogenously also and cause the same impact as typical hairpin loop derived mi RNAs.²² A wide

range of endogenous regulatory small RNAs exist within the animal cell system, originating from antisense transcripts (nati-si RNA), being generated from degradation products (rasi-RNAs) or piwi RNAs.²³⁻²⁶ A lot of regulatory small RNAs including mi RNAs have been shown to have the origin in repetitive elements.^{27,28} Several non-coding RNAs like snoRNAs, tRNAs, rRNAs and other non-coding RNAs have been reported to produce endogenous regulatory small RNAs capable to influence phenotypes in vertebrates.²⁹⁻³¹ Besides this, endogenous regulatory small RNAs are now also reported from corners which were earlier blindly filtered out from genomic studies as a practice.^{32,33} Several non-coding RNAs like snoRNAs, tRNAs, rRNAs and other non-coding RNAs have been reported to produce endogenous regulatory small RNAs capable to influence phenotypes in vertebrates.³³⁻³⁵ To understand the mechanism and conditions of the activation of micro RNA genes, it is required to locate their core promoter regions.³⁶ Developing the promoter identification algorithm is a very challenging problem. Although computational methods have been developed for predicting core promoters of protein-coding genes, their performances are not exactly correct. The situation with micro-RNA genes is even worse, and is far from satisfactory. The main reason is that our understanding of the transcription process is incomplete. For *H. Sapiens*, only the promoters of two micro RNA genes, hsa-mir23a; 27a; 24-2³⁷ and hsa-mir-371; 372; 373,³⁷ have been identified so far. The promoter of hsa-mir-23a; 27a; 24-2 has been located by biological experiments,³⁸ while the promoter of hsa-mir-371; 372; 373³⁹ has been identified by a comparative genomic analysis. Core promoter elements are highly variable, requiring sophisticated techniques for their detection. Discovering key cis-elements of micro RNA genes is more difficult, since our knowledge about the transcription of this novel family of genes is limited. Lee et al.¹ located the promoter of mir-23a; 27a; 24-2; however, none of the canonical promoter elements were discovered in this promoter.³⁷ TATA-box was found in mmu-mir-290;291;292;293;294;295.³⁹ However, the deletion of this putative TATA-containing promoter region had almost no effect on the expression level of mir292 and the precursor to mir292 in transected cell lines Ohler et al.³⁷ Scanned the 1,000-bp upstream sequences of *Drosophila* micro RNA genes for known promoter motifs but did not detect a consistent preference for any known motifs that are enriched in protein-coding genes.³⁹

Are studied and characterized the promoters of intergenic micro RNA genes in *Caenorhabditis elegans*, *Homo sapiens*, *Arabidopsis thaliana*, and *Oryza sativa*. It is shown that most known micro RNA genes in these four species have the same type of promoters as protein-coding genes have. Many significant, characteristic sequence motifs in these core promoters were also discovered. Several of them match or resemble the known Cis-acting elements for transcription initiation. Among these motifs, some are conserved across different species while some are specific to micro RNA genes of individual species. The intergenic micro RNA genes are believed to be transcribed independently and to form a new gene family, whereas the intronic ones and the ones interspersed with mobile elements Alu in the human genome can be transcribed with their host genes.⁴⁰ Primary transcripts of the murine and human ncRNAs start from the spacer promoter (about -2000bp and -1000bp of the transcription start point) correspondingly. A number of short (150-300nt) pRNAs are generated as the result of the 2000nt mouse precursor processing.⁴¹ At the same time a human 1000nt precursor contains only one pRNA copy at its 3'-end with a potential secondary structure similar for human, mouse, rat and pig It cannot be excluded that at least one additional ncRNA

can be transcribed from the region upstream of the human's spacer promoter.⁴² Another piece of experimental evidence was from a *M. Musculus* polycistronic micro RNA gene, *mmu-mir-290*; 291; 292; 293; 294; 295 Houbaviv et al.³⁶ Found a canonical TATA-box, located at ~35, of capped and polyadenylated pri-micro RNA of this gene, and showed that this upstream region was also conserved in a *H. Sapiens* homologous gene, *hsamir-371*; 372; 373.²¹ All these results are fundamentally important; they have provided direct evidence that a micro RNA gene can be transcribed by pol II. However, a few critical questions remain unanswered. One of them is whether all known micro RNA genes of different species are class-II genes.

Mi RNA regulatory mechanisms

It is shown that short intergenic RNA molecules covering the rDNA promoter bind to NoRC (nucleolar remodeling complex) and this association with IGS transcripts is required for NoRC binding to chromatin and heterochromatin formation. RNase a treatment shows that RNA is involved in the association of NoRC with rRNA genes. It was also shown in electrophoresis mobility shift assays (EMSA) experiments, that TIP5 derivatives containing the TAM domain retarded RNA, indicating that the TAM domain mediates the interaction of NoRC with RNA. According to,^{43,44} intergenic RNAs are transcribed by Pol I from a spacer promoter ~2000nt upstream of the murine major gene promoter, and long primary transcripts are processed into small intermediates 150-250 nucleotide (nt) RNAs that overlap the rDNA promoter (pRNA) that appear to be rapidly degraded unless bound to TIP5 (TIF interacting protein 5), the large subunit of the chromatin re modelling complex NoRC. NoRC dependent repression of Pol I transcription is not species specific although mammalian rDNA promoters share little sequence homology. Upon searching for the remodeling activity that alters the chromatin structure of the rDNA promoter, NoRC was identified and purified. This complex plays not only an essential role in heterochromatin formation and transcriptional silencing but also provides evidence for a link between chromatin remodeling, histone modifications, and de novo methylation. NoRC interacts with the N-terminal part of TTF-I, and this interaction enables TTF-I to bind to its cognate sequence upstream of the gene promoter. The interaction between TTF-I and NoRC brings NoRC to the rDNA promoter. NoRC binds near the promoter and subsequently represses rDNA transcription. This repression requires the presence of NoRC prior to the recruitment of the transcription initiation factors and depends upon the presence of the histone H4 tail. The results suggest that TTF-I may recruit different remodeling activities to rDNA which either activate or repress Pol I transcription. NoRC is a member of the ISWI family of ATP-dependent chromatin remodeling complexes. It consists of Tip5 and Snf2H, the mammalian homolog of the ATPase ISWI. RNase a treatment shows that RNA is involved in the association of NoRC with rRNA genes. It was also shown in electrophoretic mobility shift assays (EMSA) experiments, that TIP5 derivatives containing the TAM domain retarded RNA, indicating that the TAM domain mediates the interaction of NoRC with RNA. Computational approach showed that RNAs corresponding to rDNA promoters of several mammals, including human, mouse, rat and pig, can fold into a common secondary structure. The most striking feature of this hypothetical structure is a hairpin that contains rDNA sequences from nucleotides -127 to -49 in mouse and from -137 to -50 in human pRNA. Transcriptional repressor CTCF also known as 11-zinc finger protein or CCCTC binding transcription factor in humans is encoded by the CTCF gene. CTCF is involved in insulator activity. Absence of CTCF in cultured cells resulted in the decreased association of UBF

with rDNA and in nucleolar fusion. CTCF may load UBF onto rDNA, thereby forming part of a network that maintains rDNA genes poised for transcription. Although special features of the rRNA expression regulation are poorly studied in humans, a number of efforts have been made for a selective action on rRNA transcription in cancer cells. It was found that some G-rich libraries (composed of G+T or G+C nucleotides) strongly inhibited cancer cell growth while sparing non-malignant cells. The question remains as to whether the deregulation of rRNA synthesis itself could trigger cell transformation,^{45,46} or whether increased rRNA synthesis plays a secondary, but necessary, part in tumor genesis. Potential targets for anticancer therapeutic strategy are protein kinases, such as ERK/RSK, mTOR, and CK2, which are often hyper, activated in cancer cells and are known to be required for rRNA transcription. Several approved anticancer drugs have been shown to inhibit rRNA synthesis, albeit not necessarily with the required selectivity. There is a growing list of additional factors with oncogenic and tumor suppressor activity implicated in the modulation of RNA Pol I during malignancy.⁴⁷ Although there are multiple ways to inhibit Pol I transcription and pre-rRNA processing, the vast majority do it in a nonselective way. Cylene Pharmaceuticals identified a small molecule that selectively inhibited Pol I transcription, CX5461.⁴⁸ CX5461 inhibits Pol I transcription at the initiation step by interfering with SL1/rDNA promoter binding. In vitro characterization identified cell lines derived from hematologic malignancies with those possessing wild-type p53 being particularly susceptible to CX5461. Normal cells were found to be resistant.⁴⁸

Mi RNAs in *H. Sapiens* RIGS

Intensified cloning efforts have revealed numerous additional mi RNA genes in mammals, fish, worms, and flies. The RNA products of these genes resembled the *Lin-4* and *Let-7* RNAs. They were ~22nt endogenously expressed RNAs and potentially processed from one arm of a stem loop precursor. They were generally conserved in evolution - some quite broadly, others only in more closely related species. But unlike *Lin-4* and *Let-7* RNAs, many of them were not expressed in distinct stages of development but were more likely to be expressed in particular cell types. More than a third of the *C. Elegans* mi RNAs have easily recognized homologs among the human miRNAs.⁴¹ When comparing distant lineages, considerable expansion or contraction of gene families is apparent, the most striking example being the *Let-7*, which has four identified members in *C. Elegans* and at least 15 in human, but only one in *Drosophila*.^{40,41} Regulation of the rRNA transcription in Homo sapiens is as yet poorly studied. Silencing of ribosomal RNA genes (rDNA) requires binding of the chromatin remodelling complex NoRC to RNA that is complementary to the rDNA promoter. NoRC-associated RNA (pRNA) folds into a conserved stem-loop structure that is required for nucleolar localization and rDNA silencing. Mutations that disrupt the stem-loop structure impair binding of TIP5, the large subunit of NoRC, to pRNA and abolish targeting of NoRC to nucleoli. To understand the mechanism and conditions of the activation of microRNA genes, it is required to locate their core promoter regions, but developing the promoter identification algorithm is a very challenging problem. Transcripts originating from a spacer promoter located upstream from the pre-ribosomal RNA transcription start site have been shown to be important in ribosomal RNA gene (RDNA) silencing.⁴⁹ The data suggested that short-lived intergenic transcripts are processed into 150-250 nt. RNAs that overlap the rDNA promoter. This "promoter RNA" (pRNA) is stabilized by binding to TIP5, the large subunit of the chromatin remodeling complex NoRC. Depletion of pRNA causes

translocation of NoRC from nucleoli to the nucleoplasm, whereas 'refeeding' with ectopic pRNA restores nucleolar localization. This targeting process requires both the hairpin structure of pRNA and sequences upstream from the hairpin. In the search for structural motifs that are shared by pRNA from various species, a computational approach showed that RNAs corresponding to rDNA promoters of several mammals, including human, mouse, rat and pig, can fold into a common secondary structure. The most striking feature of this hypothetical structure is a hairpin that contains rDNA sequences from nucleotides -127 to -49 in mouse and from -137 to -50 in human pRNA. A considerable difference in the rIGS pre-promoter region organization in the human and other known vertebrates makes it interesting to compare nucleotide sequences of the regulatory RIGS regions in the human and evolutionary intermediate great apes. It turned out that the RIGS pre-promoter region part employed for the comparison with its orthologs from *Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla* and *Pongo pygmaeus* reveals a high extent of homology.⁵⁰

Unexpectedly we disclosed that the fragment ~100bp downstream -2000bp in the human RIGS reveals high (about 60%) homology to the universally known *Let-7 gene* from the *C. elegans* (our unpublished data). We have studied transcription ability of the RIGS region between -2140 and -1300bp using 7 overlapping primer pairs. The results obtained showed this entire region to be transcriptionally active. A special test taking into account the order of the primers addition in the course of the RNA reverse transcription has shown that ncRNA transcripts are coding on the leader chain that corresponds to the major rRNA product.⁵¹ This transcript (~200bp long) was localized upstream the transcription start point (TSP) of the rRNA for the brain glioma, line 313p, between the Alu-element nearest to TSP, and (CCCT)_n microsatellite cluster. The transcription activity of the region between -2140, and -1300bp upstream TSP was also studied for some other cancer cells lines, namely H-C, M, IMG K2476, K562, and 313p. The seven pairs of overlapping primers were used in these experiments and it was shown that the rIGS region between -2140 and -1300bp is able to produce RNA. Some problems with transcription activity arose only in the (CCCT)_n microsatellite region which most probably is able to form G quadruplexes.⁵² It is interesting that the region between -1840 and -1700 exhibits a high extent of homology with miRNAs from ATAD2, a large family of ATP ases, containing ATP binding sites, as well as the gene GCN5L2 which has been shown to interact with Ku70, TAF9, Ku80, and DDB1, the proteins that take part in the NORC assembling and action. This region is enabling to form hair-pin characteristic for miRNAs, and present in all rDNA repeats revealing 100% conservatively. However a question of possible functions and transcription modulation of this putative miRNA remains unsolved.

Conclusion

In the last years it has become increasingly clear that the mammalian transcriptome is highly complex and includes a large number of small non-coding RNAs (sncRNAs) and long noncoding RNAs (lncRNAs). The three classes of sncRNAs, namely short interfering RNAs (siRNAs), microRNAs (mi RNAs) and PIWI-interacting RNAs (piRNAs) have been extensively studied and are involved in pathways leading to specific gene silencing and the protection of genomes against virus and transposons, for example. SncRNAs and lncRNAs play critical roles in defining DNA methylation patterns, as well as chromatin remodeling thus having a substantial

effect in Epigenetics. The identification of some overlaps in their biogenesis pathways and functional roles raises the hypothesis that these molecules play concerted functions *In vivo*, creating complex regulatory networks where cooperation with regulatory proteins is necessary. The implications of biogenesis and gene expression deregulation of sncRNAs and lncRNAs were also highlighted in a number of human diseases like cardiac ageing,⁵² gastric, and other types of cancer.^{53,54} Over the recent years, Next Generation Sequencing (NGS) technologies targeting the micro RNA transcriptome revealed the existence of many different RNA fragments derived from small RNA species other than micro RNA. Although initially discarded as RNA turnover artifacts, accumulating evidence suggests that RNA fragments derived from small nucleolar RNA (snoRNA) and transfer RNA (tRNA) are not just random degradation products but rather stable entities, which may have functional activity in the normal and malignant cell. New findings describing the detection and alterations in expression of snoRNA-derived (sdrRNA) and tRNA-derived (tRF) RNAs were done. It is possible that there exist a number of interactions of sdrRNAs and tRFs with the canonical micro RNA pathways in the cell.⁵⁵ The increasing appreciation of the central role of non-coding RNAs (mi RNAs and long non-coding RNAs) in chronic and degenerative human disease makes them attractive therapeutic targets. This would not be unprecedented: the bacterial ribosomal RNA is a mainstay for antibacterial treatment, while the conservation and functional importance of viral RNA regulatory elements has long suggested they would constitute attractive targets for new antivirals. Oligonucleotide-based chemistry has obvious appeals but also considerable pharmacological limitations that are yet to be addressed satisfactorily. Recent studies identifying small molecules targeting non-coding RNAs may provide an alternative approach to oligonucleotide methods. New structural and chemical principles for targeting RNA with small molecules are intensively devising today.⁵⁶ Although, the field of ncRNAs has been growing fast we are still far from understanding the complexity and the mechanisms underlying the establishment of the regulatory networks between RNAs and proteins. It seems likely that the crossroads of different human pathologies range from cancer to neurodegenerative and immune diseases. Finally the continued understanding of the molecular mechanisms and signaling pathways where ncRNAs participate should offer new diagnostic strategies and open new avenues for therapies.

Acknowledgements

This work was supported by the Russian Fund of Fundamental Investigations (RFFI), grant number 16-04-00178, and Molecular and Cell Biology of the Presidium of the Russian Academy of Sciences.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Lee R, Feinbaum R, Ambros V. A short history of a short RNA. *Cell*. 1993;116(2 Suppl):89–92.
2. Lagos Quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed RNAs. *Science*. 2001;294(5543):853–858.
3. Pasquinelli AE, Reinhart BJ, Slack F, et al. Conservation across animal phylogeny of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature*. 2000;408:86–89.

4. Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*. 2001;294(5543):862–864.
5. Lim LP, Glasner ME, Yekta S, et al. Vertebrate microRNA genes. *Science*. 2003;299(5612):1540.
6. Harpreet K Saini, Anton J Enright, Sam Griffiths Jones. Annotation of mammalian primary microRNAs. *BMC Genomics*. 2008;9:564.
7. Altuvia Y, Landgraf P, Lithwick G, et al. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res*. 2005;33(8):2697–2706.
8. Brennecke J, Hipfner DR, Stark A, et al. bantam encodes a developmentally regulated micro RNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell*. 2003;113(1):25–36.
9. Xu P, Vernoooy SY, Guo M, et al. The *Drosophila* microRNA mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol*. 2003;13(9):790–795.
10. Johnston RJ, Hober O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature*. 2003;426:845–849.
11. Chen CZ, Li L, Lodish HF, et al. MicroRNAs modulate hematopoietic lineage differentiation. *Science*. 2004;303(5654):83–86.
12. Lagos Quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed si- RNAs. *Science*. 2001;294(5543):853–858.
13. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. 2004;116(2):281–297.
14. Heikham R, Shankar R. Flanking region sequence information to refine microRNA target predictions. *J Biosci*. 2010;35(1):105–118.
15. Lim LP, Lau NC, Weinstein EG, et al. The micro-RNAs of *Caenorhabditis elegans*. *Genes Dev*. 2003;17(8):991–1008.
16. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol*. 2006;13(12):1097–1101.
17. Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*. 2005;11(3):241–247.
18. Prlc A, Down TA, Kulesha E, et al. Integrating sequence and structural biology with DAS. *BMC Bioinformatics*. 2007;8:333.
19. Griffiths Jones S, Saini HK, Van Dongen S, et al. miRBase: tools for microRNA genomics. *Nucleic Acids Research D*. 2008. p. 154–158.
20. Jha A, Chauhan R, Mehra M, et al. miR-BAG: bagging based identification of microRNA precursors. *PLoS One*. 2012;7(9):e45782.
21. Jha A, Panzade G, Pandey R, et al. A legion of potential regulatory sRNAs exists beyond the typical microRNAs microcosm. *Nucleic Acids Res*. 2005;43(18):8713–8724.
22. Jha A, Chauhan R, Mehra M, et al. miR-BAG: bagging based identification of microRNA precursors. *PLoS One*. 2012;7(9):e45782.
23. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. *Cell*. 2009;136(4):642–655.
24. Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. *Genes Dev*. 2003;17(4):438–442.
25. Vagin VV, Sigova A, Li C, et al. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*. 2006;313(5785):320–324.
26. Brennecke J, Aravin AA, Stark A, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 2007;128(6):1089–1103.
27. Talmor Neiman M, Stay R, Klipcan L, et al. Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. *Plant J*. 2006;48(4):511–521.
28. Borchert GM, Holton NW, Williams JD, et al. Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. *Mob Genet Elements*. 2011;1(1):8–17.
29. Smalheiser NR, Torvik VI. Mammalian microRNAs derived from genomic repeats. *Trends Genet*. 2005;21(6):322–326.
30. Maute RL, Schneider C, Sumazin P, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl Acad Sci U S A*. 2013;110(4):1404–1409.
31. Auyeung VC, Ulitsky I, McGeary SE, et al. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*. 2013;152(4):844–58.
32. Maute RL, Schneider C, Sumazin P. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl Acad Sci U.S.A*. 2013;110(4):1404–1409.
33. Smalheiser NR, Torvik VI. Mammalian microRNAs derived from genomic repeats. *Trends Genet*. 2005;21(6):322–326.
34. Lee YS, Shibata Y, Malhotra A, et al. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev*. 2009;23(22):2639–2649.
35. Xuefeng Zhou, Jianhua Ruan, Guandong Wang, et al. Characterization and Identification of MicroRNA Core Promoters in Four Model Species. *PLoS Comput Biol*. 2007;3(3):e37.
36. Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. 2004;23(20):4051–4060.
37. Ohler U, Yekta S, Lim L, et al. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA*. 2004;10(9):1309–1322.
38. Houbaviv H, Dennis L, Jaenisch R, et al. Characterization of a highly variable eutherian microRNA gene. *RNA*. 2005;11(8):1245–1257.
39. Lee Y, Jeon K, Lee J, et al. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO J*. 2002;21(17):4663–4670.
40. Zeng Y, Cullen BR. Sequence requirements for micro RNA processing and function in human cells. *RNA*. 2003;9(1):112–123.
41. Ender C, Krek A, Friedlander MR. A human snoRNA with microRNA-like functions. *Mol Cell*. 2008;32(4):519–528.
42. Mayer C, Schmitz KM, Li J, et al. Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell*. 2006;22(3):351–361.
43. Santoro R, Grummt I. Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation. *Mol Cell Biol*. 2005;25(7):2539–2546.
44. Maggi LB, Weber JD. Nucleolar adaptation in human cancer. *Cancer Invest*. 2005;23(7):599–608.
45. Trere D, Ceccarelli C, Montanaro L, et al. Nucleolar size and activity are related to pRb and p53 status in human breast cancer. *J Histochem Cytochem*. 2004;52(12):1601–1607.
46. Hannan KM, Sanij E, Rothblum LL, et al. Dysregulation of RNA polymerase I transcription during disease. *Biochim Biophys Acta*. 2013;829(3–4):342–360.
47. Drygin D, Lin A, Bliesath J, et al. Targeting RNA polymerase I with an oral small molecule CX- 5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res*. 2011;71(4):1418–1430.

48. Santoro R, Grummt I. Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation, *Mol Cell Biol*. 2005;25(7):2539–2546.
49. Netchvolodov KK, Boiko AV, Ryskov AP, et al. Evolutionary Divergence of the pre-promoter region of ribosomal DNA in the Great Apes. *DNA Sequence*. 2006;17:378–391.
50. Sadova AA, Cherepanova MD, Kupriyanova NS, et al. Mapping of non-coding RNAs in the human Ribosomal Intergenic Spacer. *Actual problems of humanitarian and natural sciences*. 2013;3:51–56.
51. Netchvolodov KK, Kurova VS, Kononichin AS, et al. DNA-dependent protease complexes with single-stranded oligo-(AGGG)₆: Identification and possible role in the ribosomal RNA transcription modulation. *Dokl Biochem Biophys*. 2009;424:118–121.
52. Gomes AQ, Nolasco S, Soares H. Non-Coding RNAs: Multi-Tasking Molecules in the Cell. *Int J Mol Sci*. 2013;14(8):16010–16039.
53. Wang H, Bei Y, Shi J, et al. Non-Coding RNAs in Cardiac Aging. *Cell Physiol Biochem*. 2015;36(5):1679–1687.
54. Li PF, Chen SC, Xia T, et al. Non-coding RNAs and gastric cancer. *World J Gastroenterol*. 2014;20(18):5411–5419.
55. Martens Uzunova ES, Olvedy M, Jenster G. Beyond microRNA--novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett*. 2013;340(2):201–211.
56. Shortridge MD, Varani G. Structure based approaches for targeting non-coding RNAs with small molecules. *Curr Opin Struct Biol*. 2015;30:79–88.