FUNCTIONAL AND EVOLUTIONARY IMPLICATIONS OF SINGLE NUCLEOTIDE SUBSTITUTIONS IN HUMAN MICRORNAS ACROSS PRIMATES

María López Valenzuela

TESI DOCTORAL UPF / 2014

DIRECTOR DE LA TESI

Dra. Yolanda Espinosa Parrilla

Institut de Biologia Evolutiva

Ciències Experimentals i de la Salut

upf. Universitat Pompeu Fabra *Barcelona*

Para el taperclub. De lejos.

ACKNOWLEDGEMENTS

A Yolanda y bioevo en general: un agradecimiento académico, respetuoso y alegre.

A la familia y la banda que hicieron de los últimos años una de las épocas más bonitas de mi vida: un agradecimiento fraternal y cómico-festivo.

A quienes me apoyaron en las últimas semanas: el agradecimiento eterno del sobreviviente.

Y a Marc, que me ha enseñado que hay vida después de la tesis: un agradecimiento con todo mi corazón.

ABSTRACT

MicroRNAs (miRNAs) are major contributors to phenotypic diversity and have a demonstrated role in evolution. We performed an analysis on how miRNA nucleotide substitutions may have contributed to human divergence from their closest primate relatives. We first studied the functional implications of a substitution located in the seed region of mir-1304, reported to differ between present-day humans and Neandertals. We found a large change in the set of targets predicted for the two alleles, suggesting an important functional evolution for mir-1304. Among the few target genes predicted for the ancestral allele we found two genes involved in tooth development, ENAM and AMTN. Functional analysis revealed that these genes are differentially regulated by the two mir-1304 alleles suggesting that this change may have contributed to the modulation of phenotypic differences found between present-day humans and Neandertal dentitions. The thesis also offers a global view of the genetic diversification occurred in miRNA regions since the split of humans and chimpanzees. After careful alignment of all human miRNA sequences to chimpanzee and orangutan genomes, their substitution rates were calculated and compared between miRNA categories and with neutral sequences. Despite of the high conservation of these regulators, primate-specific miRNAs showed significantly higher substitution rates than more conserved miRNAs. This suggests that miRNA-driven evolution in primates could be partially sustained by mutation in novel, primate-specific miRNAs and that different miRNA categories show different evolutionary constraints and thus shall not be considered as an homogeneous group.

RESUMEN

Los microRNAs (miRNA) contribuyen de manera importante a la diversidad fenotípica y tienen un papel demostrado en la evolución. En la presente tesis se realizaron análisis sobre la manera en que las sustituciones nucleotídicas en los miRNAs pudieran haber contribuido a la divergencia entre el humano y sus parientes primates más cercanos. Primero se estudiaron las implicaciones funcionales de un cambio localizado en la "seed-region" del mir-1304, que según se ha reportado difiere entre los humanos de hoy en día y los Neandertales. Al predecir los genes diana para estas dos versiones de mir-1304 encontramos un enorme cambio en número, lo que sugiere una importante evolución para mir-1304 a raíz de esta substitución nucleotídica. Entre los pocos genes diana previstos para el alelo ancestral, se encontraron dos genes implicados en el desarrollo de los dientes, ENAM y AMTN. Su análisis funcional reveló que estos genes pueden ser regulados diferencialmente por los dos alelos mir-1304, lo que sugiere que este cambio puede haber contribuido a la modulación de las diferencias fenotípicas encontradas entre las denticiones de los seres humanos de hoy en día y los Neandertales. La tesis también ofrece una visión global de la diversificación genética en las secuencias de miRNAs desde la separación de los seres humanos y los chimpancés. Después de alinear cuidadosamente todos los genes de miRNA humanos a los genomas del chimpancé y el orangután, se calcularon las tasas de substitución y se compararon agrupados por categorías y con respecto a referencias neutrales. En general se observó una alta conservación en estos reguladores, sin embargo los miRNAs primate-específicos mostraron significativamente más altas tasas de sustitución que los miRNAs más con una distribucion filogenética más amplia. Esto sugiere por una parte que la evolución impulsada por miRNAs en primates podría, al menos parcialmente, deberse a mutaciones en los miRNAs más jóvenes y específicos de los primates y por otro lado que los microRNAs no son una clase de reguladores homogénea sino que dentro de ella existen categorías sujetas a diferentes presiones evolutivas.

PREFACE

The present thesis aimed to improve the knowledge of the role of miRNAs as contributors to the diversification of the primate lineage, in particular for human and their closest relatives, Neanderthal and chimpanzee. By the starting time of this thesis the main repository of miRNAs counted only 750 human miRNA precursors. Currently, 7 releases after, this number has more than doubled reaching 1881 known human miRNA precursors and 2588 mature miRNAs. The discovery of miRNAs in other species has also increased greatly, although not at the same pace of human, and provides a rich resource for evolutionary analyses. Also along these years great advances such as the largely awaited completion of the Neandertal genome and the suppressive finding of the Denisovan have revolved the world of ancient DNA study and added more pieces to the intriguing puzzle of human evolution. Apart from extinct humans some of the extant primates have also had their genomes sequenced for the first time and some genome assemblies have been greatly improved. This thesis has profited from all of these advances and has tried to integrate them to broaden the vision about the role of miRNAs in the recent evolution of humans, but also points that a lot more remains to be done.

Table of Contents

1. INTRODUCTION

1.1 Non-protein-coding RNA

In addition to DNA sequence of genes that code for proteins, the genomes of all cellular life forms contain varying proportions of non-protein-coding sequences. This fraction is abundant on the genomes of higher eukaryotes but very small on prokaryotic genomes where the foundations of molecular biology were first elucidated, and so historically these extensive sequences that do not code for protein have been considered inert evolutionary debris. Nevertheless much of these DNA sequences are transcribed, the RNAs produced from these sequences are known as non-coding RNA s (ncRNAs).

About twenty years ago the collection of known ncRNAs was still scarce, composed mostly of RNAs with generic housekeeping roles such as transfer RNAs, ribosomal RNAs and small nucleolar RNAs (Bussotti, Notredame, & Enright, 2013). By then only a few RNAs with regulatory function, like lin-4 (R. C. Lee, Feinbaum, & Ambros, 1993) which controls the developing timing of C. elegans or XIST involved in the inactivation of X chromosome in mammals (C. J. Brown et al., 1992), were known. Later, the finding of the prominent role of RNAs in gene silencing (Fire, Xu, Montgomery, & Kostas, 1998) greatly accelerated the pace of discovery of novel functional ncRNAs, in particular short sequences of approximately 20-30 nucleotides known as short non-coding RNAs (sncRNAs).

The identification of new ncRNAs has grown as new technologies such as next generation sequencing or tiling arrays were developed and became more available. Also ambitious collaborative projects promoted by international consortia of researchers have been launched with the aim of fully characterize the transcriptome profiles of various organisms and even to detail the differences between tissues, developmental stages or pathological states. Examples of such a large endeavors are the ENCODE project which provided a detailed landscape of transcription in human cells. Such studies result in the identification of thousands of novel regulatory sncRNAs as well as the pervasive transcription of long ncRNAs (lncRNAs) (Derrien et al., 2012).

ncRNAs range from those with distinctive structures and canonical biogenesis pathways, like microRNAs (miRNAs), to the ones with diverse functions or none yet described (Figure 1. Diverse classes of RNAs). The biological relevance of great part of these ncRNA remains to be determined, but increasing evidence shows that ncRNA species discovered in recent years can perform a wide repertoire of biological functions. ncRNAs appear to comprise yet another layer of complexity in the control of gene expression in physiology and development, including epigenetics, chromatin architecture, transcription and RNA splicing, editing, translation and turnover They also play a significant role in disease conditions and constitute an unexplored world of genetic variation both within and between species reviewed in (J. Li & Zhang, 2012; Morris & Mattick, 2014; Patil, Zhou, & Rana, 2013).

1.2 RNA interference

miRNA-mediated gene regulation, the focus of this thesis, belongs to a wider system known as RNA interference (RNAi), which is briefly revised below. The RNAi system is an ancient gene regulation program present in a wide range of eukaryotic organisms. RNA interference is a simple and effective method for silencing gene expression. It consists, in general terms, of small single strand RNA molecules that coupled to effector proteins destabilize longer RNA targets that are recognized in a sequence-specific manner (Shabalina & Koonin, 2008).

Although RNAi is an eukaryotic signature, similar mechanisms exist in prokaryotes. Some regulatory RNAs have been found in bacteria (Majdalani, Vanderpool, & Gottesman, 2005) and although the effector proteins involved in the prokayiotic systems are not homologous to the eukaryotic ones, still show some functional analogies to RNAi mechanisms like those led by microRNAs (miRNAs) and small interfering RNAs (siRNAs) in eukaryotes (Aiba, 2007; Sorek, Kunin, & Hugenholtz, 2008). These eukaryotic systems are described on later sections. The effector protein of one of these RNAi-like mechanisms in prokaryotes is the RNA chaperonin Hfq which is not homologous to any protein involved in

eukaryotic RNAi but rather to the core domain of an eukaryotic spliceosome protein (Aiba, 2007). Another pathway known as CAS (CRISPR Associated System) serves as an anti phage system in both bacteria and archea. The fundamental difference of this pathway with the eukaryotic RNAi is that the exogenous genetic content of the phage is integrated to the prokaryotic genome into CRISPR repeats, that have some resemblance to Piwi-interacting RNA (piRNA) clusters detailed below, and then are transcribed as sncRNAs (Barrangou et al., 2007; Makarova, Grishin, Shabalina, Wolf, & Koonin, 2006; Sorek et al., 2008).

The RNAi protein machinery seems to have three different prokaryotic roots; while the helicase domain of Dicer and Ago has an archeal origin the RNAse domain of Dicer shows a bacterial origin and the RNA dependent RNA polymerase (RdRP) derives from viruses (Shabalina & Koonin, 2008).

An overview of the main proteins involved in this system and their RNA counterparts is given below.

1.2.1 The proteins in RNAi

Eukaryotic RNAi systems have a prominent role in gene regulation as well as in transposon silencing, antiviral defense and chromosomal modification. Interestingly, all this diversity of function is attained through the joint action of small non coding RNAs and multiple paralogous versions of just a few key proteins, mainly AGO-PIWI protein and a Dicer-like protein that typically contains an helicase and a RNAse III domain. These proteins have suffered numerous duplications along the eukaryotic evolution. Different paralogous in association with different sncRNA lead to a combinatorial wide range of functions .

a) Dicers

Functional ncRNAs are typically 20 - 30 nt long but usually need to be excised from larger precursors. This is the function of Dicer enzymes, a large family of RNase III endonucleases that are divided in three classes. Those in class I have only one catalytic domain while class II and III have at least two and differ in their amino

terminal domains. Drosha, also involved on the first steps of miRNA processing, is a class II RNase III with proline rich and arginine-serine (RS) domains.

Dicers on the other hand pertain to class III endonucleases, are characterized by a DEXDCH ATPase on their amino terminus followed by a DUF283 domain, a PAZ (PIWI - AGO - Zwille) domain and two tandem endonucleolytic RNase III domains. Some Dicers also contain up to two dsRBD on their carboxyl ends (Carthew & Sontheimer, 2009). The crystallographic structure of Dicer has been resolved shedding light into their mode of function (Macrae et al., 2006). PAZ domains, are specialized in binding RNA duplexes with ~2nt 3' overhangs. The cargo RNA bind the PAZ domain and then extends approximately two helical turns to reach the processing center located on a cleft between the two RNase III domains each of these cleavages one strand leaving appropriate overhangs that will be needed later in the process of interference (Carthew & Sontheimer, 2009; MacRae, Zhou, & Doudna, 2007). ATPase domains remain enigmatic, some Dicers lack this domains and their function is not impaired, as it is the case of the human Dicer. In contrast the disruption of Dicer 2 in Drosophila has been reported to abolish RNA processing (Carthew & Sontheimer, 2009).

Dicers are found throughout eukaryotes, with the number of homologous varying from group to group. In organisms that have multiple paralogous copies Dicers tend to specialize in processing one kind of sncRNA. This is the case of Drosophila where Dicer1 cleaves miRNAs and Dicer2 is involved in the siRNA biogenesis . Plants have four versions of Dicer and each one couples with a different kind of sRNA (Xie, Allen, Wilken, & Carrington, 2005) with Dicer1 involved specifically on miRNA processing (Baulcombe, 2004). On the other hand, mammals as well as cold-blooded vertebrates have only one Dicer protein, which has direct orthologous in urochordates and deuterostomes. The only human Dicer process both miRNAs and siRNAs. As for protostomes, nematodes as *Caenorhabditis elegans* have only one Dicer while Drosophila and some species of mosquito present two Dicers (Murphy, Dancis, & Brown, 2008), again with Dicer1 in charge of miRNA processing (Y. S. Lee et al., 2004). Dicers in Drosophila are one of many examples of lineage-specific gene duplications

followed by sub-functionalization, in this case towards miRNA biogenesis) over the course of eukaryotic evolution.

b) Argonautes

The Argonaute (AGO) protein family is a highly diverse one, with numerous variants identified in plants, fungi, vertebrates and invertebrate animals (Murphy et al., 2008). It is the largest group of proteins specifically involved in RNAi. The number of Argonaute paralogous identified in different organisms ranges from one in *Saccharomyces pombe* to more than twenty in *Caenorhabditis elegans*. Ten members have been identified in *Arabidopsis thaliana*, five members in *Drosophila melanogaster* (Meister et al., 2004). Phylogenetic analyses shows two groups within the family, the AGO clade and PIWI clade. Metazoan AGOs are monophyletic while plant and fungal AGOs seem to form separated groups. For the PIWI cluster phylogenetic analyses suggest that the multiple copies present in mammals arose from vertebrate-specific duplication events (Murphy et al., 2008).

Eight AGOs are present in the human genome; four of them belong to the wider PIWI clade (PIWIL1 to 4) and associate to piRNAs (Carmell, Xuan, Zhang, & Hannon, 2002; Sasaki, Shiohama, Minoshima, & Shimizu, 2003) while the other four pertain to the AGO clade (AGO1, to 4) and show orthologous in other mammals and vertebrates. These AGOs preferentially bind to shorter sncRNAs: miRNAs and siRNAs. Phylogenetic analysis suggests that there was an early chordate radiation of the AGO gene family, possibly with the miRNA component of AGO2 as the ancestor of the other 3 AGOs. In human AGO1, AGO3 and AGO4 are located close to each other on chromosome 1 which also suggests a common evolution from a series of concurrent gene duplications. Among the four mammalian AGO proteins only AGO2 possesses slicer activity whereas the Drosophila Ago1 and Ago2 proteins carry weak and robust slicer activities, respectively (Patil et al., 2013). On Drosophila AGO1 and AGO2 compose both form part of RNA-induced silencing complexes (RISCs) but have different sRNA specificities, AGO1 is preferentially associated to miRNAs and AGO2 to siRNA (Okamura, Ishizuka, Siomi, & Siomi, 2004).

Proteins in the AGO clade are expressed ubiquitously in many organisms while PIWI proteins are predominantly expressed on germline cells, although they have also been found to a much lesser extent on somatic cells (Lin & Yin, 2008; Yin & Lin, 2007). Both AGO1 and AGO2 are localized in mRNA decay centers known as cytoplasmic bodies or p-bodies (Sen & Blau, 2005). Roles for AGO3 and AGO4 are unclear but they might somehow support cell differentiation in multicellular organisms (Hengst, Cox, Macosko, & Jaffrey, 2006).

AGO proteins are characterized by three structural domains: At the N-term the PAZ domain, also present in Dicer, recognizes and binds the 3' end of sRNAs duplexes and is also involved in protein-protein interactions (Lingel, Simon, Izaurralde, & Sattler, 2004; Ma, Ye, & Patel, 2004; Yan et al., 2003). PIWI domain is a cleft with endonuclease (slicer) activity similar to that in RNase H (Song, Smith, Hannon, & Joshua-Tor, 2004; Y. Wang, Sheng, Juranek, Tuschl, & Patel, 2008). The MID domain anchors the 5' monophosphate of sRNA to the protein ensuring multiple cycles of target cleavage (Ma et al., 2004; J. S. Parker, Roe, & Barford, 2005). Recently the crystal structure of human AGO2 loaded with miR-20a was solved and showed for the first time that the guide RNA spans through the entire protein in contact with every domain (Elkayam et al., 2012).

1.2.2 Small non-coding RNAs

As mentioned above numerous of sncRNAs are currently known and more are discovered continuously. sncRNAs involved in RNAi vary in length, structures, biogenesis pathways, associated effector proteins and biological roles but some similarities allow to recognize three major classes of sncRNAs: miRNAs, siRNAs and piRNAs.

a) small interfering RNAs (siRNAs)

Small interfering RNAs (siRNAs) are derived from long double-stranded precursor RNAs (dsRNAs) and unlike miRNAs are fully complementary to their targets. Depending on the source of dsRNA precursor, siRNAs can be further divided into exogenous and

endogenous siRNAs. Virus-derived RNAs (vsiRNA) are an example of exogenous siRNAs. The siRNA pathway is a key constituent of the antiviral defense mechanism in flies, worms and plants. Infection of host cells by RNA viruses initiates viral genome replication and the formation of long dsRNA intermediates that are specifically recognized and processed into vsiRNAs by Dicer proteins and mounted into RISCs to slice other viral long RNAs. This strategy to destroy viral RNA provides specific and effective defense against invading viral pathogen on this organisms. vsiRNAs have also been identified in mammalian cells infected with RNA viruses; however, it is not yet known whether such RNAs have gene silencing functions or whether the siRNA pathway is required for antiviral immunity in mammals. siRNAs can also be derived from long dsRNAs transcribed from plant or animal genomes, in which case they are considered endogenous siRNAs (endo-siRNAs). They were first detected in plants and C. elegans but were later found also in flies and mammals, suggesting that they are common to most eukaryotes. Endo-siRNAs are commonly derived from transposons and repetitive elements and function to control the expression of these elements [reviewed in (Ghildiyal & Zamore, 2009)].

b) MicroRNAs (miRNAs)

miRNAs are an abundant class of sncRNAs present in most eukaryotic lineages. Their typical length is 21-22 nucleotides and function as repressors of mRNAs. In contrast to siRNAs, miRNAs are derived from their own genes and show only partial complementarity to their targets.
Section 1.3 of this introduction offers a more detailed description on miRNA biogenesis, functions and biological relevance.

c) PIWI-interacting RNAs (piRNAs)

piRNAs were originally found on Drosophila and their biogenesis and functions are best understood on that model (Aravin et al., 2003). piRNAs are primarily expressed on the animal germline where they act as guardians of the genome protecting it from the hazardous effects of transposons (Vagin et al., 2006). Unlike miRNAs and siRNAs, piRNAs are produced from long single-stranded precursors in a Dicer-independent manner. Also they are longer (24 - 29 nt) and lack the complementary sequence found in

other sncRNA precursors (Houwing et al., 2007; Vagin et al., 2006). piRNAs are frequently located clustered on determined genomic regions. On Drosophila these regions are also rich in transposons, both active and also fragmented and thus not able to move. The majority of these clusters are transcribed bidirectionally, which produces sense transposons and complementary antisense precursors of piRNAs. Once excised from their precursors piRNAs couple with PIWI proteins, a subclass of the AGO family, into piRISCs able to disrupt complementary transposons by the slicer activity of PIWI [reviewed in (Ishizu, Siomi, & Siomi, 2012; Malone & Hannon, 2009)]. Interestingly, piRNAs can also repress transposon expression by forming heterochromatic structures at the transposon loci (Le Thomas et al., 2013).

1.3 miRNA overview

1.3.1 miRNA biogenesis pathway

(The pathway here outlined describes the biogenesis of animal miRNAs. On other organisms some steps may differ or be performed by different enzymes.)

miRNAs are considered small non coding RNAs but, with few exceptions, they are derived from much longer precursors, transcribed by RNA polymerase II, that may contain more than one miRNA located on adjacent loci (Y. Lee et al., 2004; Rodriguez & Griffiths-Jones, 2004). These transcripts bear sequences of some complementarity that allow them to fold into hairpins (o duplexes) with an approximately 30 nt stem that can have mismatches and non canonical unions G:U. The base of the stem serves as the binding site for DGCR8 (DiGeorge syndrome critical region gene 8 protein, known as Pasha on *Drosophila melanogaster* and *Caenorhabditis elegans*) which recruits RNase III Drosha to cleave the primary transcript into individual hairpins of ~70 nucleotides in length (Han et al., 2006). These processed hairpins are known as precursor miRNA or pre-miRNAs (Figure 2. The miRNA hairpin structure.).

Some miRNAs, known as mirtrons, coincide exactly with intron signalling sequences and so, instead of being processed as above, they use the splicing machinery to produce pre-miRNAs (Houwing

et al., 2007; Okamura, Hagen, Duan, Tyler, & Lai, 2007; J. G. Ruby, Jan, & Bartel, 2007). Both canonical and mirtron pre-miRNAs are exported through the nuclear pore by Exportin-5. On the cytoplasm they are cleaved by RNA endonuclease type III Dicer into double stranded duplexes of 21-22 nt (Du & Zamore, 2005). One of the strands, known as mature miRNAs -3p and -5p, is preferentially loaded into RNA Induced Silencing Complexes (RISC) composed of enzymes from Argonaute family (AGO1 to AGO4) together with other complementary proteins. Mature miRNAs then guides the RISC to target transcripts that are partially complementary to them to finally execute their repressive function (Krol, Loedige, & Filipowicz, 2010. Figure 3. The miRNA biogenesis pathway).

1.3.2 Recognition of miRNA target genes

miRNA-directed gene repression is driven by sequence complementarity between the mature miRNA and the messenger RNA (mRNA). In mammals this base pairing is not perfect but comes to target recognition. First, most of the sequences nevertheless several general rules have been observed when it complementary to miRNAs, target sites, are found on the 3'UTR of the messenger. Secondly, as the base pairing is not perfect it allows the formation of bulks that apparently have influence on whether the RISC acts on a particular mRNA via RNA cleavage or via translation repression (See section section 1.3.3 Mechanisms of miRNA-mediated gene repression). The third rule is the need of a perfect and contiguous pairing of nucleotides 2 to 8 from the 5' end of the mature miRNA. This stretch of sequence is called the seed region of the miRNA and is thought to be the most important region for target gene recognition . Recognition is improved on target sites (nucleotides opposite to the seed region on the messenger) surrounded by an A residue at 5' and a A or U at 3'. Another rule is to have a moderately good complementarity to the miRNA 3′ side to stabilize the interaction specially if the complementarity to the seed region is not perfect (Figure 4. Target recognition rules). Other factors as the accessibility of the target site or the presence of multiple sites for the same miRNA on a given messenger can also affect the repression. All these rules, together with the stability of the secondary structure formed by the duplex miRNA-mRNA and

the conservation of the target site sequences are the basis for the algorithms that predict miRNA target sites.

1.3.3 Mechanisms of miRNA-mediated gene repression

The molecular details of miRNA function are not well understood yet. Numerous studies on different systems (on different organisms and both *in vivo* and *in vitro*) have made it possible to discern several pathways and mechanisms; however the conclusions obtained by these studies are sometimes controversial and difficult to interpret and generalize. Some of these apparent contradictions may derive from the inherent differences of the experimental approaches applied but it is also possible that miRNA-mediated regulation indeed functions through diverse and not so straightforward pathways. Despite controversies it is accepted that miRNAs control gene expression posttranscriptionally mainly by affecting the translation or compromising the stability of mRNAs (Figure 5. Mechanisms of miRNA function).

Several experiments provide evidence that miRNA repression can act at the initiation step of translation. miRISCs affect this step by competing with elongation factor eIF4E for m7Gcap binding (Kiriakidou et al., 2007) and they can also prevent the assembly of the mRNA to the 40S ribosome subunit (Mathonnet et al., 2007; Thermann & Hentze, 2007). It has been suggested that miRISC interaction with eIF6 may prevent the recruiting of the 60S ribosomal subunit, although the mechanisms of this kind of repression are not yet elucidated (Chendrimada et al., 2007).

Although there is evidence that supports miRNA repression at initiation stage, some mRNAs with target sites can still associate with active polysomes in the presence of miRNAs (Maroney, Yu, Fisher, & Nilsen, 2006; Petersen, Bordeleau, Pelletier, & Sharp, 2006). These observations are incompatible with the initiation repressor model and suggest that it may not be the only way through which miRNAs affect translation. Petersen and colleagues proposed that at post-initiation stages miRNAs may induce a premature termination of translation, especially in mRNA containing IRES (Petersen et al., 2006). Also in some cases miRNAs may not

terminate but rather decelerate the translation elongation (Maroney et al., 2006). These mechanisms have not been proven to be mutually exclusive and thus one can mask the effects of the other making it difficult to disentangle the precise mode of action of miRNAs at repression of translation.

Early observations stated that miRNAs regulated gene expression through translation obstruction without affecting the levels of the mRNAs targeted. However later studies on transcriptomes have reported that miRNA repression is associated in many cases to RNA destabilization. An initial evidence of the miRISC role on RNA degradation comes from its localization on p-bodies, cellular structures where mRNA degradation occurs (Eulalio, Behm-Ansmant, & Izaurralde, 2007; R. Parker & Sheth, 2007). Also, Argonaute1 has been proven to interact with proteins of known implication on mRNA degradation such as GW182 (Behm-Ansmant et al., 2006; Till et al., 2007). Furthermore, miRNA repression is prevented by the depletion of the CCR4-NOT deadenilatyion complex and also by the knockout of proteins DCP1 and DCP2 from the decapping complex (Behm-Ansmant et al., 2006; Eulalio, Rehwinkel, et al., 2007) suggesting that miRNA effect on mRNA levels is dependent on decapping and deadenylation.

The determinants of whether the repression of certain mRNA targets goes through destabilization at one or another stage of translation are not known although accessory proteins interacting with mRNA or RISC and the degree of complementarity between mRNA and miRNA sequences are thought to play a role. Studies on *Drosophila melanogaster* have identified target mRNAs that undergo strong translational repression while their transcriptome levels remain unaffected, while other targets seem to be repressed entirely by degradation or by a combination of both pathways (Behm-Ansmant et al., 2006; Eulalio, Rehwinkel, et al., 2007). Moreover when comparing the same couple of miRNA-mRNA, repression may be observed in a tissue but not in another, as shown in studies on zebrafish embryos where some targets of miR-430 were repressed in somatic cells but not in cells from the germline (Mishima et al., 2006). Similar results have been found comparing mRNA reporters targeted by let-7 in different mammalian cells (Schmitter et al., 2006).

Finally, the finding that mature miRNAs can be relocated to the nucleus or be secreted out of the cell suggests that miRNAs may function through other still unknown mechanisms (Hwang, Wentzel, & Mendell, 2007; Valadi et al., 2007).

1.3.4 Relevance of miRNA mediated regulation

There is evidence that a large fraction of protein-coding genes are targets for miRNA-mediated regulation (John et al., 2004; Lewis, Burge, & Bartel, 2005). A single miRNA can regulate numerous target genes and affect this way the protein production of the whole cell (Baek et al., 2008; Selbach et al., 2008). Reciprocally, many mRNAs contain target sites for diverse miRNAs (Schnall-Levin et al., 2011) and so different miRNAs may have a synergic effect on one single target gene. The importance of miRNAs in regulating protein-coding genes is well established nowadays; interestingly, in addition to protein-coding genes, miRNAs can also target other kinds of RNAs as shown by Hansen and colleagues (Hansen et al., 2011). The interactions between miRNAs and other classes of ncRNAs have started to emerge recently but suggest intricate mechanisms of co-regulation. Because of their large number of targets, miRNAs occupy important positions in regulatory networks where they tune multiple processes and function as bridges between different cellular pathways (Ooi et al., 2011; Plaisier, Pan, & Baliga, 2012). Studies on regulatory networks revealed that miRNAs establish recurrent motifs in concert with transcription factors, another key element of gene regulation. Apparently these motifs, such as feedback and feed-forward loops, act to enhance the robustness of gene regulation (Tsang, Zhu, & van Oudenaarden, 2007; Yu, Lin, Zack, Mendell, & Qian, 2008).

The wide networks of miRNA regulation reach almost every physiological and developmental process studied, examples of such are animal development and differentiation (Christodoulou et al., 2010; Plasterk, 2006), pluripotency (Leonardo, Schultheisz, Loring, & Laurent, 2012), epithelial to mesenchimal transition and metastasis (Bracken, Gregory, Khew-Goodall, & Goodall, 2009), testis differentiation (Rakoczy et al., 2013), and neural plasticity and memoir (Bredy, Lin, Wei, Baker-Andresen, & Mattick, 2011) to cite a few.

On the other hand, miRNA deregulation has been associated to various complex human diseases including cardiovascular disease, several kinds of cancers (Croce, 2009) and mental disorders (Muiños-Gimeno et al., 2010). In more general terms there is evidence to suggest a role of miRNAs in orchestrating the signalling pathways of cellular stress response whose deregulation often leads to disease (Mendell & Olson, 2012). Apart from complex diseases a few cases of syndromic developmental disorders can also be traced back to mutations or deletions of particular miRNAs (de Pontual et al., 2011; Mencía et al., 2009). The first to be described was an autosomal dominant progressive hearing loss associated to a mutation in the seed region of miR-96 (Mencía et al., 2009). Later studies on mice concluded that this miRNA is needed for postnatal maturation of hair cells the failure of which leads to their degradation (Kuhn et al., 2011). There are also several cases of miRNA SNP variants (see Section 1.4.4 miRNA variability in primates and humans) associated with complex disorders, mostly with cancer. . A well-studied example is a SNP (rs11614913) on miR-196-a2 that has been associated with several kinds of cancer such as lung, breast, liver and gastric cancer (Hoffman et al., 2009; X.-D. Li, Li, Song, & Liu, 2010; Peng et al., 2010; Tian et al., 2009).

A mutation on a miRNA can potentially affect its processing and incorporation to RISC or change its set of targets [see Section 1.4.2 Diversification of miRNA genes], but reciprocally a change on a mRNA can create or destroy a miRNA recognition site (Figure 9. Outcomes of miRNA variation). Although the loss or gain of a target site on a mRNA is likely to produce a minor effect compared with a mutation on a miRNA, several functional changes of this kind have been reported. One of the most celebrated is a variant observed in sheep where a point mutation on the 3' UTR of the Myostatin gene creates a target site for miR-1 and miR-206. The output of this new regulation was the increased muscularity that characterizes the sheep strain in which the mutation was found (Clop et al., 2006).

In addition to changes on the target genes and the miRNAs themselves, variations on the proteins in charge of processing miRNAs or in their mechanisms of action are also likely to conduce to disease. This is the case of DiGeorge syndrome. Individuals

affected with this sydnrome show, among other symptoms, cognitive and behavioural impairments. DiGeorge syndrome is linked to a deletion in 22q11.2. In humans this loci contains several genes including DGCR8, a protein that interacts with Drosha at the initial steps of miRNA biogenesis. Fenelon and colleagues observed an abnormal brain function in heterozygous mutant mice Dgcr8 (+/-) (Fénelon et al., 2011). Their results place DGCR8 as a candidate to partially explain the molecular mechanisms beneath DiGeorge syndrome and highlight the importance of the proteins of miRNA biogenesis in brain function.

Knockout studies also stress the importance of miRNA processing factors. The individual knockout of Dicer, DGCR8, Drosha and Ago2 have all proven deleterious in mice, who die during early gestation (Park, Choi, & McManus, 2010). To better understand the role of miRNA regulation CRE-inducible knockouts of these factors have been developed and made it possible to refine the specific tissues and development times in which miRNAs are acting. Later knockouts further detailed specific phenotypes derived from the lack of individual miRNAs.

1.4 Evolution of miRNA repertories

1.4.1 Birth of novel miRNA genes

There are few requirements for a miRNA gene to be functional. Namely, a transcription origin and a sequence capable of forming a simple secondary structure that once transcribed can be processed by Drosha and Dicer enzymes and mounted into a RISC. Currently it is accepted that mammalian genomes are pervasively transcribed (Birney et al., 2007) and numerous structures with miRNA like properties are predicted to occur at least in the human genome (Bentwich et al., 2005). This means that there is plenty of material from which a miRNA can arise *de novo* (Liu et al., 2008; Lu, Fu, et al., 2008; Lu, Shen, et al., 2008). Nevertheless, the structure of the hairpin must fit precisely into Drosha and Dicer and this is why it is unlikely that the *de novo* appearance of a novel miRNA occurs in one single event and it must rather pass through transitional forms

until reaching the correct structure (Berezikov et al., 2011; Liu et al., 2008). Several sources for new miRNAs have been described (Figure 6. Genomic sources of novel miRNA genes).

Many miRNAs reside within introns of protein-coding genes (Rodriguez & Griffiths-Jones, 2004). Introns are an excellent location for new miRNAs to arise because they already provide them with the functional transcriptional unit of the hosting gene. Consistently with this, younger, species-specific miRNAs are more often located in introns than ancient miRNAs are [(Campo-Paysaa, Sémon, Cameron, Peterson, & Schubert, 2011), see Section 3.2 Differences in evolutionary rates among miRNAs in the human and chimpanzee genomes]. Interestingly many intronic miRNAs are transcribed from their own promoters instead of using the promoter of the hosting gene (Isik, Korswagen, & Berezikov, 2010; Martinez et al., 2008; Ozsolak et al., 2008). Various other transcripts, such as tRNAs, snoRNAs or pseudogenes can also evolve miRNA-like structures within their sequences (Burroughs et al., 2011).

Another mechanism through which miRNAs are born is related to gene duplication events. Groups of miRNAs with a high sequence similarity or even identical sequence are frequently found on genomes. These paralogous copies arise through gene duplication events, which are considered a major source of novel miRNAs. These events can place the outcome copies in tandem, in which case they can be transcribed as a single unit frequently refered as clustered miRNAs (Figure 7. Localization of miRNA genes), or on distant locations with independent transcription origins (Hertel et al., 2006). Later one mutation can lead to a neofunctionalization of the duplicated miRNAs thus creating novel miRNAs (J. Ruby, Stark, & Johnston, 2007). When the seed region of these paralogous miRNAs is conserved they are considered a miRNA family (Ambros et al., 2004).

Transposons can also lead to the apparition of novel miRNAs. Extensive sets of miRNAs derived from transposons and retrotransposons have been identified on the genomes of human, rhesus and mouse (Yuan, Sun, Liu, & Xie, 2011). Transposon-derived miRNAs tend to be lineage-specific and younger when compared with miRNAs with other genomic origins (Piriyapongsa, Mariño-Ramírez, & Jordan, 2007; Yuan et al., 2011).

Another way a novel miRNA can emerge is by antisense transcription of an existing miRNA gene. Given that the hallmark hairpin structure of a miRNA precursor is made up of an imperfect palindromic sequence, its antisense transcription can lead to the formation of a different but still processable new hairpin. The bidirectional transcription of a miRNA gene sequence has been observed in many cases (Berezikov et al., 2011; J. Ruby et al., 2007).

1.4.2 Diversification of miRNA genes

Once a miRNA is established in a genome, its sequence can be altered in such a way that it changes its mode of action converting it into a different miRNA. The function of non-coding RNAs depends directly on their sequence as it determines their secondary structure and defines the complementary partners it will associate to, may they be proteins, DNA or other RNA molecules.

In the case of miRNAs, mutations in their sequence may result in alterations of their processing, compromise their loading into the RISC or affect the set of targeted genes and so mutations on the sequence are of central importance when considering miRNA evolution (Figure 8. Diversification of miRNA genes).

Apart from direct mutations on the sequence, other mechanisms of miRNA diversification have been discerned. Given its structure and biogenesis process, a single miRNA precursor can produce two mature miRNA strands (See section 1.3.1 miRNA biogenesis pathway and Figure 2. The miRNA hairpin structure). Early low coverage cloning experiments often failed in capturing both strands (5' and 3') and so one of the strands (referred as miRNA*, complementary o passenger strand) was regarded as inert and easily degradable. More recent deep sequencing experiments have shown that although one strand is indeed more abundant in most cases, the other strand can also reach physiological levels of expression (J. Ruby et al., 2007) and may have some biological relevance (Okamura et al., 2008). Furthermore, the usage of one mature or the other can change from tissue to tissue in an organism, (as it has been observed in flies (Marco, Hui, Ronshaugen, & Griffiths-Jones,

2010; Okamura et al., 2008) and worms (de Wit, Linsen, Cuppen, & Berezikov, 2009) or in distinct stages of development (Chiang et al., 2010) and it can also be different between orthologous genes on different species (Marco et al., 2010). The mechanisms beneath mature switching are not fully understood (it is suspected that thermodynamical stability and accessory proteins may have a role in it) but the passenger strand is surely a valid substrate for miRNA diversification.

Another mechanism that produces different mature miRNAs from a single precursor is known as seed shifting. It is the consequence of the variation on the location of Drosha or Dicer cleavages on the precursor (Berezikov, 2011; J. Ruby et al., 2007). Mature variants produced this way are known as isomiRs (Morin et al., 2008), they have the same sequence except for their terminal nucleotides. The addition or subtraction of even 1 nucleotide at the 5' termini of a mature miRNA can lead to a shift of the seed (defined as the nucleotides 2 to 7 from the 5' of the mature miRNA) and so produce a change in the set of target genes. Accordingly 3' isoMIRs that do not affect the position of the seed are much more frequently found than 5' isomirs on sRNA deep sequence data (Berezikov et al., 2011; Chiang et al., 2010).

Generally, for a particular miRNA one of the isomiRs is dominant above the others but this dominance can change during evolution. This change of dominance is one of the ways in which duplicated miRNAs can achieve divergence. Examples of seed shifting have been described in *Drosophila melanogaster* (J. Ruby et al., 2007). Orthologous miRNAs on different species can also show a different isomiR preference as it is the case of miR-100, a widely spread metazoan miRNA. Notably, this miRNA has a predominant form in all bilaterians but another one on the anemone *Nematosella vectensis* (Grimson et al., 2008).

Another mechanism of miRNA change is hairpin shifting which is the appearance of a new hairpin composed by one arm from an original miRNA hairpin and the other from the adjacent upstream or downstream sequence. This way a precursor miRNA may conserve one existing mature miRNA while the other is completely new. Hairpin shifting has been observed on roughly half of the miRNA families present in nematodes (de Wit et al., 2009).

Finally another way in which a miRNA sequence can be modified is through RNA editing, the site-specific modification of an RNA sequence that yields a product different from that encoded in the DNA. Most RNA editing in human cells is the conversion from adenosine to inosine (A-to-I), also expressed as an A-to-G change, since inosine is recognized as guanosine by the cell machinery. A-to-G editing is catalyzed by Adenosine Deaminases (ADARs) that act on double stranded RNAs, such as miRNA precursors (Blow et al., 2006; Nishikura, 2010). Although miRNA editing is not very frequent and has proven difficult to distinguish from artefacts of sequencing errors, some cases of have been convincingly reported especially on neuronal tissue (Alon et al., 2012; Kawahara et al., 2008). miRNA editing can interrupt the biogenesis from the primary forms to the mature ones. For example, editing in site +3 of the primary miR-151 blocks the cleavage by Dicer, resulting in the accumulation of edited precursor and preventing the formation of the mature form (Kawahara, Zinshteyn, Chendrimada, Shiekhattar, & Nishikura, 2007). Editing can also modify the spectrum of target genes that are recognized by the miRNA as has been shown for an edition on the seed region of miR-376a-5p, which significantly changed the set of target genes repressed by this miRNA (Kawahara, Zinshteyn, Sethupathy, et al., 2007).

1.4.3 Evolution of miRNA repertories

To date no miRNA has been found on any prokaryotic species, miRNAs seem to be an eukaryotic innovation. Within Eukarya they are absent in fungi, but numerous in plants and especially in metazoans (Cerutti & Casas-Mollano, 2006; Hertel et al., 2006). These two groups show diverse families of miRNA genes, which seem to have evolved independently as none of them is shared between plant and animals. Plant miRNAs tend to have longer hairpins and the complementarity between them and their target genes is nearly perfect and expands beyond the seed region in contrast with animal miRNAs in which wobbles and mismatches are permitted and even required (See Section 1.3.2 Recognition of miRNA target genes). Interestingly, despite of their independent origin two of the oldest conserved miRNA families in these groups (miR166-miR165 in plants and let-7 in animals) are both involved

in crucial symmetry-establishing and differentiation events during development (Shabalina & Koonin, 2008).

Few miRNAs have been identified in the phylum Poriphera and those present are unique to this group and not overlapping any other miRNA reported (Robinson et al., 2013; Wheeler et al., 2009). Furthermore, sponge miRNAs are structurally different from eumetazoan miRNAs and some of their characteristics resembles more to plant miRNAs. For example the situation of the mature sequence, which in eumetazoans tends to be not farther than 10 nt from the loop, in the Poriphera miRNAs is located at least 10 nucleotides and often 30 nucleotides apart from the loop [Wheeler 2009]. This supports the idea of an independent origin of a whole new miRNA repertorie not only across Eukarya but also within Metazoa. On an early study by Sempere and colleagues the miRNA complement of three species of cnidarias was characterized and only two miRNAs found (mir10 and mir100) were shared with other animals. None of the deutero-specific or protostome-specific miRNAs (as established in the mentioned study) was found in Cnidarians.

The animals in Nephrozoa, the clade encompassing Deuterostomes and Protostomes, are characterized by the presence of organs. In contrast with simpler animal forms as the mentioned cnidarians or sponges, Nephrozoans posses numerous miRNAs genes. In general an steady increase in miRNA number is observed as the morphological complexity of animals increases but there are also some outstanding expansions in miRNA acquisition rate along evolution time most notoriously: at the split of bilaterians (Sempere, Cole, McPeek, & Peterson, 2006), at the time of emergence of vertebrates (Heimberg, Sempere, Moy, Donoghue, & Peterson, 2008) and at the rise of eutherians (Hertel et al., 2006). In a more recent study, Guerra et al. corroborated these expansions and further detailed others, most notably in rodents and primates, with a significant miRNA gain in great apes (Guerra-Assunção & Enright, 2012). They also reported some less numerous gains in insects and nematodes and stated that significant losses of miRNAs occur only in few clades, such as frogs, marsupials, squirrel and hedgehog.

Interestingly, some of these expansions coincide with dramatic changes in body plans or morphological complexity, which has led

to hypothesize that miRNAs contribute substantially to animal evolution. This idea is reinforced by the prominent role of miRNAs in development that has been extensively documented during the last years and also with the fact that newly emerged miRNAs are often expresse in novel organs such as the vertebrate's blood vessels and somites (Giraldez et al., 2005; Kloosterman, Wienholds, de Bruijn, Kauppinen, & Plasterk, 2006; Wienholds & Plasterk, 2005; Wienholds et al., 2005; Zhao, Samal, & Srivastava, 2005).

Comparative studies carried out in metazoans have found that miRNAs differ from other genetic entities in three characteristics. First, they present remarkable sequence conservation. When *Drosophila melanogaster* miRNAs were assessed they showed sequence conservation levels even higher than elongation factor 1a and histone H3 genes, which are often used to elucidate the deepest branches on the tree of life. Similar results where found when comparing human miRNAs with those of other eutherians sharing a common ancestor approximately 100 Mya. Again, the level of nucleotide diversity of miRNAs was extremely low and even comparable to the rate of change of DNA sequences coding for 18S rRNA (Sempere et al., 2006). Although these results are based on the limited amount of miRNAs known at the time, the remarkable sequence conservation of established miRNAs has been confirmed by subsequent studies. Secondly, miRNA genes are continuously added to metazoan genomes which results in lineage-specific sets that when reconstructed correlate with established phylogenetic relationships. And lastly, once miRNA genes are established on a genome they are rarely lost. All together these evolutionary trends have allowed to reconstruct the origin of various miRNA gene families in great detail.

Currently miRBase (Release 21, june 2014), the main repository for miRNAs (Kozomara & Griffiths-Jones, 2011), houses data from 223 species and new miRNAs are continually discovered and incorporated, specially since the popularization of deep RNA sequencing. Human is by far the most deeply explored species with 1881 miRNA precursors and 2588 matures described to date. Human miRNAs outnumber those in closely related species (*Pan troglodytes*, 655 precursors; *Gorilla gorilla*, 352 precursors; *Pongo pygmaeus*, 642 precursors; *Macaca mulatta*, 619 precursors) and even model organisms (*Mus musculus*, 1193 precursors;

Caenorhabditis elegans, 250 precursors; *Drosophila melanogaster*, 256 precursors; *Arabidopsis thaliana*, 325 precursors). Some species of economic relevance have also been surveyed for the presence of miRNAs. This is the case of cow (*Bos taurus*, 808 precursors), soy (*Glycine max*, 573 precursors), rice (*Oryza sativa*, 592 precursors) and the silkworm (*Bombyx mori*, 487 precursors).

1.4.4 miRNA variability in primates and humans

In an early comparative study Bereznikov et al. sequenced 122 miRNAs in 10 different primate species and found a strong conservation in the stem of the miRNA hairpins, this constraint was more relaxed in the loop of the hairpin and markedly higher that the conservation of the sequence immediately flanking the miRNA (Berezikov et al., 2005).

The strong inter-species conservation in miRNAs observed by Bereznikov among primates is also observed intra-species in human. miRNAs in the human genome show less variation than their flanking regions as shown on a screening of SNPs by Saunders et al. (Saunders, Liang, & Li, 2007). The caveats of work based on SNP are ascertainment bias and lack of full coverage of sequence variation. These problems are overcomed in a study by Quach et al in which the investigators resequenced 92 genomic regions containing 117 miRNAs in different populations (Quach et al., 2009). This approach allowed them to perform statistical tests to undercover the action of natural selection over miRNAs in humans. On their study they found that miRNA variation is globally constrained. They also report that the variation is higher on the loop region and lower on one of the arms of the hairpins (allegedly the dominant mature form) where no mutations were found. Interestingly the complementary regions showed very few mutations. This conservation may ensure the stability of the hairpin but could also represent that this region is acting as a genuine mature miRNA itself (Okamura et al., 2008; Quach et al., 2009).

The signature of purifying selection is patent on most miRNAs indicating that most mutations are selected against because of their deleterious consequences, but some mutations on miRNAs could be advantageous and reach high frequencies by the action of positive

selection, or be neutral and rise in frequency by genetic drift. Interestingly, Quach's study showed evidences of positive selection in European and East Asian populations for a small-RNA-rich island on chromosome 14 enriched in miRNAs and snoRNAs and with no linkage disequilibrium with any protein-coding gene. Nevertheless the signal was not located in the miRNA region but rather in the region corresponding to the snorRNAs.

Likewise the miRNA recognition sites on the 3'UTR also show a strong constraint in humans (Chen & Rajewsky, 2006). Together these data indicates that a strong complementarity between the miRNA and its target site on the mRNA is necessary to establish a proper regulation.

Figure 1. Diverse classes of RNAs. tRNA, transfer ribonucleic acid; rRNA, ribosomal ribonucleic acid; snRNA,small nuclear ribonucleic acid; lincRNA, long intergenic non-coding ribonucleic acid; RNAi,Interference RNAs; piRNA, piwi-interecting ribonucleic acid; siRNA, short-interfering ribonucleic acid and miRNA, micro ribonucleic acid.
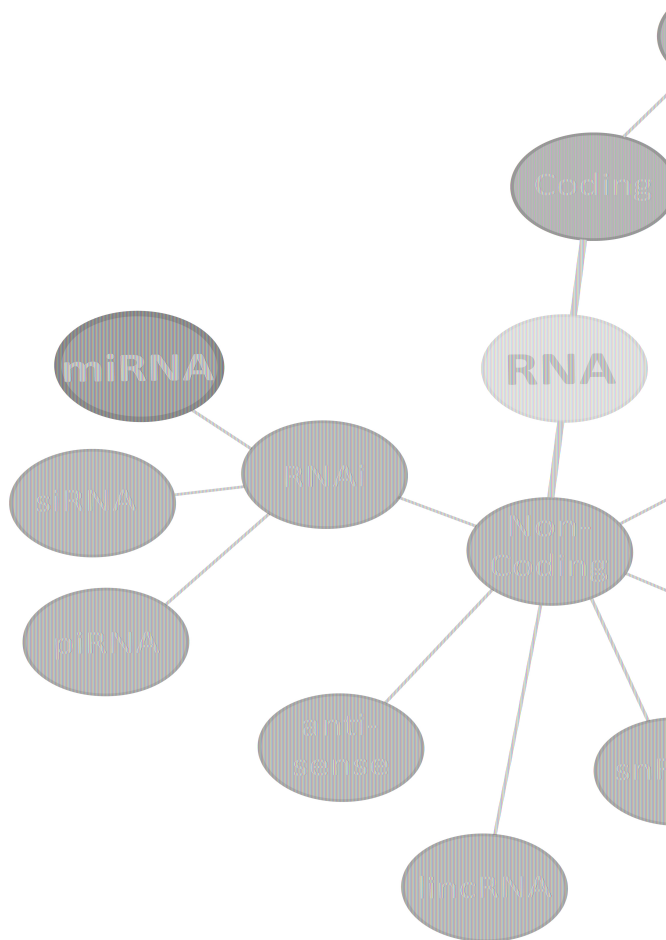
Figure 2. The miRNA hairpin structure. The miRNA precursor (also refered as pre-miRNA or hairpin) is approximately 85 nucleotides. It is composed of an imperfect duplex between the mature miRNA (miR) and its complementary sequence (miR*), an adjacent stem, and a terminal loop. Adapted from Quach 2009.
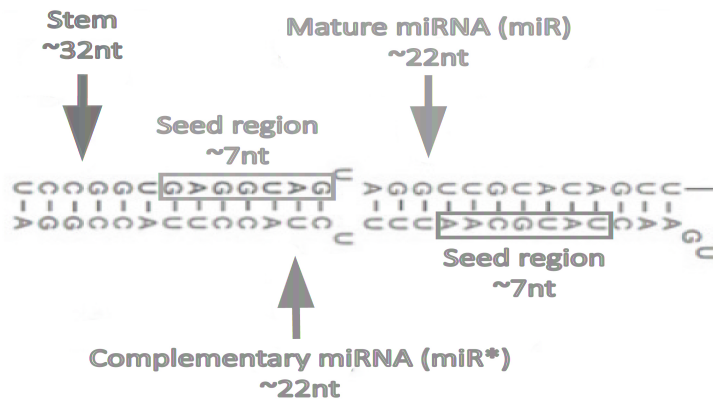
Figure 3. The miRNA biogenesis pathway. Canonical pathway of animal miRNA processing. Adapted from He & Hannon, 2004.
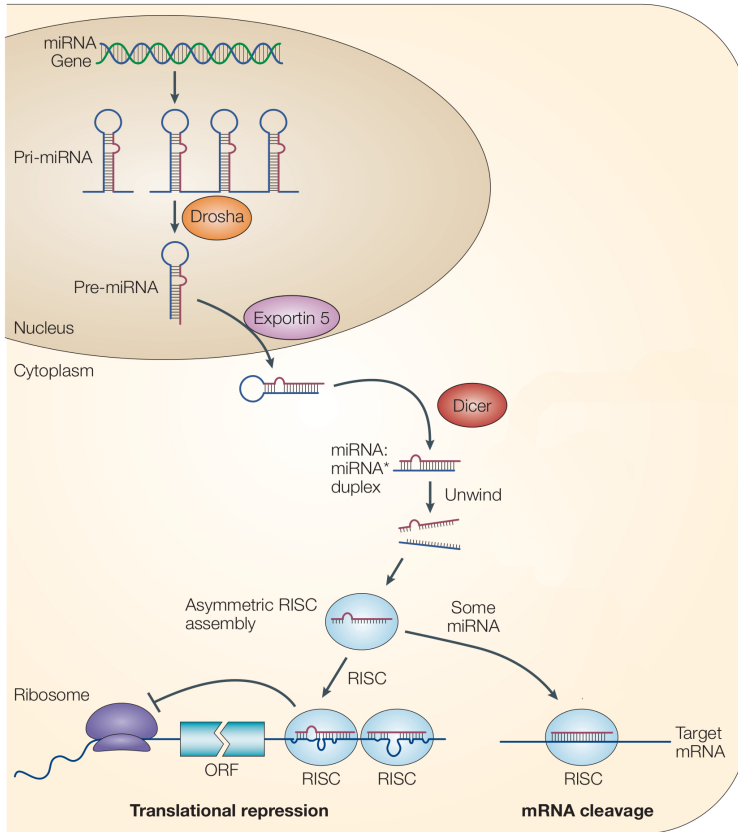
Figure 4. Target recognition rules. MicroRNAs (miRNAs) interact with their mRNA targets by base pairing. Metazoan miRNAs base pair with their targets imperfectly, following a set of rules. One rule for miRNA–target base pairing is perfect base pairing of miRNA 'seed' region. An A residue across position 1 of the miRNA, and an A or U across position 9, improve the site efficiency. Bulges or mismatches must be present in the central region of the miRNA–mRNA duplex. There must be reasonable complementarity to the miRNA 3′ half to stabilize the interaction particularly when matching in the seed region is suboptimal. From Berezikov 2011.
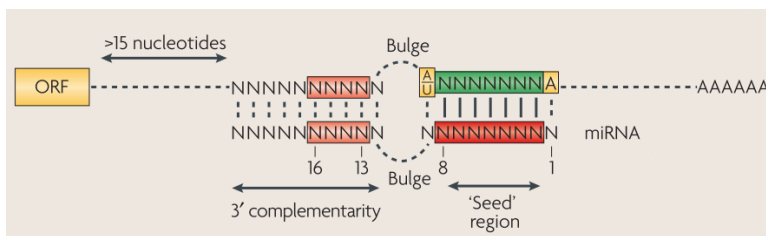
Figure 5. Mechanisms of miRNA function. miRNA mediated repression can occur through different mechanisms: miRNAs may induce deadenylation of target mRNAs and their relocalization to P-bodies; they may repress translation at initiation and/or elongation stages and also proteolytic cleavage has been proposed as a mean of miRNA repression although it is less sustained than the other mechanisms. From (Filipowicz, Bhattacharyya, & Sonenberg, 2008)
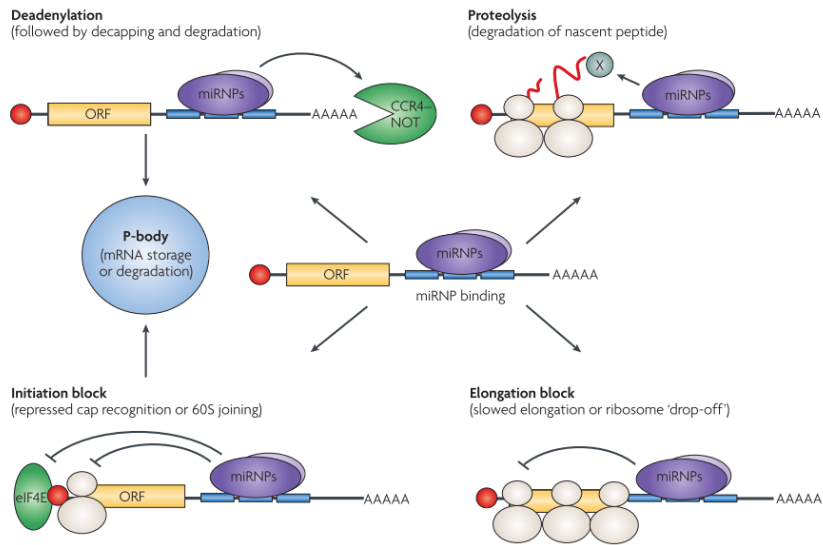
Figure 6. Genomic sources of novel miRNA genes. (a) Novel miRNA genes can emerge by local or non-local duplication of existing miRNA genes. (b) Introns are a frequent source of unstructured transcripts that can gradually evolve into novel intronic miRNA through a stage of miRNA-like hairpin. (c) In de novo emergence of miRNAs, a transcriptional unit provides a source of initially unstructured transcript that evolves into a novel miRNA gene. (d) Transposable elements can provide novel transcriptional units for the evolution of miRNA-like hairpins into novel miRNA genes. (e) Structured transcripts, such as tRNA and small nucleolar RNA (snoRNA), can evolve into novel miRNAs through the miRNA-like hairpin stage. (f) Antisense transcription of existing miRNA loci can lead to the formation of miRNA hairpins with novel mature miRNA and miRNA*. From Berezikov 2011.
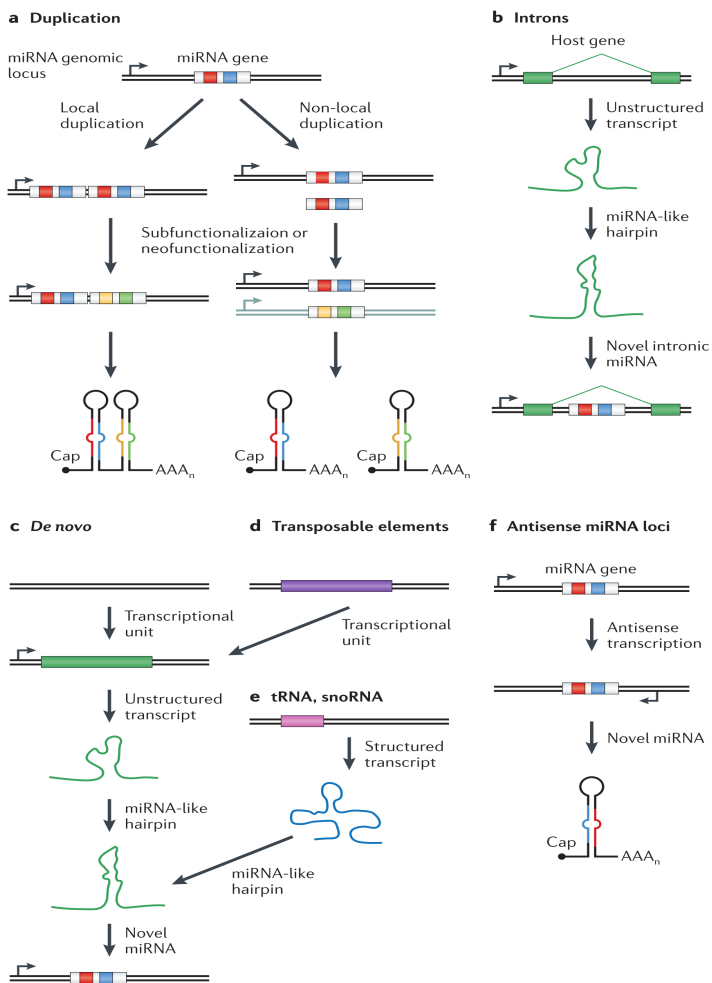
Figure 7. Localization of miRNA genes. miRNAs are encoded in genomes either (a) as independent transcriptional units with their own promoters (solo miRNAs), or (b) as clusters of several miRNA genes transcribed as a single pri-miRNA. (c) A substantial fraction of animal miRNA genes are located in introns of protein-coding genes. Adapted from Berezikov 2011.
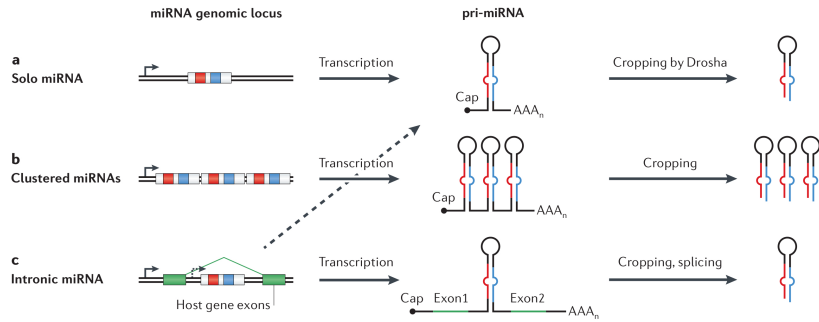
Figure 8. Diversification of miRNA genes. (a) Mutations that affect the mature miRNA sequence can directly lead to changes in the target repertoire: specifically, if they give rise to a new seed sequence (yellow). (b) In seed shifting, the dominant mature miRNA form is shifted owing to changes in Drosha and Dicer processing. As a consequence of the changed 5′ end of the mature miRNA, a new seed sequence (yellow) comes into play. (c) In arm switching, changes somewhere in the miRNA precursor lead to a shift in the ratio of miRNA to miRNA* (d) In hairpin shifting, a novel hairpin is evolved from an adjacent region, keeping the same mature miRNA (red) but introducing a novel miRNA* sequence (yellow), which then could acquire regulatory function through loading into miRISC. (e) RNA editing results in post-transcriptional conversion of adenosines to inosines (A to I) in the double-stranded parts of an miRNA precursor. This can affect processing of the precursor and can introduce changes into mature miRNA sequence; in particular, editing of the seed region (yellow) can result in a changed target repertoire. Adapted from Berezikov 2011.
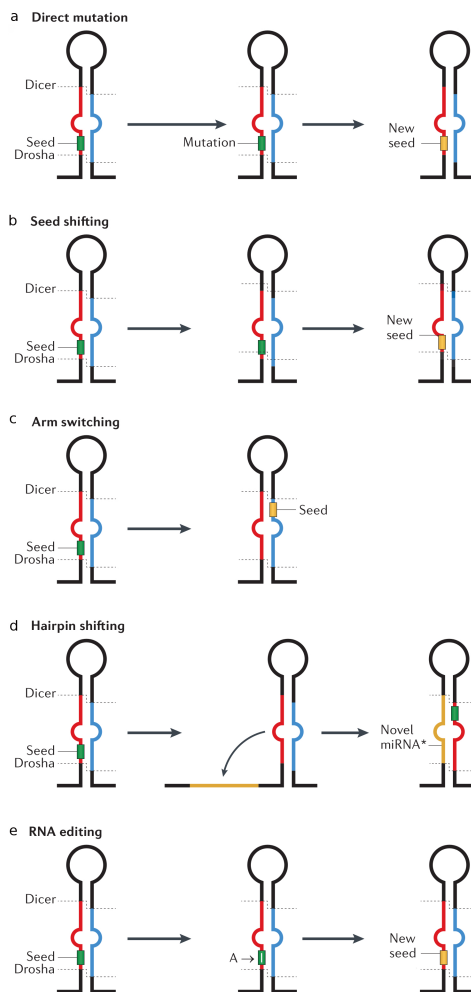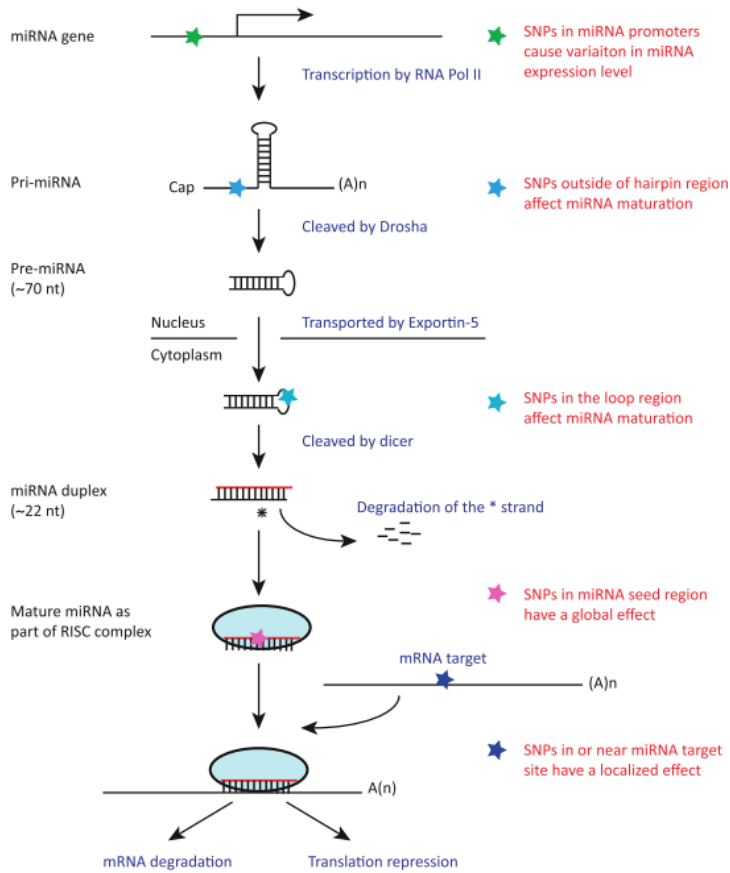
Figure 9. Outcomes of miRNA variation. The stars in the figure represent single nucleotide polymorphisms (SNPs) within miRNA and mRNA sequences. For each type of SNP, the effect on miRNA processing and miRNA-mediated regulation are annotated on the right side of the figure. From Jingjing Li 2012.

2. OBJECTIVES

General objective

This thesis was conceived as an evolutionary study with the objective to compare, at the genomic and individual genic level, miRNAs among present-day humans and their closest relatives: extant non-human primates, particularly chimpanzee, and extinct humans such as the Neanderthal.

Particular objectives

- To identify the differences at the sequence level between human miRNAs and their putative orthologous miRNAs in the Neanderthal and chimpanzee genomes.

- To survey the functional implications of a nucleotide substitution in the seed region of the miRNA miR-1304, the unique miRNA found to diverge between present-day humans and Neanderthals.

- To investigate the action of natural selection in human miRNAs through a comparative genomics approach using the human and chimpanzee genomes.

-To classify and evaluate the total number of human miRNAs into relevant categories regarding their genomic location, aggregation in clusters, duplication status and phylogenetic distribution.

- To assess the contribution of this categorization in explaining the variance of miRNA evolutionary rates between human and chimpanzee.

3. RESULTS

3.1 An Ancestral miR-1304 Allele Present in Neanderthals Regulates Genes Involved in Enamel Formation and Could Explain Dental Differences with Modern Humans.

Available on http://mbe.oxfordjournals.org/content/29/7/1797.long

Lopez-Valenzuela M, Ramírez O, Rosas A, García-Vargas S, de la Rasilla M, Lalueza-Fox C, Espinosa-Parrilla Y. An ancestral miR-1304 allele present in Neanderthals regulates genes involved in enamel formation and could explain dental differences with modern humans. Mol Biol Evol. 2012 Jul;29(7):1797-806. doi: 10.1093/molbev/mss023.

3.2 Differences in evolutionary rates among microRNAs in the human and chimpanzee genomes

Maria Lopez-Valenzuela, Gabriel Santpere, Natalia Petit-Marty, Arcadi Navarro, Yolanda Espinosa-Parrilla

4. DISCUSSION

The depart of this thesis was the observation of the great quantity and variety of human miRNAs that have been discovered along the last decade compared with the repertoire of miRNAs in other species of primates. In general terms we wanted to assess the significance of such a huge diversification and, furthermore, compare the human repertoire with that of closely related species. The idea was to discern if this finding was just an artefact derived from the large amount of resources dedicated to the study of humans compared with other species. This kind of reflection poses big questions such as: What is a real miRNA? Do all described miRNAs exist? How can we tell if a particular sequence is a miRNA? How can the function of a miRNA be assessed?

One approach that we took to begin to tackle these questions was trying to discern the action of natural selection in human miRNAs through a comparative genomic approach. We assessed the differences in substitution rates among miRNAs considered in broad categories that took into account their genomic localization, aggregation in clusters, duplication status and phylogenetic distribution. This was a whole genome analysis that departed from the total collection of human miRNAs reported on miRBase. It is worth to remark that along the work of this thesis miRBase, the repository of all known miRNAs (www.mirbase.org/), has passed from having 750 human miRNAs (Sanger release 14.0) to the current 1881 (June 2014; Sanger release 21.0). This exponential increase in the number of known miRNAs has been one of the challenges of this thesis: to keep our datasets updated according to the latest version of miRBase available while being in the middle of our analysis.

Many of these new miRNAs present very low concentrations and have been found in very particular tissues and biological scenarios, frequently pathological (Nygaard et al., 2009: teratomes and breast cancer; Stark et al., 2010: melanome; H. Wang et al., 2013: serum of sepsis patients). These circumstances raise a crucial question about the real existence of these miRNAs

This is a big concern among investigators studying miRNAs (M. Brown, Suryawanshi, Hafner, Farazi, & Tuschl, 2013; Meng, Shao,

Wang, & Chen, 2012), to the point that recently miRBase incorporated a tag for "High confidence miRNA" on their records. A miRNA is considered of high confidence if either (a) it has at least 10 reads mapping to each arm, or (b) it has at least 5 reads mapping to each arm and at least 100 reads mapping in total. This tag aims to prevent that a bona fide miRNA gene set gets swamp into a sea of dubious annotations, a situation that would be specially compromising for comparative studies as the ones presented on this thesis. In addition to the "High confidence status" a more light-hearted survey is available on each miRNA datasheet on miRBase, in which anybody can answer to the question "Do you believe this miRNA is real?"

There are some basic acceptance criteria for miRNA identification, namely: a minimum read count for mature miRNA, the prediction of a hairpin fold with overlap between mature and complementary miRNA and mapping to a limited number of locations in the genome. Some authors suggest that this criteria should extended to what they consider signatures of a functional miRNA. For example Liang and colleagues propose a miRNA candidate shall have a mature more conserved than the rest of the hairpin and stress the importance of the expression level (Liang & Li, 2009).

Our comparative analysis found two miRNAs that are clear examples of this problematic issue of the existence, miR-3691 and miR-4686. They have significant results on the Tajima's test of relative rates of evolution between human and chimpanzee, but have not yet been described in the chimpanzee genome. Without further expression analyses, we can not confirm if these are real miRNAs existing in chimpanzee or if they have a part on primate evolution.

Similar is the case of miR-1304, for which we performed a functional study in this thesis. This particular miRNA differs between the reference human and the first draft of Neandertal genome (Green et al., 2010) at one nucleotide position, precisely in its seed region. Although we found functional difference between the two variants of miR-1304, this is a relatively novel primate-specific miRNA that by the time of publication of the paper was

reported only in human tissues and at a very low level. As the target genes we found for mir-1304 where typical of ameloblast cells we wanted to test it expression on that particular tissue, but unfortunately we could not finally obtain these precious cells. mir-1304 expression in other species apart from human was later confirmed in an extensive sequencing study (Dannemann, Nickel, Lizano, Burbano, & Kelso, 2012) in which mir-1304 was found in rhesus macaque tissues.

The discovery of miRNAs on species other than human is important for comparative analysis. With this aim we set up an sRNA sequencing study with samples from different species (chimpanzee, gorilla, rhesus macaque and mouse) and tissues (brain, heart, muscle, liver). This study is still ongoing and we hope it will provide further evidence of the existence of miRNAs already reported in primates and maybe add some new miRNAs as well. This would help to reduce the gap between the number of miRNAs reported for human and for the rest of primates, although it is foreseeable that this gap will not disappear completely given that, as mentioned before, the majority of novel human miRNAs are discovered on clinical studies of conditions of medical interest that are not likely to be tested on other species.

Apart from the concern about the existence of reported miRNAs another important question regards about differences in the sequence of orthologous miRNAs: What are the evolutionary and functional implications of the single nucleotide substitutions in miRNAs among primates? To try to solve this question we compared human miRNAs with their putative orthologous on chimpanzee and orangutan. We find that many miRNAs are identical among these three species, and those that diverge show very few differences and are mainly novel miRNAs present only in primate species. Many of these primate-specific miRNAs are probably not yet established into regulatory networks and may represent transient phases that may or may not get fixated over time. As reported by Meunier, half of the newly born miRNAs families in mammals would be deleterious and lost by purifying selection (Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, 2013).

Also, we have explored some characteristics of miRNAs (clustering, localization, age) in relation with substitution rates, but other comparisons are possible and may shed light on the function of miRNA. Primate-specificity was the characteristic that better explained the differences in substitution rates in our data set, but it may be masking another characteristic, dependent or independent, such as expression level.

The results of this thesis point that even though the forces of miRNA-driven evolution in primates could partially rely in neo-functionalization of existing miRNAs due to mutations that change the spectrum of target genes, it is probably more due to the apparition of new miRNAs or to differences in miRNA expression due to mutations or RNA modifications.

5. CONCLUSIONS

- The functional study of a single nucleotide substitution in the primate-specific miR-1304 suggests that a functional difference among the ancestral and derived versions of this miR-1304 related to the regulation of a cluster of dental genes that could, at least partially, underlie phenotypic differences between the *Homo sapiens* and *Homo Neanderthalensis* species. The observed miR-1304-mediated regulation illustrates how a change in a regulatory molecule could have an impact on a very particular phenotypic trait and potentially contribute to differentiation of closely related species.

-The comparative study of substitution rates among individual miRNAs and in groups of miRNAs classified in different categories in human and chimpanzee showed a general strong constraint in miRNA genes. However, novel primate-specific miRNAs evolved faster than miRNAs present beyond the primate group, indicating that miRNA-driven evolution in primates could be in part sustained by mutation in these novel, primate-specific miRNAs.

-miRNAs are an heterogeneous class of genetic regulatory elements, for which evolutionary forces have acted differently. The evolutionary approach performed helped to identify miRNA attributes that have been specially accelerated along primate evolution.

-The miRNA field is a rapidly moving area of research. The discovery of new miRNAs increases at an astonishing pace. Some of the characteristics first attributed to the miRNAs (regarding their mode of action, localization, etc.) are now being revised as more and more miRNAs are discovered. Interestingly, with these revisions comes the realization that the general characteristics retrieved from early studies carried out with limited miRNA sets may not be applicable to the numerous new miRNAs recently discovered.

6. REFERENCES

Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. *Current Opinion in Microbiology*, *10*(2), 134–9. doi:10.1016/j.mib.2007.03.010

Alon, S., Mor, E., Vigneault, F., Church, G., Locatelli, F., Galeano, F., … Eisenberg, E. (2012). Systematic identification of edited microRNAs in the human brain. *Genome Research*. doi:10.1101/gr.131573.111

Ambros, V., Bartel, B., Bartel, D. P., Krichevsky, A. M., King, K. S., Donahue, C. P., … Griffiths-jones, S. A. M. (2004). A uniform system for microRNA annotation. *RNA (New York, N.Y.)*, *9*, 277–279. doi:10.1261/rna.2183803.One

Aravin, A. a, Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., … Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. *Developmental Cell*, *5*(2), 337–50.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, *455*(7209), 64–71. doi:10.1038/nature07242

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., … Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–12. doi:10.1126/science.1138140

Baulcombe, D. (2004). RNA silencing in plants. *Nature*, *431*(7006), 356–63. doi:10.1038/nature02874

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., & Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & Development*, *20*(14), 1885–98. doi:10.1101/gad.1424106

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., … Bentwich, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genetics*, *37*(7), 766–70. doi:10.1038/ng1590

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nature Reviews Genetics*, *12*(12), 846–860. doi:10.1038/nrg3079

Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R. H. a, & Cuppen, E. (2005). Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, *120*(1), 21–4. doi:10.1016/j.cell.2004.12.031

Berezikov, E., Robine, N., Samsonova, A., Westholm, J. O., Naqvi, A., Hung, J.-H., … Lai, E. C. (2011). Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. *Genome Research*, *21*(2), 203–15. doi:10.1101/gr.116657.110

Birney, E., Stamatoyannopoulos, J. a, Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., … de Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. doi:10.1038/nature05874

Blow, M. J., Grocock, R. J., van Dongen, S., Enright, A. J., Dicks, E., Futreal, P. A., … Stratton, M. R. (2006). RNA editing of human microRNAs. *Genome Biology*, *7*(4), R27. doi:10.1186/gb-2006-7-4-r27

Bracken, C. P., Gregory, P. a, Khew-Goodall, Y., & Goodall, G. J. (2009). The role of microRNAs in metastasis and epithelial-mesenchymal transition. *Cellular and Molecular Life Sciences : CMLS*, *66*(10), 1682–99. doi:10.1007/s00018-009-8750-1

Bredy, T. W., Lin, Q., Wei, W., Baker-Andresen, D., & Mattick, J. S. (2011). MicroRNA regulation of neural plasticity and memory. *Neurobiology of Learning and Memory*, *96*(1), 89–94. doi:10.1016/j.nlm.2011.04.004

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, *71*(3), 527–42.

Brown, M., Suryawanshi, H., Hafner, M., Farazi, T. a, & Tuschl, T. (2013). Mammalian miRNA curation through next-generation sequencing. *Frontiers in Genetics*, *4*(August), 145. doi:10.3389/fgene.2013.00145

Burroughs, A. M., Ando, Y., de Hoon, M. L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., & Daub, C. O. (2011). Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biology*, *8*(1), 158–177. doi:10.4161/rna.8.1.14300

Bussotti, G., Notredame, C., & Enright, A. J. (2013). Detecting and comparing non-coding RNAs in the high-throughput era. *International Journal of Molecular Sciences*, *14*(8), 15423–58. doi:10.3390/ijms140815423

Campo-Paysaa, F., Sémon, M., Cameron, R. A., Peterson, K. J., & Schubert, M. (2011). microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evolution & Development*, *13*(1), 15–27. doi:10.1111/j.1525-142X.2010.00452.x

Carmell, M. a, Xuan, Z., Zhang, M. Q., & Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes & Development*, *16*(21), 2733–42. doi:10.1101/gad.1026102

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642–55. doi:10.1016/j.cell.2009.01.035

Cerutti, H., & Casas-Mollano, J. A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics*, *50*(2), 81–99. doi:10.1007/s00294-006-0078-x

Chen, K., & Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. *Nature Genetics*, *38*(12), 1452–6. doi:10.1038/ng1910

Chendrimada, T. P., Finn, K. J., Ji, X., Baillat, D., Gregory, R. I., Liebhaber, S. a, … Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. *Nature*, *447*(7146), 823–8. doi:10.1038/nature05841

Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. C., Spies, N., Baek, D., … Bartel, D. P. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes & Development*, *24*(10), 992–1009. doi:10.1101/gad.1884710

Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., … Arendt, D. (2010). Ancient animal microRNAs and the evolution of tissue identity. *Nature*, *463*(7284), 1084–8. doi:10.1038/nature08744

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., … Georges, M. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, *38*(7), 813–8. doi:10.1038/ng1810

Croce, C. M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews. Genetics*, *10*(10), 704–14. doi:10.1038/nrg2634

Dannemann, M., Nickel, B., Lizano, E., Burbano, H. a, & Kelso, J. (2012). Annotation of primate miRNAs by high throughput sequencing of small RNA libraries. *BMC Genomics*, *13*(1), 116. doi:10.1186/1471-2164-13-116

De Pontual, L., Yao, E., Callier, P., Faivre, L., Drouin, V., Cariou, S., … Amiel, J. (2011). Germline deletion of the miR-17∼92 cluster causes skeletal and growth defects in humans. *Nature Genetics*, *43*(10), 1026–30. doi:10.1038/ng.915

De Wit, E., Linsen, S. E. V, Cuppen, E., & Berezikov, E. (2009). Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Research*, *19*(11), 2064–74. doi:10.1101/gr.093781.109

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–89. doi:10.1101/gr.132159.111

Du, T., & Zamore, P. D. (2005). microPrimer: the biogenesis and function of microRNA. *Development (Cambridge, England)*, *132*(21), 4645–52. doi:10.1242/dev.02070

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., & Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. *Cell*, *150*(1), 100–10. doi:10.1016/j.cell.2012.05.017

Eulalio, A., Behm-Ansmant, I., & Izaurralde, E. (2007). P bodies: at the crossroads of post-transcriptional pathways. *Nature Reviews. Molecular Cell Biology*, *8*(1), 9–22. doi:10.1038/nrm2080

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.-F., Doerks, T., … Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes & Development*, *21*(20), 2558–70. doi:10.1101/gad.443107

Fénelon, K., Mukai, J., Xu, B., Hsu, P.-K., Drew, L. J., Karayiorgou, M., … Gogos, J. a. (2011). Deficiency of Dgcr8, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4447–52. doi:10.1073/pnas.1101219108

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews. Genetics*, *9*(2), 102–14. doi:10.1038/nrg2290

Fire, A., Xu, S., Montgomery, M., & Kostas, S. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(February), 806–811.

Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nature Reviews. Genetics*, *10*(2), 94–108. doi:10.1038/nrg2504

Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., … Schier, A. F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science (New York, N.Y.)*, *308*(5723), 833–8. doi:10.1126/science.1109020

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, *328*(5979), 710–22. doi:10.1126/science.1188021

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., … Bartel, D. P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, *455*(7217), 1193–7. doi:10.1038/nature07415

Guerra-Assunção, J. A., & Enright, A. J. (2012). Large-scale analysis of microRNA evolution. *BMC Genomics*, *13*, 218. doi:10.1186/1471-2164-13-218

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., … Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, *125*(5), 887–901. doi:10.1016/j.cell.2006.03.043

Hansen, T. B., Wiklund, E. D., Bramsen, J. B., Villadsen, S. B., Statham, A. L., Clark, S. J., & Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *The EMBO Journal*, *30*(21), 4414–22. doi:10.1038/emboj.2011.359

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews. Genetics*, *5*(7), 522–31. doi:10.1038/nrg1379

Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. J., & Peterson, K. J. (2008). MicroRNAs and the advent of vertebrate morphological

complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(8), 2946–50. doi:10.1073/pnas.0712259105

Hengst, U., Cox, L. J., Macosko, E. Z., & Jaffrey, S. R. (2006). Functional and selective RNA interference in developing axons and growth cones. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(21), 5727–32. doi:10.1523/JNEUROSCI.5229-05.2006

Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., … Stadler, P. F. (2006). The expansion of the metazoan microRNA repertoire. *BMC Genomics*, *7*, 25. doi:10.1186/1471-2164-7-25

Hoffman, A. E., Zheng, T., Yi, C., Leaderer, D., Weidhaas, J., Slack, F., … Zhu, Y. (2009). microRNA miR-196a-2 and breast cancer: a genetic and epigenetic association study and functional analysis. *Cancer Research*, *69*(14), 5970–7. doi:10.1158/0008-5472.CAN-09-0236

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., … Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, *129*(1), 69–82. doi:10.1016/j.cell.2007.03.026

Hwang, H.-W., Wentzel, E. a, & Mendell, J. T. (2007). A hexanucleotide element directs microRNA nuclear import. *Science (New York, N.Y.)*, *315*(5808), 97–100. doi:10.1126/science.1136235

Ishizu, H., Siomi, H., & Siomi, M. C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes & Development*, *26*(21), 2361–73. doi:10.1101/gad.203786.112

Isik, M., Korswagen, H. C., & Berezikov, E. (2010). Expression patterns of intronic microRNAs in Caenorhabditis elegans. *Silence*, *1*(1), 5. doi:10.1186/1758-907X-1-5

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., & Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biology*, *2*(11), e363. doi:10.1371/journal.pbio.0020363

Kawahara, Y., Megraw, M., Kreider, E., Iizasa, H., Valente, L., Hatzigeorgiou, A. G., & Nishikura, K. (2008). Frequency and fate of microRNA editing in human brain. *Nucleic Acids Research*, *36*(16), 5270–80. doi:10.1093/nar/gkn479

Kawahara, Y., Zinshteyn, B., Chendrimada, T. P., Shiekhattar, R., & Nishikura, K. (2007). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Reports*, *8*(8), 763–9. doi:10.1038/sj.embor.7401011

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., & Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science (New York, N.Y.)*, *315*(5815), 1137–40. doi:10.1126/science.1138050

Kiriakidou, M., Tan, G. S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P. T., & Mourelatos, Z. (2007). An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell*, *129*(6), 1141–51. doi:10.1016/j.cell.2007.05.016

Kloosterman, W. P., Wienholds, E., de Bruijn, E., Kauppinen, S., & Plasterk, R. H. A. (2006). In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nature Methods*, *3*(1), 27–9. doi:10.1038/nmeth843

Kozomara, A., & Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, *39*(Database issue), D152–7. doi:10.1093/nar/gkq1027

Krol, J., Loedige, I., & Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews. Genetics*, *11*(9), 597–610. doi:10.1038/nrg2843

Kuhn, S., Johnson, S. L., Furness, D. N., Chen, J., Ingham, N., Hilton, J. M., … Marcotti, W. (2011). miR-96 regulates the progression of differentiation in mammalian cochlear inner and outer hair cells. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2355–60. doi:10.1073/pnas.1016646108

Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., … Tóth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & Development*, *27*(4), 390–9. doi:10.1101/gad.209841.112

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–54.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, *23*(20), 4051–60. doi:10.1038/sj.emboj.7600385

Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., & Carthew, R. W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, *117*(1), 69–81.

Leonardo, T. R., Schultheisz, H. L., Loring, J. F., & Laurent, L. C. (2012). The functions of microRNAs in pluripotency and reprogramming. *Nature Cell Biology*, *14*(11), 1114–21. doi:10.1038/ncb2613

Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, *120*(1), 15–20. doi:10.1016/j.cell.2004.12.035

Li, J., & Zhang, Z. (2012). miRNA regulatory variation in human evolution. *Trends in Genetics : TIG*, 1–9. doi:10.1016/j.tig.2012.10.008

Li, X.-D., Li, Z.-G., Song, X.-X., & Liu, C.-F. (2010). A variant in microRNA-196a2 is associated with susceptibility to hepatocellular carcinoma in Chinese patients with cirrhosis. *Pathology*, *42*(7), 669–73. doi:10.3109/00313025.2010.522175

Liang, H., & Li, W.-H. (2009). Lowly expressed human microRNA genes evolve rapidly. *Molecular Biology and Evolution*, *26*(6), 1195–8. doi:10.1093/molbev/msp053

Lin, H., & Yin, H. (2008). A novel epigenetic mechanism in Drosophila somatic cells mediated by Piwi and piRNAs. *Cold Spring Harbor Symposia on Quantitative Biology*, *73*, 273–81. doi:10.1101/sqb.2008.73.056

Lingel, A., Simon, B., Izaurralde, E., & Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Structural & Molecular Biology*, *11*(6), 576–7. doi:10.1038/nsmb777

Liu, N., Okamura, K., Tyler, D. M., Phillips, M. D., Chung, W. J., & Lai, E. C. (2008). The evolution and functional diversification of animal microRNA genes. *Cell Research*, *18*(10), 985–996. doi:10.1038/cr.2008.278.The

Lu, J., Fu, Y., Kumar, S., Shen, Y., Zeng, K., Xu, A., … Wu, C.-I. (2008). Adaptive evolution of newly emerged micro-RNA genes in Drosophila. *Molecular Biology and Evolution*, *25*(5), 929–38. doi:10.1093/molbev/msn040

Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., … Wu, C.-I. (2008). The birth and death of microRNA genes in Drosophila. *Nature Genetics*, *40*(3), 351–5. doi:10.1038/ng.73

Ma, J.-B., Ye, K., & Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, *429*(6989), 318–22. doi:10.1038/nature02519

MacRae, I. J., Zhou, K., & Doudna, J. a. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nature Structural & Molecular Biology*, *14*(10), 934–40. doi:10.1038/nsmb1293

Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., … Doudna, J. a. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science (New York, N.Y.)*, *311*(5758), 195–8. doi:10.1126/science.1121638

Majdalani, N., Vanderpool, C. K., & Gottesman, S. (2005). Bacterial small RNA regulators. *Critical Reviews in Biochemistry and Molecular Biology*, *40*(2), 93–113. doi:10.1080/10409230590918702

Makarova, K. S., Grishin, N. V, Shabalina, S. a, Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, *1*, 7. doi:10.1186/1745-6150-1-7

Malone, C. D., & Hannon, G. J. (2009). Small RNAs as guardians of the genome. *Cell*, *136*(4), 656–68. doi:10.1016/j.cell.2009.01.045

Marco, A., Hui, J. H. L., Ronshaugen, M., & Griffiths-Jones, S. (2010). Functional shifts in insect microRNA evolution. *Genome Biology and Evolution*, *2*, 686–96. doi:10.1093/gbe/evq053

Maroney, P. a, Yu, Y., Fisher, J., & Nilsen, T. W. (2006). Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nature Structural & Molecular Biology*, *13*(12), 1102–7. doi:10.1038/nsmb1174

Martinez, N. J., Ow, M. C., Reece-Hoyes, J. S., Barrasa, M. I., Ambros, V. R., & Walhout, A. J. M. (2008). Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. *Genome Research*, *18*(12), 2005–15. doi:10.1101/gr.083055.108

Mathonnet, G., Fabian, M. R., Svitkin, Y. V, Parsyan, A., Huck, L., Murata, T., … Sonenberg, N. (2007). MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science (New York, N.Y.)*, *317*(5845), 1764–7. doi:10.1126/science.1146067

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, *15*(2), 185–97. doi:10.1016/j.molcel.2004.07.007

Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., … Moreno-Pelayo, M. A. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics*, *41*(5), 609–13. doi:10.1038/ng.355

Mendell, J. T., & Olson, E. N. (2012). MicroRNAs in stress signaling and human disease. *Cell*, *148*(6), 1172–87. doi:10.1016/j.cell.2012.02.005

Meng, Y., Shao, C., Wang, H., & Chen, M. (2012). Are all the miRBase-registered microRNAs true? *RNA Biology*, *9*(3), 249–253.

Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, K. H. (2013). Birth and expression evolution of mammalian microRNA genes. *Genome Research*, 1–12. doi:10.1101/gr.140269.112.Freely

Mishima, Y., Giraldez, A. J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A. F., & Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Current Biology : CB*, *16*(21), 2135–42. doi:10.1016/j.cub.2006.08.086

Morin, R., O'Connor, M., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A., … Marra, M. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research*, *18*(4), 610.

Morris, K. V, & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews. Genetics*, *15*(6), 423–37. doi:10.1038/nrg3722

Muiños-Gimeno, M., Espinosa-Parrilla, Y., Guidi, M., Kagerbauer, B., Sipilä, T., Maron, E., … Estivill, X. (2010). Human microRNAs miR-22, miR-138-2, miR-148a, and miR-488 Are Associated with Panic Disorder and Regulate Several Anxiety Candidate Genes and Related Pathways. *Biological Psychiatry*. doi:10.1016/j.biopsych.2010.10.010

Murphy, D., Dancis, B., & Brown, J. R. (2008). The evolution of core proteins involved in microRNA biogenesis. *BMC Evolutionary Biology*, *8*, 92. doi:10.1186/1471-2148-8-92

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annual Review of Biochemistry*, *79*, 321–49. doi:10.1146/annurev-biochem-060208-105251

Nygaard, S., Jacobsen, A., Lindow, M., Eriksen, J., Balslev, E., Flyger, H., … Litman, T. (2009). Identification and analysis of miRNAs in human breast cancer and teratoma samples using deep sequencing. *BMC Medical Genomics*, *2*, 35. doi:10.1186/1755-8794-2-35

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., & Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, *130*(1), 89–100. doi:10.1016/j.cell.2007.06.028

Okamura, K., Ishizuka, A., Siomi, H., & Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & Development*, *18*(14), 1655–66. doi:10.1101/gad.1210204

Okamura, K., Phillips, M. D., Tyler, D. M., Duan, H., Chou, Y., & Lai, E. C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nature Structural & Molecular Biology*, *15*(4), 354–63. doi:10.1038/nsmb.1409

Ooi, C. H., Oh, H. K., Wang, H. Z., Tan, A. L. K., Wu, J., Lee, M., … Tan, P. (2011). A densely interconnected genome-wide network of microRNAs and oncogenic pathways revealed using gene expression signatures. *PLoS Genetics*, *7*(12), e1002415. doi:10.1371/journal.pgen.1002415

Ozsolak, F., Poling, L. L., Wang, Z., Liu, H., Liu, X. S., Roeder, R. G., … Fisher, D. E. (2008). Chromatin structure analyses identify miRNA promoters. *Genes & Development*, *22*(22), 3172–83. doi:10.1101/gad.1706508

Park, C. Y., Choi, Y. S., & McManus, M. T. (2010). Analysis of microRNA knockouts in mice. *Human Molecular Genetics*, *19*(R2), R169–75. doi:10.1093/hmg/ddq367

Parker, J. S., Roe, S. M., & Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, *434*(7033), 663–6. doi:10.1038/nature03462

Parker, R., & Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. *Molecular Cell*, *25*(5), 635–46. doi:10.1016/j.molcel.2007.02.011

Patil, V. S., Zhou, R., & Rana, T. M. (2013). Gene regulation by non-coding RNAs. *Critical Reviews in Biochemistry and Molecular Biology*, *9238*, 1–17. doi:10.3109/10409238.2013.844092

Peng, S., Kuang, Z., Sheng, C., Zhang, Y., Xu, H., & Cheng, Q. (2010). Association of microRNA-196a-2 gene polymorphism with gastric cancer risk in a Chinese population. *Digestive Diseases and Sciences*, *55*(8), 2288–93. doi:10.1007/s10620-009-1007-x

Petersen, C. P., Bordeleau, M.-E., Pelletier, J., & Sharp, P. a. (2006). Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell*, *21*(4), 533–42. doi:10.1016/j.molcel.2006.01.031

Piriyapongsa, J., Mariño-Ramírez, L., & Jordan, I. K. (2007). Origin and evolution of human microRNAs from transposable elements. *Genetics*, *176*(2), 1323–37. doi:10.1534/genetics.107.072553

Plaisier, C. L., Pan, M., & Baliga, N. S. (2012). A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. *Genome Research*, *22*(11), 2302–14. doi:10.1101/gr.133991.111

Plasterk, R. H. a. (2006). Micro RNAs in animal development. *Cell*, *124*(5), 877–81. doi:10.1016/j.cell.2006.02.030

Quach, H., Barreiro, L. B., Laval, G., Zidane, N., Patin, E., Kidd, K. K., … Quintana-Murci, L. (2009). Signatures of purifying and local positive selection in human miRNAs. *American Journal of Human Genetics*, *84*(3), 316–27. doi:10.1016/j.ajhg.2009.01.022

Rakoczy, J., Fernandez-Valverde, S. L., Glazov, E. a, Wainwright, E. N., Sato, T., Takada, S., … Wilhelm, D. (2013). MicroRNAs-140-5p/140-3p modulate Leydig cell numbers in the developing mouse testis. *Biology of Reproduction*, *88*(6), 143. doi:10.1095/biolreprod.113.107607

Robinson, J. M., Sperling, E. a, Bergum, B., Adamski, M., Nichols, S. a, Adamska, M., & Peterson, K. J. (2013). The identification of microRNAs in calcisponges: independent evolution of microRNAs in basal metazoans. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, *320*(2), 84–93. doi:10.1002/jez.b.22485

Rodriguez, A., & Griffiths-Jones, S. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Research*, *14*(10A), 1902–1910. doi:10.1101/gr.2722704.

Ruby, J. G., Jan, C. H., & Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*, *448*(7149), 83–6. doi:10.1038/nature05983

Ruby, J., Stark, A., & Johnston, W. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome ...*, *17*, 1850–1864. doi:10.1101/gr.6597907.

Sasaki, T., Shiohama, A., Minoshima, S., & Shimizu, N. (2003). Identification of eight members of the Argonaute family in the human genome , *82*, 323–330. doi:10.1016/S0888-7543(03)00129-0

Saunders, M. A., Liang, H., & Li, W. (2007). Human polymorphism at microRNAs and microRNA target sites. *PNAS*, *104*(9), 3300–3305.

Schmitter, D., Filkowski, J., Sewer, A., Pillai, R. S., Oakeley, E. J., Zavolan, M., … Filipowicz, W. (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Research*, *34*(17), 4801–15. doi:10.1093/nar/gkl646

Schnall-Levin, M., Rissland, O. S., Johnston, W. K., Perrimon, N., Bartel, D. P., & Berger, B. (2011). Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. *Genome Research*, *21*(9), 1395–403. doi:10.1101/gr.121210.111

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, *455*(7209), 58–63. doi:10.1038/nature07228

Sempere, L., Cole, C., McPeek, M., & Peterson, K. (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. … *Evolution*, *588*(April), 575–588. doi:10.1002/jez.b

Sen, G. L., & Blau, H. M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature Cell Biology*, *7*(6), 633–6. doi:10.1038/ncb1265

Shabalina, S., & Koonin, E. (2008). Origins and evolution of eukaryotic RNA interference. *Trends in Ecology & Evolution*, *23*(10), 578–587. doi:10.1016/j.tree.2008.06.005.Origins

Song, J.-J., Smith, S. K., Hannon, G. J., & Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science (New York, N.Y.)*, *305*(5689), 1434–7. doi:10.1126/science.1102514

Sorek, R., Kunin, V., & Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews. Microbiology*, *6*(3), 181–6. doi:10.1038/nrmicro1793

Stark, M. S., Tyagi, S., Nancarrow, D. J., Boyle, G. M., Cook, A. L., Whiteman, D. C., … Hayward, N. K. (2010). Characterization of the Melanoma miRNAome by Deep Sequencing. *PloS One*, *5*(3), e9685. doi:10.1371/journal.pone.0009685

Thermann, R., & Hentze, M. W. (2007). Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature*, *447*(7146), 875–8. doi:10.1038/nature05878

Tian, T., Shu, Y., Chen, J., Hu, Z., Xu, L., Jin, G., … Shen, H. (2009). A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *18*(4), 1183–7. doi:10.1158/1055-9965.EPI-08-0814

Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., … Ladurner, A. G. (2007). A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nature Structural & Molecular Biology*, *14*(10), 897–903. doi:10.1038/nsmb1302

Tsang, J., Zhu, J., & van Oudenaarden, A. (2007). MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Molecular Cell*, *26*(5), 753–67. doi:10.1016/j.molcel.2007.05.018

Vagin, V. V, Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science (New York, N.Y.)*, *313*(5785), 320–4. doi:10.1126/science.1129333

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., & Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, *9*(6), 654–9. doi:10.1038/ncb1596

Wang, H., Zhang, P., Chen, W., Jie, D., Dan, F., Jia, Y., & Xie, L. (2013). Characterization and Identification of novel serum microRNAs in sepsis patients with different outcomes. *Shock (Augusta, Ga.)*, *39*(6), 480–7. doi:10.1097/SHK.0b013e3182940cb8

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., & Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature*, *456*(7219), 209–13. doi:10.1038/nature07315

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. a, Holstein, T. W., Heber, S., & Peterson, K. J. (2009). The deep evolution of metazoan microRNAs. *Evolution & Development*, *11*(1), 50–68. doi:10.1111/j.1525-142X.2008.00302.x

Wienholds, E., Kloosterman, W. P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., … Plasterk, R. H. a. (2005). MicroRNA expression in zebrafish embryonic development. *Science (New York, N.Y.)*, *309*(5732), 310–1. doi:10.1126/science.1114519

Wienholds, E., & Plasterk, R. H. a. (2005). MicroRNA function in animal development. *FEBS Letters*, *579*(26), 5911–22. doi:10.1016/j.febslet.2005.07.070

Xie, Z., Allen, E., Wilken, A., & Carrington, J. C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana.

Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., & Zhou, M.-M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, *426*(6965), 468–74. doi:10.1038/nature02129

Yin, H., & Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature*, *450*(7167), 304–8. doi:10.1038/nature06263

Yu, X., Lin, J., Zack, D. J., Mendell, J. T., & Qian, J. (2008). Analysis of regulatory network topology reveals functionally distinct classes of microRNAs. *Nucleic Acids Research*, *36*(20), 6494–503. doi:10.1093/nar/gkn712

Yuan, Z., Sun, X., Liu, H., & Xie, J. (2011). MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. *PloS One*, *6*(3), e17666. doi:10.1371/journal.pone.0017666

Zhao, Y., Samal, E., & Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, *436*(7048), 214–20. doi:10.1038/nature03817