UNIVERSITA' DEGLI STUDI DI CATANIA

Scuola "FACOLTA' DI MEDICINA"
Dipartimento di Medicina Clinica e Sperimentale

Dottorato di Ricerca in *Medicina Molecolare*

XXVI Ciclo – 2010/2014

Dario Veneziano

Knowledge bases, computational methods and data mining techniques with applications to A-to-I RNA editing, Synthetic Biology and RNA interference

Tesi di Dottorato

Coordinatore e Relatore:

Chiar.mo Prof. Alfredo FERRO

Tutor:

Dr. Alessandro LAGANA'

To my precious Nikki

and my loving family

Contents

Introduction

The central dogma of molecular biology has represented for over fifty years the basis of genetics [1]. It essentially described the genetic information flow of life in which DNA and protein, as repository and functional incarnation of that information respectively, have been perceived as the two main actors in the cell's life cycle, confining RNA simply to the role of template for protein synthesis. Nevertheless, this view of the biological role of RNA, initially apparently exhaustive, has been over time subjected to challenges, as firstly suggested by Gilbert in 1986 [2].

Within the context of this new hypothesized "RNA World", one of the most significant biological discoveries of the past two decades was made, that is, the existence of several types of RNAs, each with their specific functions in eukaryotic cells [3,4].

As the ENCODE project has confirmed, most of the human genome is in fact transcribed, but only a very small fraction of it encodes for proteins [5,6]. Indeed, the larger remaining portion of the transcribed genomic output is represented by a diverse family of untranslated transcripts that play crucial roles in many biochemical cellular processes [7]. A category of such RNA molecules which has stood out in the midst of this vast panorama

and whose discovery has pioneered the investigation of a new level of understanding in gene regulation has been that of microRNAs.

The discovery of these small regulatory RNAs, whose role has been crucial in a variety of physiological and pathological processes, not only has radically changed previous understanding of the mechanism behind the regulation of gene expression but, at the same time, has opened up new frontiers in the development of new therapeutic approaches which are more specific and less toxic, especially against all those diseases posing serious challenges to traditional therapies. The huge amount of new data produced daily urges the need for algorithms and computational tools that allow their proper interpretation as well as the formulation of novel hypotheses about the genesis and function of these small regulatory molecules.

Additional discoveries in recent years have unveiled how other mechanisms within the 'RNA world' have eventually revealed to be connected to each other. That is the case of the RNA editing phenomenon in relation to miRNA-mediated gene regulation.

All of this could not have taken place without the aid of computer technology, as it contributed greatly to the advancements of the biological and medical sciences.
Until the second half of twentieth century, the connection between Biology and Computer Science was not so tight and data was usually collected on perishable materials such as paper and stored up in filing cabinets.
With the advent of Bioinformatics the landscape changed profoundly. Bioinformatics, also known as Computational Biology, is a relatively novel field that aims at the

resolution of biological problems through computational approaches. This interdisciplinary science pursues two particular objectives among many: on one hand, the construction of biological databases in order to rationally store the ever increasing amount of data that is becoming more available, and, on the other, the development and application of algorithms in order to extract predicting patterns and mine novel information otherwise impossible to attain from such data.

This thesis will present novel results on both of the above aspects. In fact, the doctoral research work described in this thesis has had as its objective the development of heuristics and data mining techniques for the collection and analysis of data concerning the mechanisms of post-transcriptional regulation and RNA interference, as well as the connection of the RNA A-to-I editing phenomenon to miRNA-mediated gene regulation.

In particular, efforts have been aimed at the development of a database for the prediction of RNA A-to-I edited miRNA binding sites; an algorithm for the design of synthetic miRNAs with high specificity; and a knowledge base provided with data mining algorithms for the functional annotation of microRNAs, serving as a unified resource for miRNA research.

Chapter 1

Post Transcriptional Gene Regulation and RNA Interference

The discovery of the first short ncRNA capable of acting as endogenous regulator of gene expression was made by Ambros et al. in 1993, while investigating the role of lin-4 in the developmental timing of C. elegans [8]. That accomplishment revealed only a glimpse of the much broader reality uncovered by Fire and Mello five years later when they reported the capability of exogenous double stranded RNAs to silence genes in a specific manner, disclosing the mechanism we know today as RNA interference (RNAi) [9]. In 1999, a similar process was discovered to take place in plants, as short RNA sequences (~20-25 nt) were found capable to bind their mRNA targets through perfect base complementarity [10]. Since then, a revolution has occured in the field of RNA biology thanks to the characterization of RNAi as an innate biological process through which the expression of specifically targeted genes can be modulated and/or silenced, ushering in a new world of research and potential therapeutic applications. Specifically, RNAi is a naturally occuring mechanism resulting in a sequence-specific, post-transcriptional down-regulation of gene expression induced by double-stranded RNA (dsRNA) homologous to the target gene. This regulation can occur at different levels of the gene expression process, including transcription, mRNA processing and translation. The small RNA molecules responsible

for RNAi originate from endogenous and exogenous double strand precursors and play a specific role in guiding effector protein complexes toward their nucleic acid targets by partial or full complementarity bonds [11].

Although several classes of regulatory ncRNAs have been identified, the most relevant ones can essentially be represented by two groups, according to their origin, structure, associated effector proteins and function: small interfering RNAs (siRNAs), principally of exogenous origin in animals, and microRNAs (miRNAs), which are endogenous genome products. These molecules, apparently present only in eukaryotes and in some viruses, are in fact the most abundant regulatory molecules in terms of both phylogeny and physiology.

1.1 siRNA and miRNA

In 2001 the first major characterization for these small regulatory ncRNA molecules was able to functionally distinguish the two major groups previously mentioned: miRNAs clearly emerged as regulators of endogenous genes, while siRNAs as defenders of the integrity of the genome, neutralizing exogenous nucleic acids such as originating from viruses and transposons [12]. Subsequently, in 2004, it was discovered that a specific protein complex, later termed as RNA-Induced Silencing Complex (RISC), was found associated to single-stranded forms of miRNAs and siRNAs [13]. In both cases, the complementarity of the bases mediated the targeting, thus identifying the target genes. As research progressed, their differences emerged.

siRNA is a class of double-stranded RNA molecules, 20-25 base pairs in length, whose role appears to be involved in the defense of the cell and maintenance of genome integrity through the silencing of exogenous nucleic acids and undesired transcripts (such

as transposones and repetitive elements) [11,14]. Although their post-transcriptional-gene-silencing role in plants and in some simple animal species was shown to be of endogenous origin, siRNAs in animals, instead, are mainly exogenous molecules obtained from perfectly complementary long double stranded precursors coming from viruses and transposons [10]. As it was first demonstrated in nematodes [9] and later in mammalians [15], siRNA-mediated RNAi can be obtained either by simple dsRNAs of about 21 nucleotides (nt) with two-nucleotide 3' overhang or by stably expressed short hairpin RNAs (shRNAs), which are processed into siRNAs [16,17]. Generally, shRNA is transcribed in cells from a DNA template as a single- stranded RNA molecule (about 50–100 bases). The complementary regions are spaced by a hairpin loop, therefore the name "short hairpin" RNA [18]. Intracellular presence of siRNAs which are perfectly complementary to their target mRNAs is of crucial importance for the induction of RNAi and leads to mRNA degradation.

While siRNAs are mainly exogenous molecules, miRNAs instead are a class of RNAi inducers which derive from partially complementary double strand hairpin precursors of endogenous origin. Once processed, they are small single stranded RNAs (20-22 nt long) able to modulate post-transcriptional gene silencing through repression, and at times degradation, of specific mRNA target molecules [12]. Specifically, it has been suggested that destabilization of the target mRNA is the predominant reason (≥84%) for reduced protein levels by endogenous miRNAs [19]. It has been estimated that miRNA-coding genes represent 1% of the total gene population, regulating the expression of at least 50% of the genes in a cell at the post-transcriptional level [20] thus constituting the biggest class of regulatory molecules and the most dominating in most somatic tissues. They are present in plants, higher eukaryotes and in some viruses. In light of this, it is not

surprising that their dysregulation is often associated, if not even the cause, of several human diseases such as cancer, as extensively documented by Croce [21].

From here on, in spite of their originating difference, common features, such as the length of their mature products and their sequence-specific inhibitory functions, suggest that siRNA and miRNA have similar processing and common mechanisms. However, a fundamental difference between miRNAs and siRNAs consists in the type of binding, which is considered a key factor in their regulatory function: miRNAs bind their target with partial complementarity, allowing bulges and loops in duplexes. The presence of mismatches in the central part of the duplex is usually associated to translational repression, which seems to be the default mechanism of miRNA-mediated RNAi. The cleavage of perfectly paired duplexes, which is the default RNAi mechanism in the case of siRNAs, is instead considered for miRNAs an additional feature leading to the same effect on the protein level.

The biological role of these molecules is currently being intensively elucidated and their involvement in fundamental processes, such as apoptosis, metabolism, cell proliferation and organism development has been widely demonstrated.

1.2 miRNA-coding regions in the human genome

All miRNAs discovered, either through experimental approaches or computational analysis, are stored in the *miRBase* database (www.mirbase.org), a register of miRNAs which extensively reports all their features [22,23]. Among these, important information about the genomics of miRNAs, that is, their organization in human chromosomes, is also reported. miRNAs are distributed on all chromosomes, except for the Y chromosome, often encoded in clusters by genes usually located in introns and, more rarely, in exons of

protein coding genes, as well as in intergenic regions [12]. The miRNAs coded in a cluster are often related to each other also on a functional standpoint, as it happens for the cluster of miRNAs hsa-let-7a-1, hsa-let-7f-1, and hsa-let-7d in chromosome 9.

According to their localization, miRNAs in the genome can be distinguished in:

- Intergenic MiRNAs;
- Intronic miRNAs in encoding transcripts;
- Intronic miRNAs in non-coding transcripts;
- Esonic miRNAs in non-coding transcripts.

Generally, about 70% of miRNA-coding genes are located in intergenic regions, while the residual 30% can be found in intronic sequences of specific “guest” genes. This means that miRNAs can also be located in transcriptional and independent units, even though a large number of them is processed from transcripts containing either a cluster of miRNAs or intronic sequences of a host gene.

1.3 miRNA biogenesis

The biogenesis of miRNAs consists in a multistage process under tight temporal and spatial control, which starts in the nucleus and it continues and ends in the cytoplasm. In contrast to the linear miRNA processing pathway that was initially thought to be universal for the biogenesis of all mature miRNAs (Figure 1a), multiple discoveries led to the recognition of miRNA-specific differences that open a plethora of regulatory options to express and process individual miRNAs differentially (Figure 1b and 1c). Here we cover the so-called "canonical" pathway for miRNA biogenesis, along with one of the main alternative pathways which do not involve the main mediating protein complexes of the canonical pathway.

1.3.1 Canonical biogenesis pathway

In miRNA canonical biogenesis, long primary transcripts, called pri-miRNAs, are transcribed from genomic DNA by RNA polymerase II or III [20]. The majority of miRNAs are found in intergenic regions and for this reason the transcription of intergenic miRNAs usually depends on their own promoters. On the other hand, other miRNAs genes are located in protein-coding genes resulting in the name of intragenic miRNAs: this group is divided into exonic and intronic miRNAs based on their localizations. The expression of intragenic miRNAs depends on the promoter of their host genes [20]. The miRNA must therefore be excised during its biogenesis to cause gene silencing; two endoribonucleolytic enzymes are responsible for this excision. The first endonucleolytic reaction takes place in the nucleus and is driven by two proteins associated with each other; the transcribed pri-miRs are processed to precursor miRNAs (pre-miRNAs) by the endoribonuclease Drosha in complex with the dsRNA binding protein DGCR8

(DiGeorge critical region 8). Drosha is a highly conserved 160 kDa protein containing two RNAse III domains and one double-strand RNA-binding domain. Drosha forms a huge complex, 500 kDa in *Drosphila melanogaster* (*D. melanogaster*) and 650 kDa in *Homo sapiens*, called Microprocessor and containing the co-factor DGCR8. DGCR8, known also as Pasha in *C. elegans* and *D. melanogaster*, recognizes and binds to the double strand region of the pri-miRNA, functioning as a molecular ruler: DGCR8 arranges the Drosha cut site 11bp from the base of the hairpin stem. This progression releases a pre-miRNAs that possesses a 3' extension. After being processed to pre-miRNAs, the precursors are exported to the cytoplasm via Exportin-5 (Exp5) in complex with Ran-GTP. Once exported, the pre-miRNA is processed by a second endoribonucleolytic reaction, catalyzed by RNAse III Dicer [24], which digests an ≈22nt RNA duplex with protruding 3' projections at both ends. At this stage of the processing an Argonaute protein binds the duplex miRNA/miRNA* and one strand, complementary to the target mRNA (guide strand), is selected and subsequently forms the miRNA effector (mature miRNA) as part of the miRISC (miRNA-induced silencing complex). Whereas one of the two strands is selected as guide strand according to thermodynamic properties, the remaining strand, termed also as passenger strand or miRNA* (star miRNA; many publications refer to the two strand pair as miR-3p/miR-5p, referring to the direction of the functional miRNA) is released and degraded. The so-called miRNA* was initially thought to be the strand subjected to degradation, while more recent evidences suggest that it does not simply represent a non-functional bioproduct of miRNA biogenesis, but it can be selected as a functional strand and play significant biological roles [25]. Once the processing steps are completed, the mature single stranded miRNA product is incorporated in the complex known as miRNA-containing

ribonucleoprotein complex (miRNP), or miRNA-containing RNA-induced silencing complex (miRgonaute or miR-RISC, a ribonucleoproteic complex containing Argonaute proteins, of which AGO1 and 2 have been the most widely studied). As a part of this complex, the mature miRNA is able to regulate gene expression at the post-trascriptional level, binding for the most part through partial complementarity to target mRNAs in mammals, and mainly leading to mRNA degradation or translation inhibition.

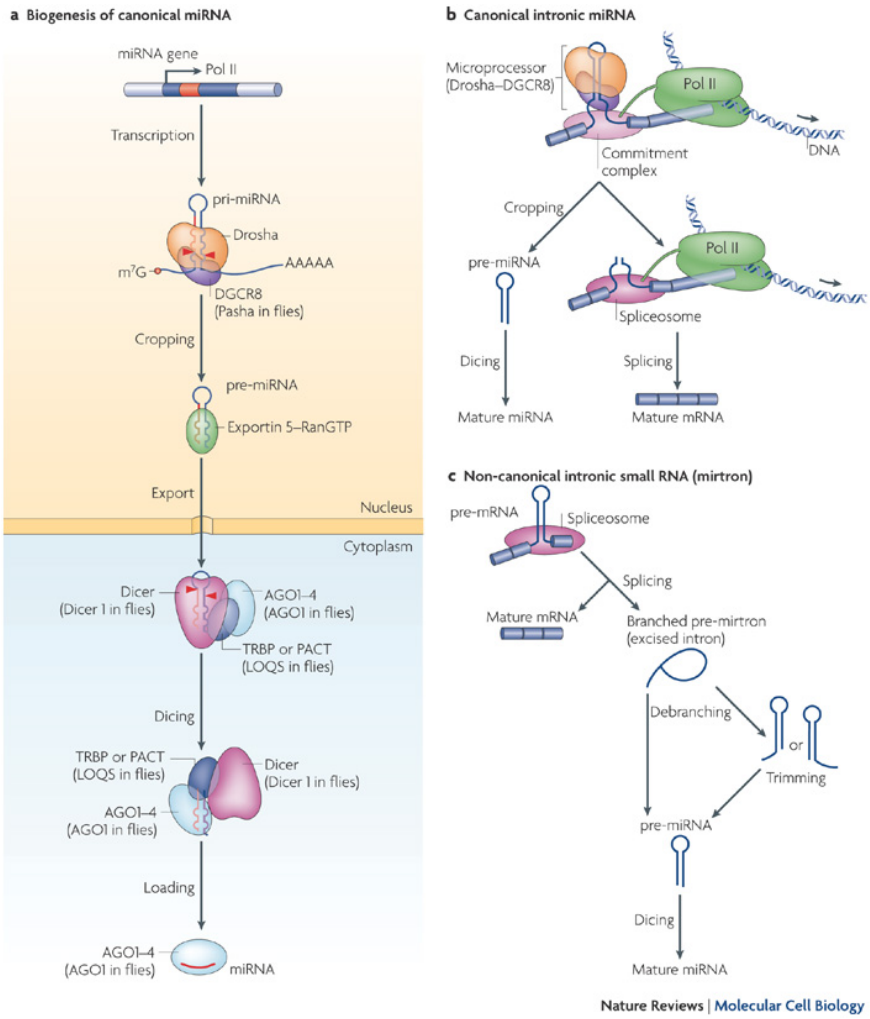

Figure 1: miRNA biogenesis pathways: a) Canonical miRNA genes are transcribed by RNA polymerase II (Pol II) to generate the primary transcripts (pri-miRNAs). The initiation step (cropping) is mediated by the Drosha–DiGeorge syndrome critical region gene 8 (DGCR8; Pasha in Drosophila melanogaster and Caenorhabditis elegans) complex (also known as the Microprocessor complex) that generates 65 nucleotide (nt) pre-miRNAs. Pre-miRNA has a short stem plus a 2-nt 3' overhang, which is recognized by the nuclear export factor exportin 5 (EXP5). On export from the nucleus, the cytoplasmic RNase III Dicer catalyses the second processing (dicing) step to produce miRNA duplexes. Dicer, TRBP (TAR RNA-binding protein; also known as TARBP2) or PACT (also known as PRKRA), and Argonaute (AGO)1–4 (also known as EIF2C1–4) mediate the processing of pre-miRNA and the assembly of the RISC (RNA-induced silencing complex) in humans. One strand of the duplex remains on the Ago protein as the mature miRNA, whereas the other strand is degraded. Ago is thought to be associated with Dicer in the dicing step as well as in the RISC assembly step. In D. melanogaster, Dicer 1, Loquacious (LOQS; also known as R3D1) and AGO1 are responsible for the same process. In flies, most miRNAs are loaded onto AGO1, whereas miRNAs from highly base-paired precursors are sorted into AGO2. The figure shows the mammalian processing pathways with fly components in brackets. b) Canonical intronic miRNAs are processed co-transcriptionally before splicing. The miRNA-containing introns are spliced more slowly than the adjacent introns for unknown reasons. The splicing commitment complex is thought to tether the introns while Drosha cleaves the miRNA hairpin. The pre-miRNA enters the miRNA pathway, whereas the rest of the transcript undergoes pre-mRNA splicing and produces mature mRNA for protein synthesis. c) Non-canonical intronic small RNAs are produced from spliced introns and debranching. Because such small RNAs (called mirtrons) can derive from small introns that resemble pre-miRNAs, they bypass the Drosha-processing step. Some introns have tails at either the 5' end or 3' end, so they need to be trimmed before pre-miRNA export. *m7G*, 7-methylguanosine. Figure and caption by Kim et al. [26].

1.3.2 Alternative biogenesis pathway

The two endoribonucleolytic cleavage steps represent the most obvious points at which functional miRNA production can be regulated. The first endoribonucleolytic digestion takes place in the nucleus as previously described and presents some unconventional processing steps of miRNAs. An alternative miRNA biogenesis pathway, called miRtron pathway, has been discovered among diverse mammals, drosophila and nematodes [27-30]: miRtrons are regulatory RNAs which get processed to form pre-miRs using the splicing machinery without Drosha-mediated cleavage (Fig. 1c). Moreover Drosha-DGCR8-mediated processing of let-7 pri-miRNA can be inhibited by Lin28B [31], although the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) binds specifically to pri-miR-18a to promote its processing [32]. In the cytoplasm, Dicer catalyzes the second endonucleolytic reaction and is essential for miRNA maturation; it's been also demonstrated that its knockdown in human cell lines marks the production of aberrant miRNAs [33] and moreover the accumulation of pre-miRNAs. A number of recent reports have described regulatory mechanisms that affect on Dicer processing, labeling Dicer as a regulatory node in the cytoplasm. One such controlling pathway modifies via 3'end uridylylation of the pre-miRNA which alternatively promotes or inhibits dicing and the pre-miRNA maturation [34,35] (Figure 2). Dicer is a very large enzyme (approximately 200 kDa) which is conserved among the species and presents several different domains: a double strand RNA-binding domain (dsRBD); two RNAse III catalytic domains, one which binds the 30-end small RNAs (PAZ domain) and a second which binds other domains with ATPasic and RNA-helicasic activity. Since Dicer does not function alone but in association with additional proteins (Figure 3), such as TRBP (transactivating response RNA-binding protein) and PACT (protein activator of PKR), we find even more

regulatory mechanisms at this level. TRBP and PACT interacting with Dicer may alter its activity to give different miRNA isoforms with new target specificities [36]. Furthermore, recent discoveries showed that members of the Argonaut family participate in the Dicer-mediated processing of pre-miRs, playing an important role in stabilizing the complex Dicer-miRNA [37]. In mammalians the Argonaut 2 (Ago2) protein complex, characterized by RNAse H activity, cooperates in the Dicer-mediated processing of some pre-miRNAs, yielding to another intermediate processing product, called AGO2-cleavd precursor miRNA (ac-pre-miRNA) [20].

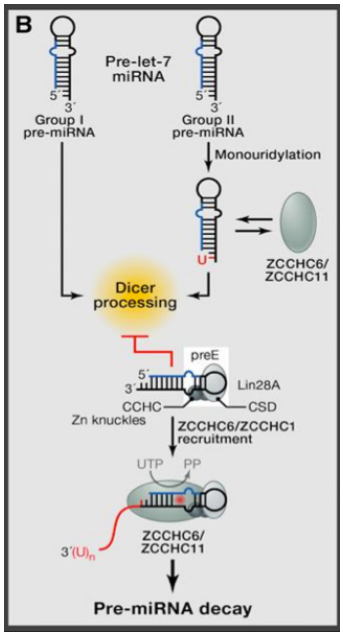

Figure 2: Modification of the 3' end of pre-miRNA by the untemplated addition of uridines catalized by cytoplasmatic TUTs. Mono-uridylylation facilitates Dicer processing, whereas Lin28 binding and oligo-uridylylation are inhibitory. Figure by Yates et al. [38]

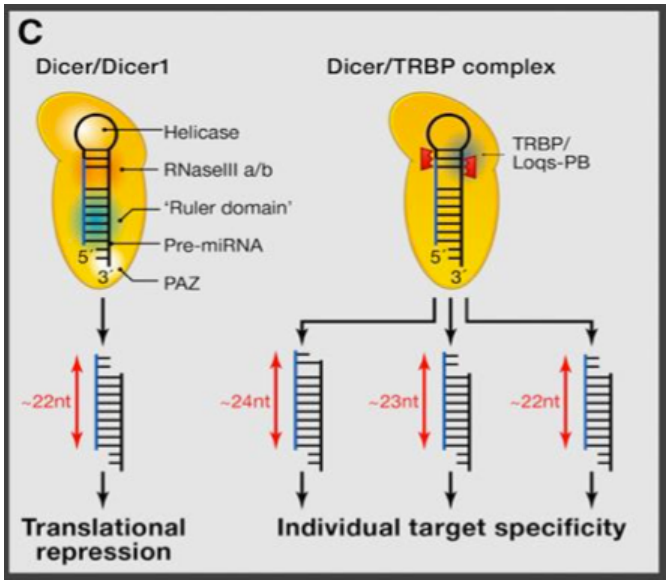

Figure 3: Tuning of Dicer cut-site selection (shifted by 1-2 nt) by its binding partners allows the generation of specific isomiRs with altered target specificities. Figure by Yates et al. [38].

1.4 Regulation of miRNA expression

The regulation of miRNA expression is fundamental to the role that these molecules play.

MicroRNA expression is regulated at various levels during biogenesis:

- *Regulation of transcription;*
- *Adjusting the processing;*
- *Editing;*
- *microRNA decay.*

The regulation of miRNA transcription is identical to that of many coding genes, thanks to the presence of the same regulatory elements (*TATA box sequences, CpG Islands, initialization elements*) at the level of their promoters.

Many *transcription factors* (**TFs**) regulate *tissue-* or *development-specific* expression, such as MYC/MYCN which stimulates the expression of the oncogenic cluster of *miR-*

17-5p in lymphoma cells [39], or *REST* which inhibits the expression of miR-124 in non-neuronal cells or neuronal progenitors through histonic de-acetylation and methylation of the promoter [40]. In turn, miRNAs can regulate the expression of TFs, creating thus circuits of positive or negative adjustment. In that instance, the total control over the quantity of miRNAs present or the TFs is what ultimately determines the final physiological effect.

The regulation of miRNA processing occurs at several levels: Drosha, Dicer and their accessory proteins. For example, some helicases of the rat and the *SMAD* proteins act at the Drosha level, controlling the production of pri-miR [20], while at the pre-miR level, the levels of Dicer are controlled and stabilized by *TRBP*, its cofactor [41].

The editing phenomenon of both the pri-miR and the pre-miR by ADAR proteins, which catalyze the conversion of adenosines into inosines, can alter their secondary structure (and therefore their stability), while the editing of the seed sequence of the mature miRNA product by other proteins would alter the recognition of the target [20].

Finally, the regulation of the stability and degradation of the mature miRNA can control the final quantity inside the cell and thus the biological effect. It was observed that miRNAs are generally more stable than second-class messengers and have a half-life, ranging from a few hours to many days. A complete control of the decay process may have a fundamental role in the development mechanisms and in the on-off switch response type, as, for example, in the development of the retina in mice, where the levels of miR-204 and miR-211 decrease rapidly in neurons but not in the glia [20].

1.5 Post Transcriptional miRNA-mediated gene regulation

Once the microRNA processing steps are completed, the mature single stranded miRNA is integrated in the miRNP or miR-RISC complex to inhibit their target genes by pairing with the 3'UTR of the messenger RNA (mRNA) (see Figure 4). The mature form of miRNA is about 22 nucleotides long but only a small region of the non-coding RNA binds to its complementary sequence in the 3'UTR of the target messenger: the 7-9 nucleotides from the 2^{nd} to 8^{th} (9^{th} or 10^{th}) nucleotide are identified as the "seed region" [12]. The seed sequence has a major role in miRNA function and its interaction with the 3'UTR target leads to repression of gene expression through three different mechanisms: (a) site-specific cleavage; (b) enhanced mRNA degradation; and (c) translation inhibition. The first one is exclusively Ago2-dependent and restricted to miRNAs whose total sequence displays a perfect or near-perfect complementarity with the target mRNA, an event which considered a very rare in mammals. The second and the third mechanisms just listed represent the common scenario in mammals and are frequently associated with mismatched miRNA/target sequences. Nonetheless the exact pathway through which miRNAs can impair translation is still debated, this mechanism together with the enhanced mRNA degradation are commonly defined as a *non-cleavage repression*, and can be carried out by any of the four components of the Ago family [42]. In addition, there are additional pathways able to control miRNA-mediated regulation of gene expression, as for instance the formation of mRNA secondary structure capable of inhibiting access to the target site for the small non-coding RNA [43]. Likewise, RNA binding proteins can cover the complementary region of mRNA targets by binding it and preventing the miRNA seed sequence from pairing. Aside form this canonical targeting regulation, more recent studies report that miRNAs can also bind the 5'-untraslated region (5' UTR) or the

Open Reading Frame (ORF) [44,45] and, even more remarkably, they can up-regulate translation upon growth arrest condition [46]. It is not surprising that each miRNA has the potential to target a large number of genes, since these small RNA molecules are controlled by highly developed regulatory pathways and fine-tuning mechanisms in both processing and target recognition [47-49]. For this reason, miRNAs are considered versatile regulators of gene expression and to date we can count roughly 60% of mRNAs as sharing one or more sequences that are evolutionarily conserved and predicted to interact with miRNAs. Bioinformatics analysis predicts that 3' UTRs of single genes are often targeted by several different miRNAs [50].

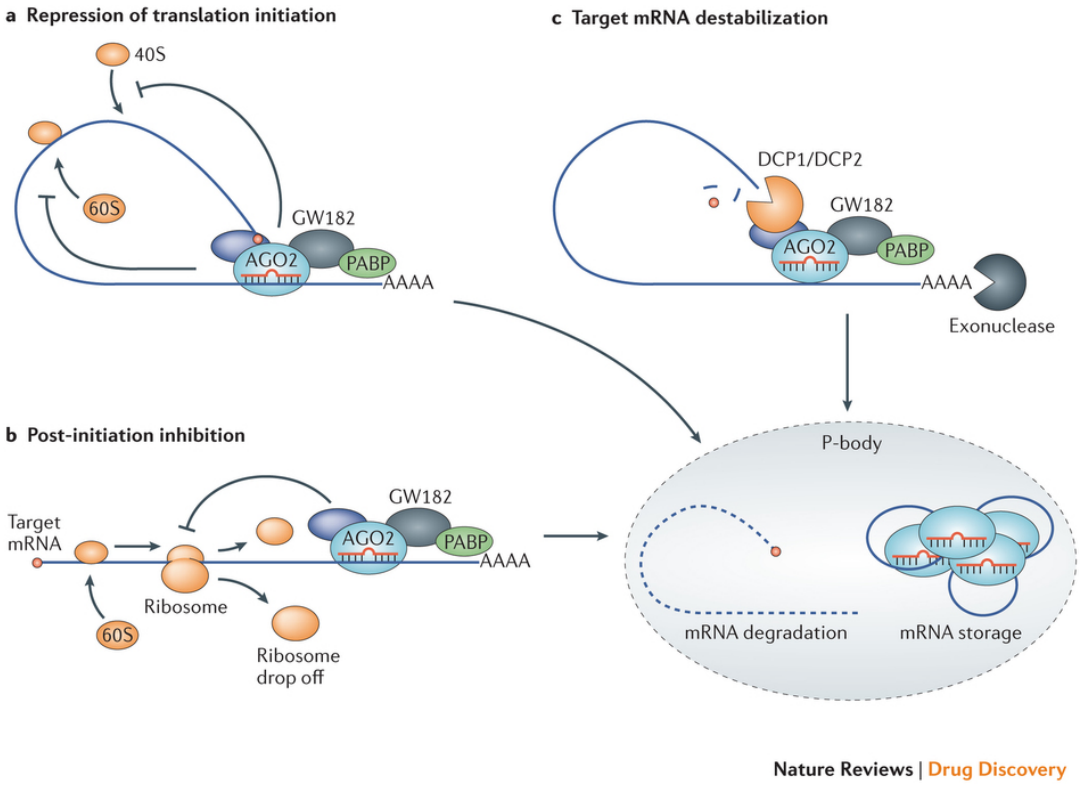

Figure 4: a | Repression of translation initiation. MicroRNA (miRNA)-mediated silencing complexes (miRISCs) inhibit the initiation of translation by affecting eukaryotic translation initiation factor 4F (eIF4F) cap recognition, 40S small ribosomal subunit recruitment and/or by inhibiting the incorporation of the 60S subunit and the formation of the 80S ribosomal complex. Some of the target mRNAs bound by the miRISC are transported into processing bodies (P-bodies) for storage and may re-enter the translation phase when induced by exogenous signals such as stress. b | Post-initiation translational repression. miRISCs may inhibit the elongation of ribosomes, causing them to drop off the mRNAs and/or facilitate the degradation of newly synthesized peptides. c | Destabilization of target mRNAs. Binding of miRISCs to target mRNAs may recruit RNA decapping and/or deadenylating enzymes that lead to mRNA destabilization. P-bodies are the key cellular organelles for the degradation and storage of targeted mRNAs. AGO2, Argonaute 2; DCP1, mRNA-decapping enzyme 1; PABP, poly(A)-binding protein. Figure and caption by Li and Rana [51].

1.6 miRNA role in development

The biological functions of miRNAs are the subject of intense study aimed to verify their involvement in various cellular processes. It has been established that miRNAs have a critical role in apoptosis, cell proliferation, resistance to stress, metabolism, and in defending the body from pathogens. MiRNAs also play an essential role in development. Models of the Dicer knockout mice provide significant evidence of the specific role of miRNAs in mammalian development. These mice, in fact, do not survive beyond the

seventh day after gastrulation and lack of pluripotent stem cells [52]. The conditional removal of the Dicer protein only in certain tissues and organs also allows to evaluate the role of miRNAs in specific contexts. This experimental approach has demonstrated the essential role of Dicer in the morphogenesis of several organs, including the lungs, limbs and muscles, and in the differentiation of T cells [53-56]. Of course, these experiments should be interpreted under the assumption that Dicer does not play any other important roles outside of the processing of miRNAs and siRNAs. However, the analysis of miRNA expression supports the assumptions made.

Currently, the role of individual miRNAs in the development and differentiation of vertebrates and invertebrates is starting to emerge. Targeted experiments have allowed, for example, to establish the importance of the miR-17-92 cluster, the removal of which causes death in mice within a few hours of birth, due to the incomplete development of the lungs and to a heart defect. The experiments also show the essential role of this cluster in the regulation of pro-apoptotic protein Bim, related to the development of B cells [57].

Other experiments prove the essential role of miRNAs in the proliferation and differentiation of stem cells. For example, the expression of miR-520H is related to the differentiation of hematopoietic stem cells [58], whereas miR-150 can direct the differentiation of megakaryocytes, bone marrow cells responsible for the production of platelets [59]. The specific expression of miR-1, miR-133 and miR-206 in muscle suggests their involvement in myogenesis [60-62], a process for which the crucial role of miR-26a has been shown, for this miRNA can regulate the EZH2 protein, a suppressor of the differentiation of skeletal muscle cells [63].

The role of miRNAs as switches for regulatory pathways, such as the mechanisms of alternative splicing, which may contribute to tissue-specificity, is also emerging. For example, the muscle-specific miR-133 is able to mute a regulatory protein of alternative splicing during the differentiation of myoblasts, in order to control the splicing of certain combinations of exons [64]. Similarly, the role of miR-124 in the development of the nervous system through the regulation of neuron-specific alternative splicing has also been demonstrated [65].

Although the regulatory mechanisms of miRNAs have not yet been fully elucidated, it the importance of these molecules in the normal development of many organs, including the heart, and their impact in many diseases, from infections to cancer, is nonetheless evident.

1.7 miRNA role in disease

Recent years have been depicted by an increase in efforts to study alterations of miRNA expression in many diseases. Evidence suggests a potential involvement of miRNAs in neurodegeneration. It has been demonstrated, for example, that patients with Alzheimer's disease are characterized by a significant under-expression of miR-107. The latter may be involved in the progression of the disease, through the regulation of BACE1, an enzyme that cuts the myeloid precursor protein, resulting in a neurotoxic amyloid peptide. The loss of miR-107 thus leads to an increase in the level of BACE1, as the bioinformatic predictions indicated and laboratory experiments have proven. This dysregulation may be one of the mechanisms responsible for the pathogenesis of Alzheimer’s disease [66].

MiRNAs may play an important role also in Parkinson's disease. A study has investigated their role in dopaminergic neurons in mammals, identifying miR-133b as specific to these

neurons, nevertheless under-expressed or even absent in the cells of Parkinson's patients [67]. This miRNA regulates the maturation and function of dopaminergic neurons through down-regulation of the expression of the transcription factor Pitx3 [67]. In addition, several studies in mammals and invertebrates suggest the involvement of miRNAs in neuro-protection, in fragile X syndrome and schizophrenia [68-70]. All these observations indicate that neurodegeneration may be the result of the alteration of various cellular pathways in which miRNAs may play a significant role.

Several studies have demonstrated an unexpected role of miRNAs in controlling various aspects of the function and dysfunction of the liver. miR-122, for example, is the most highly expressed miRNA in the liver, where it controls the response to stress by regulating the CAT-1 gene [71]. Another study has allowed the validation of targets for cellular miRNAs in the genome of HCV. Through microarray analysis performed after interferon treatment of hepatoma cell lines infected with HCV, it was shown that interferons α e β are able to inhibit the replication and infection of HCV, up-regulating the expression of many cellular miRNAs. Specifically, eight of the IFN-β-induced miRNAs (miR-1, miR-30, miR-128, miR-196, miR-296, miR-351, miR-431 and miR-448) show almost perfect complementarity of their seed to sequences of viral RNA, and the over-expression of these miRNAs reproduces the antiviral effects of IFN-β, while their suppression reduces the effects. Experiments show the direct interaction of miR-196 and miR-448 with HCV RNA. It therefore follows that mammals, through interferons, leverage on miRNAs for defense against viral infections [72].

Evidence has also been shown of the involvement of miRNAs in primary muscular disorders, which encompass various diseases, including muscular dystrophy and inflammatory and congenital myopathies. One study showed alterations in the expression

of 185 miRNAs in Duchenne muscular dystrophy, in Miyoshi myopathy and dermatomyositis. Five of them, miR-146b, miR-221, miR-155, miR-214 and miR-222, are consistently over-expressed in almost all samples analyzed, suggesting their possible involvement in a common regulatory pathway [73].

MiRNAs play a crucial role also in heart disease. Evidence describes the correlation between miR-133, which regulates protein RhoA and NELF-A / WHSC2, and hypertrophy of cardiomyocytes [74]. miR-1 is over-expressed in individuals with coronary artery disease, and the over-expression of miRNAs in the heart of rats, exacerbates the arrhythmia by silencing genes GJA1 and KCNJ2 [75]. The knockout of miR-1 can inhibit ischemic arrhythmias, suggesting a possible therapeutic application.

All these studies demonstrate the potential regulatory function of miRNAs in various cell types and tissues. MiRNAs, by modulating networks of hundreds or thousands of proteins, may be involved in the pathophysiology of many human diseases, and a single miRNA could have effects on multiple pathological pathways, due to its several targets. As shown in the previous section, miRNA targets may include genes involved in differentiation and transformation, such as transcription factors and proteins involved in the control of the cell cycle. Diseases may therefore be the result of the disruption of these pathways due to mutations in miRNA genes, in the binding sites of their targets or in the pathways that regulate their expression.

In the search for new molecular entities to be used as therapeutic tools, both miRNAs and their targets are respectively potentially targetable. Innovative therapeutic strategies may use miRNAs or anti-miRNAs as small molecules able to mimic or antagonize the action of miRNAs on multiple targets, giving rise to novel therapies for diseases that are currently difficult to treat, as it will be shown in later chapters of this thesis.

1.8 miRNA and cancer

Although studies to determine the correlation between the dysfunction of miRNAs and human diseases still have a long way to go, a large amount of data demonstrates, nonetheless, the crucial role miRNAs play in the pathogenesis of cancer.

Over the progression of the disease, cancers develop refined networks of biological events allowing them to grow and, in some cases, escape treatment. In the pathogenesis of cancer, dysregulated miRNAs may function as either tumor suppressors or oncogenes by targeting one of the six essential features of cancer progression as described by Hanahan and Weinberg: self-sufficiency in growth signals, insensitivity to anti-growth signals, apoptosis evasion, limitless replicative potential, sustained angiogenesis and tissue invasion and metastasis [76]. Experiments of gain- and loss-of-function have provided new insights into the role of miRNAs in carcinogenesis. For example, tumor suppressor miRNAs have been studied through gain-of-function approaches showing crucial roles in various cancer pathways by targeting oncoproteins such as BCL2, found regulated by miR-15a and miR-16-1 [77], MYC or RAS, both targeted by let-7 [78,79] and MCL1 (myeloid cell leukemia sequence 1) modulated by miR-29b [80]. Vice versa, to delineate the biological effects of oncogenic miRNAs, often overexpressed in cancer cells, in vitro silencing was performed using antisense oligonucleotides. For instance, a cohort of miRNAs such as miR-221, miR-222, miR-21, miR-24, miR-133, miR-17-5p, miR-19, miR-25 and miR-128 targets pro-apoptotic genes, resulting in inhibition of apoptosis. The first examples of miRNAs with oncogenic activity validated in engineered animal models were the miR-17-92 cluster and miR-155, both discovered to be over-expressed in lymph proliferative disorders, including lymphomas and leukemia [81,82]. Infection of murine hematopoietic stem cells with a retrovirus carrying the mir-17-92

cluster accelerated the development of lymphomas in Myc transgenic mice [81]. Two different studies describe the development of lymphoproliferative disease and autoimmunity in transgenic mice overexpressing miR-17-92 cluster in B cells [57,83]. Lymphocytes' higher rate of proliferation and lower rate of activation-induced cell death in these mice were partially accredited to the direct targeting of the anti-apoptotic genes Bim and Pten by the miR-17-92 cluster. Moreover, Ventura and colleagues showed that mice deficient for the miR-17-92 cluster died shortly after birth with lung hypoplasia and a ventricular septal defect [57]. Lastly, a later study showed that deletion of the whole miR-17-92 cluster reduces Myc-induced oncogenesis [84]. In contrast, miR-155 overexpression in the lymphoid compartment was sufficient to cause cancer without any other cooperative mutation or Myc expression. Costinean and colleagues developed miR-155 transgenic mice that showed polyclonal lymphoid proliferation followed by acute lymphocytic lymphoma or leukemia [85]. These data were the first to report that the dysregulation of a single miRNA can lead to malignancy. Further, it was one of the first, and still few, miRNA engineered animal models which, through knock out or transgene introduction, can provide the genetic demonstration of the causative involvement of a specific microRNA in a biological phenomenon. Regardless of notable recent studies in cancer-related microRNAs, it is unlikely that miRNAs will be found responsible for a specific phenotype by aiming at a specific target. Though many critical questions concerning microRNAs still remain to be addressed, it is largely accepted that miRNAs engage in complex interactions with the machinery that controls the transcriptome, while also simultaneously targeting multiple mRNAs. This is probably the most captivating foundation supporting the idea of using microRNAs for cancer therapeutics.

1.9 Conclusions

RNAi research is ever changing and gradually elucidating mechanisms unimaginable just a decade ago. The discovery that certain genomic regions which were previously considered as non-transcribed, instead generate a large amount of small RNAs that are actively involved in the regulation of the genome, is amazing and a true revolution in basic and applied biological research.

Although increasingly sophisticated and specific experiments are gradually revealing the subtle mechanisms underlying the biogenesis and function of small ncRNAs, many questions still need to be answered. The elucidation of such diverse variety of small RNA molecules that share many aspects but which differ in others, both evolutionarily and at the level of their biogenesis and function, still needs to be fully addressed. The existence of subclasses of these molecules, such as Mirtrons, and the reason the RNAi pathway in prokaryotes is somehow replaced by the acquired immunity provided by CRISPR molecules [86] are elements of a scenario still poorly understood. It is very plausible that new small regulatory molecules will be discovered in the near future.

The study of the involvement of miRNAs and siRNAs in regulatory pathways reveals their fundamental importance in understanding the mechanisms underlying physiopathological processes. The hypothesis that more than one half of human genes are under the control of miRNAs [20] explains their extensive involvement in many diseases, including cancer. MiRNAs can act like tumor suppressor genes as much as like oncogenes. A miRNA that regulates a tumor suppressor or a pro-apoptotic protein can act as an oncogene, promoting the inhibition of apoptosis and the enhancement of the cell cycle. Similarly, a miRNA that regulates an oncogene or anti-apoptotic protein acts as a tumor suppressor, promoting an increase in apoptosis and cell cycle arrest.

The numerous computational tools designed for the study of RNAi have proved indispensable for the understanding of the basic mechanisms as well as for the effects on the phenotype. The computational prediction of miRNA and siRNA genes and their targets, together with sophisticated data mining analyses aimed at revealing hidden correlations among regulatory RNAs, their target and the physiopathological processes in which they are involved, have been increasingly elucidating the complex molecular mechanisms in which these molecules are involed, aiding in the quest to determine the causes of many diseases and thus revealing the diagnostic and prognostic potential of these regulatory molecules for the design of novel therapeutic strategies.

Chapter 2

RNA Editing

In 1986, a process that became known as *RNA editing* was discovered by the research group of Benne Rob [87]. They found a post-transcriptional process in which mitochondrial messenger RNAs were altered by the insertion and deletion of uridine[1]. This phenomenon is explained by the fact that mitochondrial genomes of protozoa encode a small number of proteins, many of which are coded by genes which either display a disruption of the ORF or do not even have a start codon for trascription[2].

According to what the scientific community knew at the time about nucleotide modification in RNA and alternative splicing of mRNA, it was impossible to explain that mitochondrial mRNA contained insertions of one or more non-genomically encoded uridines, without any flanking consensus sequence at the site of insertion.

Despite the fact that the editing phenomenon was detected in different species at the beginning of the 1990s, real interest started around 1994. In fact, from this year on, several international conferences were organized by Harold Smith and Steve Hajduk[3], by Glenn Bjork, Ted Maden and Henri Grosjean[4], and by Paul Sloof and Rob Benne[5]. In 1993 Rob Benne wrote the first text dedicated to the theme of RNA editing [88]. In 1997, the inaugural Gordon Research Conference was dedicated to the modification and RNA

[1] *Uridine* is a nucleoside that composed by the pyrimidine base of uracil coupled with a ring of ribose. If the uracil is attached to a deoxyribose ring, you get a molecule of deoxyuridine.
[2] AUG codon encodes for methionine.
[3] 1994, Albany Conference, Rensselaerville, NY, USA.
[4] 1994, EMBO Workshop, Aussois, France.
[5] 1996, EMBO Workshop, Maastricht, The Netherlands.

editing [89]. This rapid growth has shown how the mechanisms of RNA and DNA editing are important for the biological phenomena present in the cell.

2.1 The Biological Phenomenon of RNA Editing

RNA editing is a process in which the nucleotide sequence of the transcribed RNA is altered compared to its originating genomic code. Editing is related to the insertion/deletion of nucleotides, or to base modification. The peculiarity is that the result of RNA editing is a change in the diversity and/or abundance of proteins expressed in the proteomes of organisms, in particular in their tissues or organelles.

The coordination of the editing activities is fundamental to other cellular pathways involving RNA, as, for example, transcription, processing and translation. There are different factors involved in the recognition of the RNA substrate and in the catalysis[6], such as the single enzymes involved in both the substrate identification and the catalytic activity, the macro- molecular complexes containing proteins and small RNA molecules as guides for the recognition of the substrate, along with multiple other proteins to coordinate the editing activities. When the editosome edits a nucleotide base, such as in ***A-to-I*** and ***C-to-U***, the editing factor acts in multiple sites.

In recent years it was discovered that A-to-I RNA editing can regulate the production of RNAi, thus potentially representing an important cellular mechanism in the modulation of the abundance of individual sequences within the transcriptome.

[6] *Catalysis* is the increase in the rate of a chemical reaction due to the participation of an additional substance called a **catalyst**. With a catalyst, reactions occur faster and with less energy. Because catalysts are not consumed, they are recycled. Often only tiny amounts are required.

2.2 RNA Editing in Different Organisms

This section will explain briefly how RNA editing occurs in plants, animals and viruses, in regards to its proper functioning in particular biological processes.

RNA Editing in Plants

Even if it happens rarely, the conversions C-to-U and U-to-C are the only types of RNA editing taking place in plant mitochondria and plastids. In particular, RNA editing sites are mostly found in the coding regions of the mRNA, as well as in introns and other non-translated regions [90].

Even if the precise mechanism hasn't been yet elucidated, considering the many sites requiring editing in the above organelles, some studies have suggested the involvement of gRNA and the editosome complex.

The importance of RNA editing is revealed also in the normal functioning of both the translation and respiration activities of plants [91]. RNA editing may be able to reactivate the functionality of tRNAs [92,93], as it corrects the base-pairing in these molecules [94]. Moreover, RNA editing has been connected to the production of RNA-edited proteins, embedded within a polypeptide complex involved in the respiration pathway. Finally, it is very probable that polypeptides synthesized by unedited RNAs would not work properly and would prevent the activity not only of mitochondria, but also of plastids.

RNA Editing in Animals

The process of polyadenylation (polyA) that occurs in the mitochondria of animals was the first in which the phenomenon of RNA editing was observed. The polyA is responsible for the derivation of the final 3' end in several animal mRNAs. This mRNA

downstream region is essential to complete some transcripts and ensures the correct translation of proteins in animal mitochondria.

RNA Editing in Viruses

RNA editing in viruses, such as measles, mumps, or parainfluenza, is used to give stability and to generate various proteins [95].

2.3 Editing by Deamination

The deamination is the deletion of an amino group from a molecule, resulting in the production of a molecule of ammonia[7]. The enzymes catalyzing this reaction are called *deaminases*. Moreover, for the reaction to occur, a molecule of water is required. This is the reason why it is also called *oxidative deamination*, because it oxidizes the carbon where the amino group is linked to, replacing it with a carbonyl group. In humans, deamination takes place primarily in the brain and in the liver, but it can also occur in the kidneys. This process allows the removal of the potentially harmful atoms of nitrogen that are in the amino acids. The process of deamination can occur both in the bases of deoxyribonucleotides (DNA molecule) as well as in those of ribonucleotides (RNA molecule).

[7] Ammonia is a compound of nitrogen with the chemical formula NH3. It occurs as a colorless, toxic gas with the characteristic odor.

The Deamination Process in DNA

The process of deamination occurs spontaneously. When occuring in DNA molecules, it can lead to genetic mutations, unless the damage is repaired. The most representative examples of deamination affecting DNA are described below.

Deamination of Cytosine The process of spontaneous oxidative deamination of cytosine causes the formation of uracil (see Figure 5).

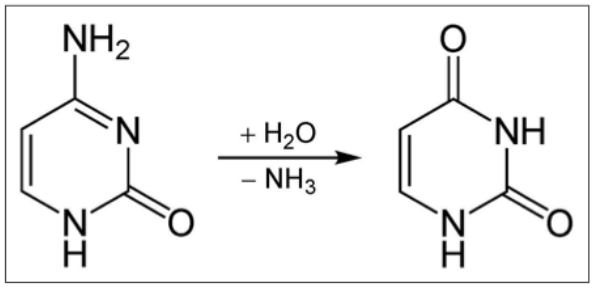

Figure 5: Spontaneous oxidative deamination of cytosine.

It can also be induced in vitro in order to distinguish between the double helix of the DNA the strand with non-methylated cytosine (thus replaced by uracil) from the one with methylated cytosine (unaltered). If the process occurs in vivo, uracil will be inserted in place of cytosine. This mismatch can be recognized and repaired by the DNA repair system; if the error is not repaired within the next DNA replication, the new molecules of synthesized DNA will contain a mutation that will no longer be repairable.

Deamination of 5-methylcytosine It is possible to find 5-methylcytosine mainly in prokaryotes. It is formed as the result of the methylation of a cytosine through an enzyme called *methyl transferase*. The deamination of this base causes the formation of thymine. In general, the DNA repair system is not able to correct this reaction, since it does not recognize the thymine as incorrect, and so the mutation persists. This defect in the repair

mechanisms contributes to the formation of rare CpG sites in eukaryotic genomes[8]. Nevertheless, there are also rare enzymes, such as *thymine-DNA glycosylase*, that are able to both recognize the mismatch deriving from this phenomenon (T-G) and fix the cytosine to thymine point mutation in one of the two daughter strands prior to passage of the replication fork [96].

Deamination of guanine The result of the deamination of guanine is the formation of the *xanthine* molecule[9]. Instead of cytosine, xanthine pairs with thymine. This process creates a mutation of post-replicative transition, in which the base pair that originally was a G-C is now transformed into an A-T base pair. This kind of mutations can be corrected by the action of the alkyladenine glycosylase enzyme during base excision repair.

Deamination of adenine The result of the deamination of adenine is the formation of the *hypoxanthine* molecule[10]. Instead of thymine, hypoxanthine is coupled selectively with cytosine. As it happens in the previous case, this process creates a mutation of post-replicative transition, so that the initial A-T base pair is transformed into the G-C base pair.

The Deamination Process in RNA

In the following sections, the most representative types of RNA editing produced by deamination of ribonucleotides base will be taken into consideration. Moreover, in Figure 6, the main effects resulting from the RNA editing process are shown.

C-to-U Editing The editing produced by the *cytidine deaminase* enzyme deaminates a base of cytidine and transforms it into a base of uridine. The *apolipoprotein*

[8] The *CpG* sites or *CG* sites are regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide in the linear sequence of bases along its length. "CpG" is shorthand for "—C—phosphate—G—", that is, cytosine and guanine separated by only one phosphate; phosphate links any two nucleosides together in DNA. The "CpG" notation is used to distinguish this linear sequence from the CG base-pairing of cytosine and guanine.
[9] *Xanthine* is a purine base. In nature it exists as methyl derivative on the various nitrogen atoms.
[10] *Hypoxanthine* is a purine derivative that occurs in nature. It is occasionally found as a component of nucleic acids

B gene (**APO B35**) in humans is an example of C-to-U editing. There are two isoforms in the human body: the *APO B100*, in the liver, and the *APO B48* synthesized exclusively in the small intestine. While the sequence of the *B100* apolipoprotein is **CAA**, when it is edited in the intestine it becomes **UAA**, which is a STOP codon. This phenomenon, however, does not occur in the liver. This concept is expressed in Figure 7.

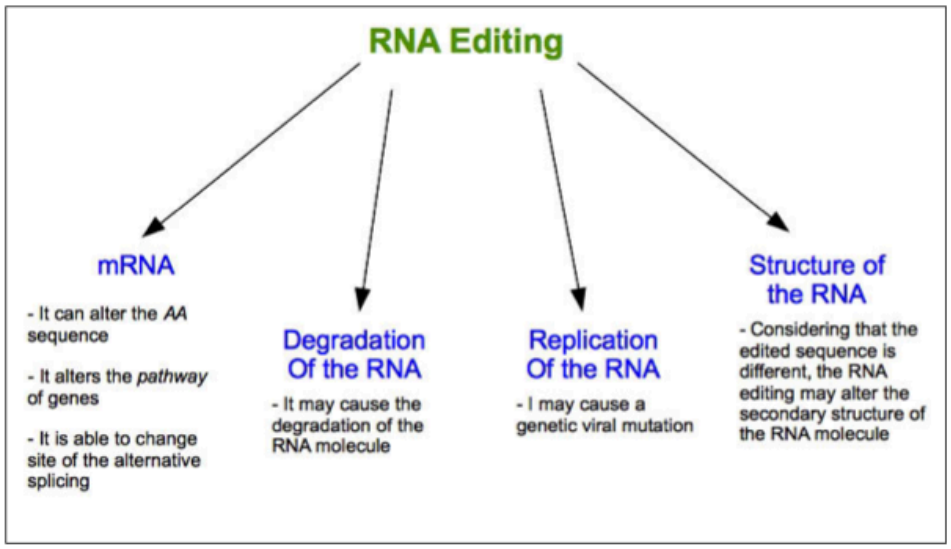

Figure 6: Possible effects caused by RNA editing.

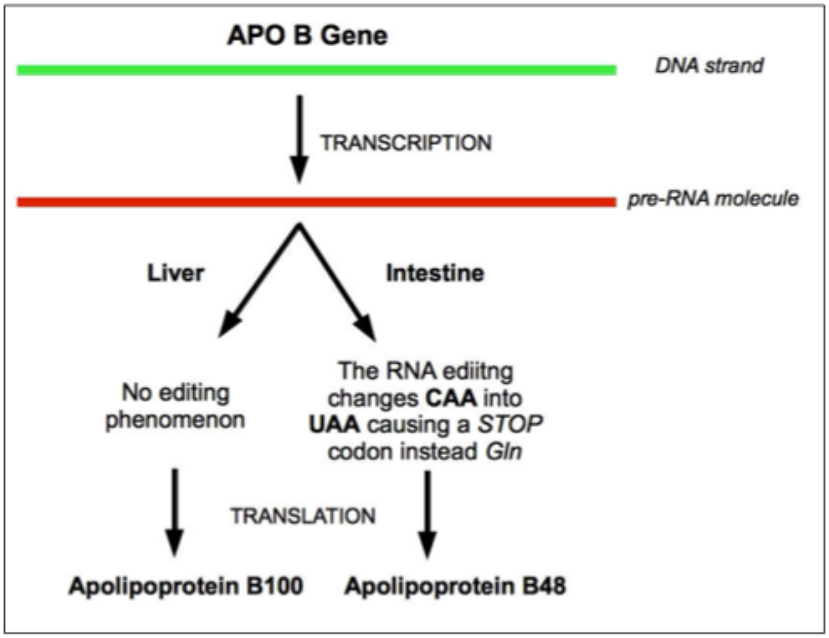

Figure 7: Example of C-to-U RNA editing in the human APO gene transcript.

A-to-I editing A-to-I RNA editing is the most studied editing phenomenon in eukaryotes and is induced by the *ADAR* family[11]. These enzymes modify a specific adenosine site into *inosine* (ergo the name *A-to-I*) in pre-mRNAs. It seems that the editing produced by ADAR occurs in all metazoans, and, specifically, it is essential for the development of mammalians. A-to-I RNA editing occurs in regions of double strand RNA (dsRNA). The action of ADAR in a double-strand region is shown in Figure 8.

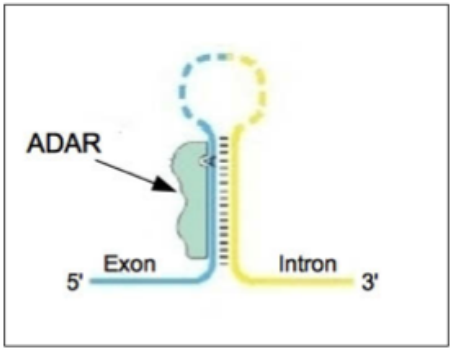

Figure 8: Example of how the ADAR acts on double-strand region.

[11] ADAR is an acronym for *adenosine deaminases acting on RNA.*

A-to-I editing can be either *specific* (if a single adenosine is edited within the dsRNA) or *promiscuous* (if the adenosines edited are up to 50%). The *specific* editing occurs within a short double-strand region (for example, those editing sites that are located in a mRNA in which the bases of the intronic sequence pair in a complementary way to the bases of the exon sequence), while the *promiscuous* editing occurs within large duplex regions (e.g. pre- or pri-miRNAs, duplexes deriving either from transgenes[12] or from viral expressions, and, finally, duplexes resulting by the pairing of repetitive regions).

The consequences resulting from the A-to-I editing may be different. This can be related to the fact that the inosine (***I***) has the same effect of guanosine (***G***) not only in the process of translation, but also in the formation of RNA secondary structure (see both Figure 9 and 10).

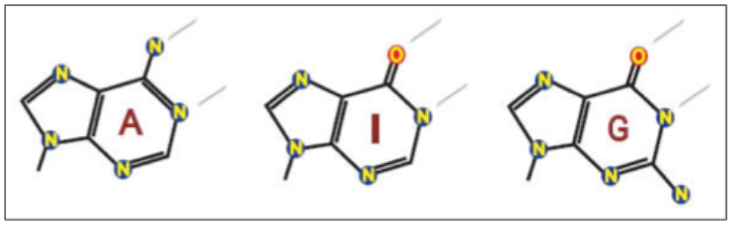

Figure 9: Molecular structures of adenine and inosine

[12] The transgene is a gene that is introduced into an organism, and this gene is alien to the entire genome of the host organism.

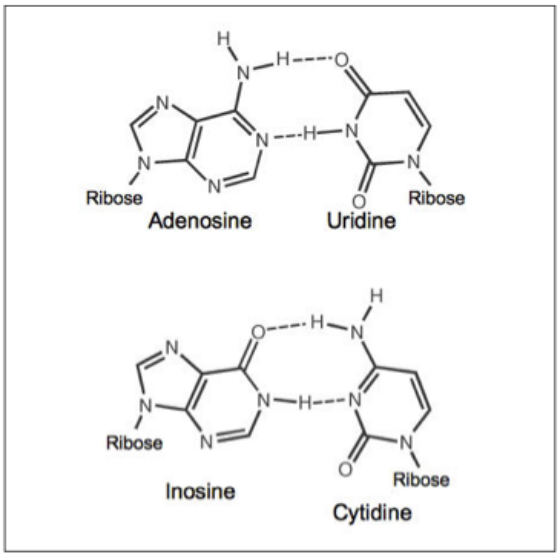

Figure 10: Inosine's interaction is similar to Guanosine's.

Among the effects of the editing phenomenon there are, for example:

- *alteration of coding capacity,*
- *alteration of the set of miRNA and siRNA targets,*
- *formation of heterocromatin,*
- *inhibition of miRNA and siRNA biogenesis,*
- *splicing alteration.*

Those can be illustrated in Figure 11.

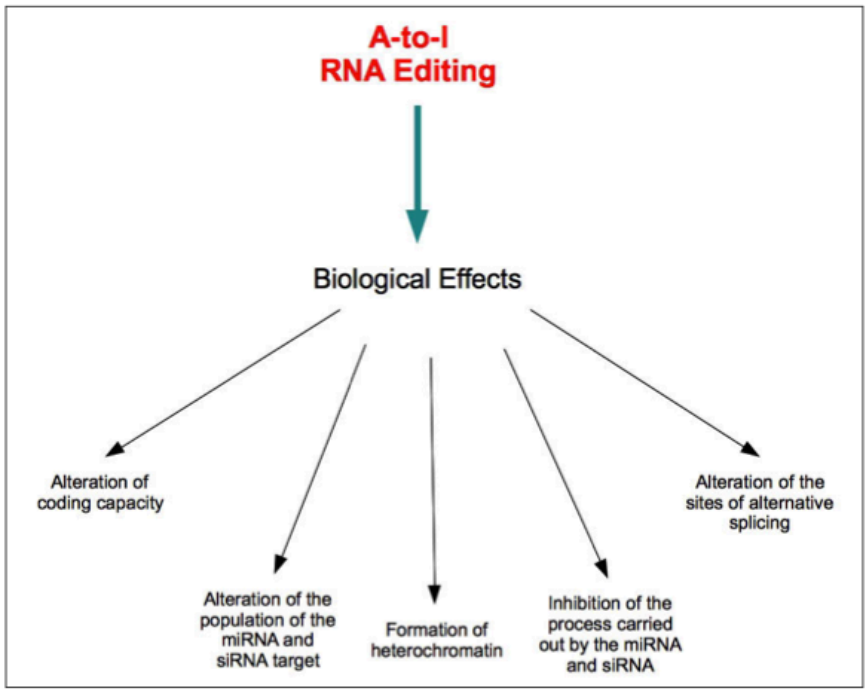

Figure 11: Main effects of A-to-I RNA editing.

2.4 A-to-I RNA editing analysis

As previously mentioned, A-to-I RNA editing can have different effects on the stability of RNA structure and RNA coding, but it is also able to affect the correct functioning of adjustment mechanisms, such as those for miRNAs and siRNAs. In this section, the history of A-to-I RNA editing, from its discovery to the present days, will be analyzed.

The origins of A-to-I RNA editing

In 1991 an A/G discrepancy was discovered between the cDNA (coding DNA) and the genomic sequences of the GluR-2 subunit[13] of mammals, due to a modification of the

[13] A protein which in humans is encoded by the gene GRIA2. It is a neurotransmitter receptor in the human brain and is activated in several physiological processes.

base at the RNA level [97]. The modification of this adenosine nucleotide converted a codon that codified glutamine in a codon encoding arginine.

Thanks to the discovery of the modification of adenosines caused by RNA editing, several other cases in transcripts of the nervous system were identified. In each of those cases, the modification of a single nucleotide, causing the substitution of an amino acid, could be connected to the change in function of the resulting protein. The simple fact that both the edited and non-edited proteins were co-expressed in the same cells, allowed scientists to realize that RNA editing was not only an important mechanism for the diversity of genetic information, but that it also had the ability to increase the complexity of both the eukaryotic trascriptome and proteome.

When editing of the mRNA encoding GluR-2 was discovered, the processes involved were unknown and two were the explanations given for the *A-to-G* change observed in the corresponding cloned cDNA. The *first* was that it was considered as the result of a process of unknown modification of the adenosine that alters the purine into another purine base equivalent to the guanosine (such as the hypoxanthine); the *second* proposed explanation was that it was caused by a mechanism involving first the removal of either the base or the entire nucleotide, and then the introduction of guanosine.

As it had already been known, the *adenosine deaminase enzyme* (also known as adenosine aminhydrolase, or **ADA**) converts the adenosine mononucleotides in *hypoxanthine* nucleotides (also called *inosines*), but it is also able to mediate the metabolism of both the eukaryotic and prokaryotic nucleotides. The ADA is an important therapeutic marker as *ADA deficiency* leads to various types of disorders of the immune system [98]. Moreover, ADA modifies adenosine mononucleotides using a mechanism of hydrolytic deamination, as illustrated in Figure 12.

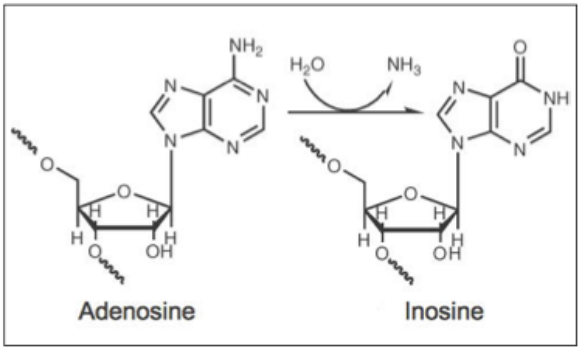

Figure 12: Transition from Adenosine to Inosine.

In addition to the modification of mononucleotides by the ADA, the modification of adenosine into inosine had been previously observed in transfer RNA (tRNA), representing a critical process for the translation of the genetic code.

Immediately before the discovery of the editing process modifying adenosines in pre-mRNAs, a new enzyme activity was discovered. It targeted adenosines incorporated in double-strand RNA molecules (dsRNA) [99,100] and through the analysis of the reaction products, it was verified that the actual molecular process was to modify adenosine into inosine. Thus, the double-strand structure of the RNA was an essential feature for the editing to occur. This is the main difference between the inner workings of the A-to-I editing and C-to-U deamination processes, involving secondary structure elements guiding the RNA modification system.

The protein responsible for this process was initially called *dsRAD*, or *Drada*, and was later renamed as *ADAR1* (in Figure 13 the tertiary structure of the protein ADAR1 is shown).

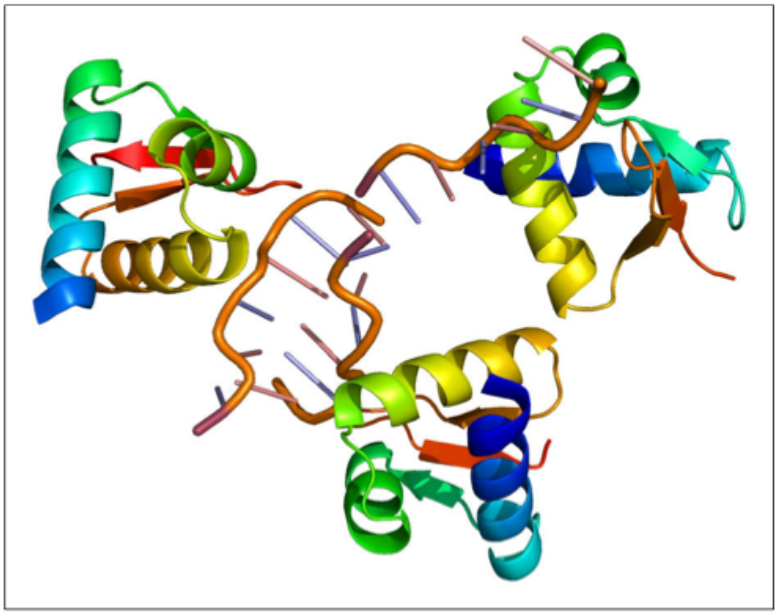

Figure 13: Tertiary structure of the ADAR1 protein.

The same protein was previously studied in several laboratories as having both a potential viral function [101]as well as being specific to the A-to-I editing activity in dsRNA of mammalian cells [102,103]. *ADAR2* and *ADAR3*, together with other similar forms in vertebrates [104], flies [105] and worms [106], were discovered after the cloning of the first ADAR (ADAR1). In Figure 14 various forms of the ADAR protein are shown.

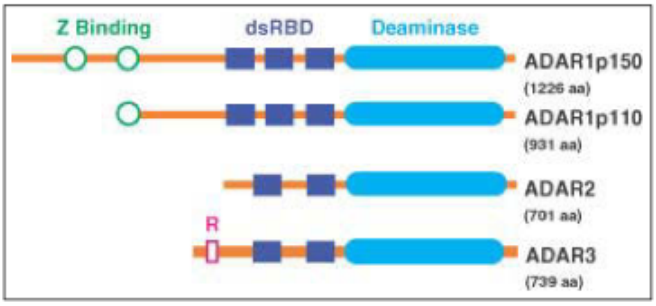

Figure 14: Comparison of ADAR proteins.

RNA Editing in miRNA-induced RNA silencing

Double-strand RNAs are the possible targets not only of the ADAR, but also of any other dsRNA-binding protein, such as those involved in the process of RNA silencing. It is easy to realize there is a connection between the processes of RNA editing and RNA silencing. There may often be a competition for the double-strand molecules between the editing machinery and the silencing one. Concerning RNA, the result may be due to the set of enzymes which first operate on the RNA molecule. According to another model, RNA editing may be a nuclear event that induces the silencing at the level of chromatin [107].

Thanks to the discovery of a few miRNAs subject to A-to-I editing [108-111], the relationship between editing and silencing is now more evident. Analyzing the characterization of the secondary structures of known ADAR targets, it can be hypothesized that miRNA molecules might undergo RNA editing. As seen above, prior to their maturation, miRNAs possess a hairpin-shaped molecular structure. Pri-miRNAs have a length of usually a few hundred nucleotides and they are first of all processed into pre-miRNA within the nucleus. The pre-miRNA, which is composed of approximately 70-90 nucleotides, is exported to the cytoplasm, where a second process generates a 20- to 22-nucleotide molecule, representing the functional mature miRNA.

The occurrence of A-to-I RNA editing in a miRNA molecule has been described first in *miRNA22* of both human and mouse [108]. The observed editing events were localized both inside and outside the *seed* of the miRNA. Depending on where the editing occurs, there can be two different consequences: either inhibiting the function of the mature miRNA or allowing the miRNA to bind to those RNA molecules to which it could not previously bind. It has been estimated that between 6% and 10% of all miRNA genes are subject to

A- to-I modification [111]. Thanks to the identification of editing events in miRNAs, it has been showed that not only miRNA transcripts are subject to post- transcriptional modification, but also that miRNA functions might not be fully elucidated solely by analyzing their genomic sequence.

Another important aspect of the editing phenomenon is illustrated in a work by Glen M. Borchert and others, published in September 2009 in the journal *Human Molecular Genetics*, in which the deamination of adenosine that occurs in human cDNA was analyzed [112]. They hypothesized that there is a relationship between the A-to-I editing events in the non-coding 3' regions (*3′ UTR*) and the *mRNA::miRNA* binding site sequence. They found meaningful correlations between A-to-I editing and the complementarity modification in miRNAs. In fact, more than 3.000 over 12.723 evaluated editing sites were found to generate complementarity to the seed sequences of a subset of human miRNAs.

In addition, the group noted editing sites within a motif 13 nucleotides long in 200 ESTs. The deamination of this motif simultaneously creates the seed matches for three microRNAs, something that would not be possible if the editing had not occurred. According to these results, one of the functions of the ADAR is to create regulatory sites for miRNAs. This means that many of them may be identified among miRNA target sites only through the examination of expressed sequences.

2.5 Conclusions

While in the past, researchers mainly focused on the study of DNA mutations in order to add information regarding the molecular pathways involved in numerous cancers, in the last decade more and more research groups have been working on the analysis of post-

transcriptional modification events, such as RNA editing. In fact, the editing machinery, occurring both in coding and non-coding RNAs, has been implicated in various human diseases [113,114]. Nowadays, it is estimated that only 1% of the mammalian genome consists in protein-coding genes, while the vast majority of the transcriptome is represented by non-coding RNAs that play a central role in the gene expression pathway, such as in transcription, translation and gene regulation [115]. A strong interest is currently growing toward understanding why and how RNA editing influences the processing of non-coding RNA.

In the past decade, surprising results have been obtained in RNA editing site discovery thanks to the application of bioinformatics approaches, as well as the contribution of RNAseq based methods in recent years. The large amount of editing sites discovered by these methodologies has given rise to the need to create public databases [116-118] containing very important information that can help to further understand the biological functions hidden within the RNA editing phenomenon. One of these databases will be discussed in the next chapter.

Chapter 3

Prediction of A-to-I edited miRNA target sites

Alterations of A-to-I editing have been associated to several human diseases, such as infections, neurological diseases and cancer [119-121]. Moreover, A-to-I editing can influence miRNA-mediated gene regulation [121]. Several cases of A-to-I editing of miRNA precursors have been reported [122,123]. This phenomenon can suppress processing by Drosha and Dicer, while the presence of inosines in the mature sequences can alter the recognition of their target sites [110]. A-to-I editing is most abundant though in the 3' untranslated regions (UTRs) of the human transcriptome [124,125]. This could affect the existing miRNA binding sites as well as generate novel binding sites [126].

The importance of RNA editing in miRNA activity suggests the need for computational tools to predict and analyze the effects of RNA editing on miRNA-mediated regulation. The present chapter will illustrate how the current knowledge on miRNA target prediction can aid in elucidating how the phenomenon of A-to-I RNA editing can influence miRNA-mediated gene regulation.

3.1 Bioinformatics Prediction of Molecular miRNA Targets

The discovery of miRNAs introduced a new paradigm in the elucidation of gene regulation systems. A primary point in understanding the functional role of miRNAs and

the complex molecular networks at the base of gene regulation, is the identification of genes regulated by miRNAs themselves. miRNA sequences are very short and are characterized by a pairing imperfection with the molecular target, contributing with complexity to the identification of mRNAs targeted by miRNAs. In recent years, the basic principles of this interaction have been extrapolated from experimental studies in order to develop numerous mathematical algorithms for the prediction in silico of the hypothetical target mRNAs. They include:

- Imperfect complementarity between the miRNA and the target's 3' UTR, with stronger bond of the 6-8 nucleotide-long seed region at the 5' rather than at the 3' end of the miRNA;
- Evolutionary conservation of the targets' 3' end sequences;
- Thermodynamic stability of the miRNA-mRNA duplex;
- Cooperativity between multiple sites in close proximity;
- Multiplicity and cooperativity of the miRNA-target interaction;
- Loss of secondary structure for the mRNA target at the miRNA binding.

Table 1 lists some of most important predictors of miRNA targets. These programs seek regions on the 3' UTR of mRNAs which display complementarity with miRNAs. There are various sequences potentially capable of binding a single microRNA, considering that not only generally the sequence complementarity between miRNAs and mRNA targets is not absolute, but also the brevity of the miRNA sequence. The programs use empirical rules to give a score to the various alignments, making use of phylogenetic prints and considering the presence or absence of more binding sites within a candidate target

mRNA. Some of the results obtained with these algorithms have been experimentally validated, allowing to significantly improve the performance of the in-silico prediction of miRNA targets.

Table 1: List of some of the most important miRNA target predictors.

Tool	Website	References
TargetScan	www.targetscan.org	49
PicTar	pictar.mdc-berlin.de	127
DIANA-microT	62.217.127.8/DianaTools/index.php?r=microtv4/index	128
miRanda	www.microrna.org	129,130
MirTarget2	mirdb.org	131
RNAhybrid	bibiserv2.cebitec.uni-bielefeld.de/rnahybrid	50

3.2 The construction of miR-EdiTar

Leveraging on the scientific evidence illustrate above, the first goal of the current doctoral thesis work consisted in the development of the miR-EdiTar knowledge base.

miR-EdiTar is a database of predicted A-to-I edited miRNA binding sites [132] and respectively: predicted miRNA binding sites that could be affected by A-to-I editing; and sites that could become miRNA binding sites as a result of A-to-I editing[14].

Prediction of A-to-I edited miRNA binding sites

The first step of this work of research is the collection of data. In fact, 1139 human 3' UTR sequences with a total of 10571 A-to-I editing sites were gathered from the first release of the *DAtabase of RNa Editing* (***DARNED***) [116]. The computational method ***miRiam*** [133] was used to predict miRNA-target interactions taking place at locations of the messenger overlapping with reported edited sites. In order to perform its predictions, miRiam exploits binding rules inferred based on experimentally validated miRNA/target

[14] miR-EdiTar is freely available online at http://microrna.osumc.edu/mireditar

pairs and the structural accessibility of the target sites. This last feature is estimated based on the local pairing probability computed by *RNAplfold* of the *Vienna RNA package* [134] with the parameters $W = 80$ (sliding window length), $L = 40$ (interactions outside the span size of 40 are not allowed) and $u = 4$ (the stretch of consecutive bases for which the probability of being unpaired is computed), as recommended in [135]. In particular, the accessibility is computed as the average probability of stretches of 4 nucleotides to be unpaired in the predicted binding site. The score of the duplex structure and its free energy are also computed. In particular, the latter is computed by using the tool *RNAduplex* from the Vienna RNA package[136]. For each affected binding sites, the accessibility, the duplex structure and the free energy are computed for both the unedited and the edited version of the duplex, in order to evaluate the effects of the editing events on the binding.

Predictions were performed on the complete set of 1922 human miRNA sequences, retrieved from *miRBase* Release 18 [137]. The edited adenosines predicted to fall in at least one miRNA binding site were 9532 out of 10571 (90%). The UTRs that had at least one edited adenosine on the seed portion of a miRNA binding site were 1102 UTRs (96.75%), while 771 (67.7%) had all their edited adenosines on the seed portion of at least one miRNA binding site. On the miRNA side, 1,664 miRNAs (86.6%) had the seed portion of at least one binding site potentially affected by editing.

The duplexes were then classified into two categories, depending on whether the edited adenosines were located on a miRNA seed portion of the binding site or not. Seed matches were classified as 6mer, 7mer-A1, 7mer-m8 and 8mer, as in [138].

Furthermore, an important aim was to find all the novel miRNA binding sites potentially generated by A-to-I editing. By changing all the edited adenosines in guanosines in the

set of the 1139 human 3' UTR sequences and repeating the above analysis, 1076 UTR sequences (94.45%) had at least one novel binding site created by editing events and 1400 miRNAs (72.8%) had at least one target site potentially created by editing.

Table 2 below summarizes the descriptive statistics:

Table 2: Overall descriptive statistics.

Edited Sites	
Edited sites on 3' UTRs	10571
Edited sites on a predicted miRNA binding site	9532
Targets	
3' UTR sequences affected by editing	1139
3' UTR sequences with at least one edited base in a predicted miRNA binding site	1102
3' UTR sequences with all their edited bases in a predicted miRNA binding site	771
3' UTR sequences with at least one novel predicted binding site created by editing	1076
miRNAs	
miRNAs reported by miRBase v.18	1922
miRNAs with predicted sites affected by editing	1664
miRNAs with predicted novel sites created by editing	1400

3.3 miR-EdiTar contents

miR-EdiTar contains a collection of predicted human miRNA binding sites in A-to-I edited 3' UTR sequences. The database contains two kinds of sites:

- *current* sites, that are those sites predicted to be miRNA binding sites but that could be affected by A-to-I editing;
- *novel* sites are those sites not predicted to be miRNA binding sites but that could become miRNA binding sites as a result of A-to-I editing.

The web site can be searched by miRNA and/or by target. Given an miRNA, the list of its predicted targets is shown in a box. When a target is selected, the corresponding interaction details are displayed on a table and available for download in comma separated value (CSV) format. The binding sites are grouped into two categories based on their type (current sites or novel sites). Several data elements are provided, such as the position of the binding site on the UTR, the seed type, the free energy of the duplex, the structural accessibility degree, the interaction score and the duplex structure. The edited bases are highlighted in bold characters and the corresponding alignment pipes are replaced with an X, indicating the potential disruption of the corresponding bond. In the case of current sites, an entry indicates whether the edited bases are located in the seed region. Moreover, the values of seed type, free energy, accessibility, interaction score and duplex structure are provided for both the edited and unedited forms of the site. Similar results can be obtained by choosing a target from the list and then selecting one of its predicted miRNAs. Check boxes can be used to filter the results visualized. In particular, users can choose to filter the interactions based on the type of predicted site (current or novel), the fact that the seed region is edited or not, the type of seed match (*6mer*, *7mer-A1*, *7mer-m8* and *8mer*) and the energy of the duplex.

Finally, miR-EdiTar is connected to *miRò*, a web environment that provides users with miRNA functional annotations inferred through their validated and predicted targets [139] which will be extensively discussed in chapter 5 of this doctoral thesis work.

3.4 Database implementation and web interface

All the data are collected and maintained up-to-date in a MySQL database (v5.1) running on an Apache server (v2.2.15). The web application was implemented in Ruby on Rails

(v2.3.5), a framework based on the MVC (Model-View-Controller) design pattern, allowing a fast development and management of the application. The queries that the database allows to perform were coded leveraging on the association mechanisms between models that the framework provides. The interface makes use of the Ajax technology to improve the usability through a fast client-side update of selections and results.

3.5 Utility and discussion

The modifications of predicted miRNA binding sites are classified into two categories, based on whether the editing events occur in the seed region or in another part of the duplex. The replacement of adenosines with inosines in the seed region can change A-U matches into G-U wobbles which are sometimes tolerated, especially in the presence of compensatory matches elsewhere in the duplex, but which have been reported to weaken the interaction or even abrogate binding [140]. This process is shown in Figure 15.

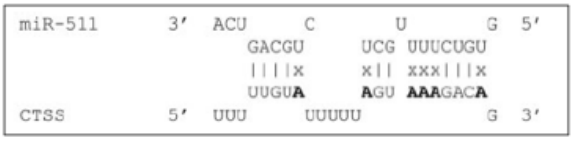

Figure 15: Predicted binding site for miR-511 on the 3' UTR of CTSS.

Editing events that occur outside of the seed binding region could also influence targeting. They might either reduce the stability of the duplex, through the introduction of G-U wobbles and mismatches, or increase it by improving the seed match or by creating new matches outside the seed area, as specified in Figure 16.

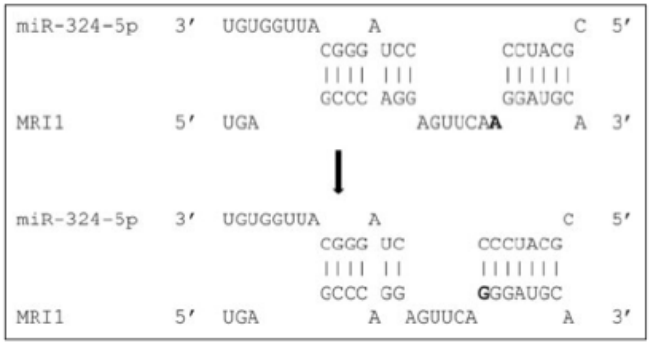

Figure 16: An edited adenosine in a potential binding site for miR-324-5p on the 3' UTR of MRI1 may improve the seed match by adding an extra CG bond and changing the seed type from 7mer-A1 to 8mer.

The presence of inosines in miRNA binding sites could also alter their secondary structure and, as a consequence, increase or reduce the chances of binding. It has been demonstrated that single nucleotide polymorphisms (SNPs) can significantly change mRNA secondary structure [141,142] and that changes in secondary structure can considerably affect the binding of miRNAs [143,144]. Therefore, it is plausible that editing events may yield similar effects (see Figure 17 and 18).

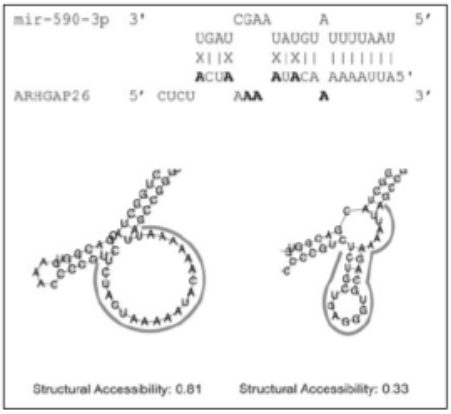

Figure 17: An example of variation of structural accessibility of predicted miRNA binding sites affected by A-to-I editing. The estimated structural accessibility of a predicted binding site for miR-590-3p in the 3' UTR of gene ARHGAP26 decreases by 40% due to editing events. The predicted interactions are shown along with the secondary structures of the unedited and edited versions of the binding sites.

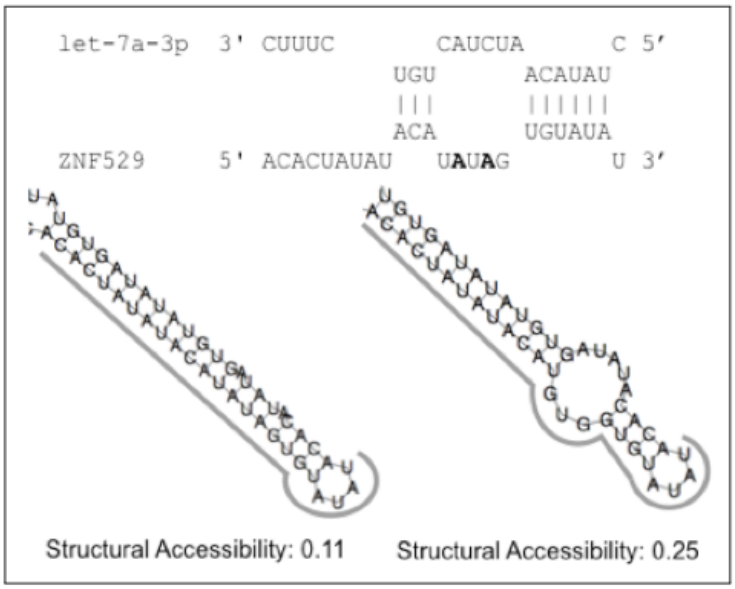

Figure 18: Example of variation of structural accessibility of predicted miRNA binding sites affected by A-to-I editing. Two edited adenosines in a non-seed area of the binding site for let-7a-3p on the 3' UTR of ZNF529 increase the estimated degree of accessibility 2.15 times. The predicted interactions are shown along with the secondary structures of the unedited and edited versions of the binding sites. Structural accessibility is computed as the average probability of stretches of 4 nucleotides to be unpaired in the predicted binding sites. Individual probabilities are calculated by the tool RNAplfold on 40 nt windows. Secondary structures of the targets are shown as computed by RNAfold on an 80 nt window encompassing the predicted binding site.

Other than affecting existing miRNA binding sites, A-to-I editing can generate novel miRNA/target interactions by either changing mature miRNA sequences or creating new sites on UTRs, as already reported by a few studies [112,123]. As a proof of principle one of the predicted novel binding sites, the gene MDM4, was validated. It is an important negative regulator of the tumor suppressor p53 [145]. The 3' UTR of MDM4 presented a cluster of 4 edited adenosines generating a novel binding site for miR-500a-3p. A fragment of the Wild Type (WT) 3' UTR of MDM4 gene containing the predicted binding site was cloned downstream of the luciferase gene on a reporter construct. A mutant version of the plasmid (MUT) mimicking the editing events was generated by replacing the adenosines reported to be edited into guanosines (Figure 19a,b).

Moreover, H460 cells (non-small cell lung carcinoma) were transfected with the luciferase reporter construct along with a precursor of miR-500a-3p or a scramble

miRNA as negative control. There was not any significant difference in the luciferase activity between cells transfected with the WT plasmid along with either the scramble miRNA or miR-500a-3p precursor. On the contrary, a 32% reduction in the luciferase activity ($P < 0.01$) was observed in cells transfected with MUT and the miR-500a-3p precursor compared to cells transfected with MUT and the scramble miRNA (see Figure 19c). This data clearly confirms that the editing process can produce new binding sites for miRNAs on specific regions of the 3'UTR of a gene.

All these hypotheses and preliminary experiments suggest a new layer of dynamic regulation in miRNA-mediated gene expression control and encourage further investigations.

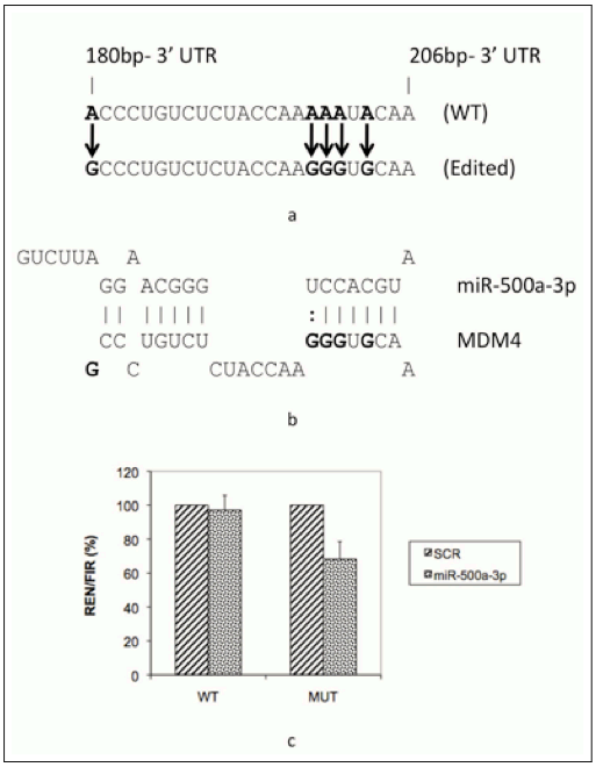

Figure 19: Experimental validation of a novel predicted site for miR-500a- 3p created by editing in the 3' UTR of MDM4. (a) A 24 nt long fragment of the 3' UTR sequence of MDM4 with 5 edited adenosines and the corresponding mutated version mimicking the editing events. (b) The predicted duplex of the miRNA/target interaction created by the editing events. (c) Renilla luciferase activity following co-transfection of a negative control miRNA (SCR) and miR-500a-3p, along with the non-edited luciferase reporter construct (WT) and its mutated version (MUT) into H460 cells. A 32% reduction in the luciferase activity ($P < 0.01$) is observed in the cells transfected with MUT and the miR-500a-3p precursor compared to the cells transfected with MUT and the negative control miRNA. No effect is observed in the cells transfected with miR-500a-3p/SCR and WT.

Chapter 4

Design of multi-site multi-target synthetic miRNAs

The discovery of small regulatory RNAs constitutes a milestone in the world of molecular biology and biomedicine. Fire and Mello's experiments on gene silencing induced by the transfection of a double-stranded RNA sequence complementary to the target gene, have earned them the Nobel Prize in Physiology or Medicine in 2006, opening new horizons on the possibility of developing new therapeutic strategies based on RNA interference. From the moment the first artificial siRNA molecules were designed primarily for gene *knock-out* experiments, research has given way to the development of *short-hairpin RNA* (**shRNA**) and *synthetic miRNAs* (or *artificial miRNAs*, **a-miRs**), which have proven to be effective in the inhibition of specific proteins, thus constituting a potential new class of "smart" drugs.

In this chapter, a brief introduction on the therapeutic potential of synthetic miRNAs will be followed by the description of ***miR-Synth***, a computational tool for the design of highly specific synthetic miRNA sequences, developed as part of this thesis work and published in 2014 in the Oxford Journal *Nucleic Acids Reasearch* [146].

4.1 Therapeutic potential of RNA interference

RNA interference (RNAi), a mechanism widely described in the first chapter of this thesis work, is a natural phenomenon taking place with the purpose of controlling and silencing gene expression, playing a crucial role in many physiological and pathological processes. The enormous power of this mechanism is now emerging as a promising therapeutic approach [147,148]. Selective gene silencing through small interfering RNAs is widely and successfully employed in functional studies and is currently being investigated as a potential tool for the treatment of various diseases, including cancer, skin diseases and viral infections. siRNA, shRNA and their optimized chemical modifications are the active silencing agents and are intended to target single mRNAs in a specific way [149].

siRNAs introduced into the cell in the form of double-stranded RNA possessing perfect complementarity with the target of interest, enter the RNAi cellular pathway directly during the loading of the RISC in the cytoplasm. *shRNA* (siRNA essentially modeled as short-hairpin) and *artificial miRNA precursors*, with the characteristic mismatch of endogenous pre-miRNAs, once introduced into the cell must first reach the nucleus. These exogenous molecules are thus capable of mediating *additional* post-transcriptional silencing, beyond normal regulation of mRNA translation, thereby reducing abnormal levels of mRNA and proteins. The partial or total silencing performed by these artificial molecules can find a wide use in various therapeutic applications. In fact, the goal of targeting multiple genes and disrupting complex signaling pathways can be reached by co-expression of multiple siRNA or shRNA which enable multiple target inhibition, along with the targeting of multiple sites on a specific gene [150]. In some diseases, for

example, the partial reduction of molecular targets not only may be sufficient to normalize the cell, but can also be preferable in order to allow the targeted proteins to continue to perform their normal functions. For other diseases, instead, the complete elimination of the target's expression may be necessary in each abnormal cell, which is not a simple task to attain. Finally, we must consider the risk of several side effects due to partial complementarity of small synthetic RNA with undesired targets. However, such effects can often be computationally predicted by simple sequence analysis.

4.1.1 RNAi and other therapeutic approaches

Two of the major advantages in the use of small RNA molecules compared to traditional drugs, are the greater simplicity and speed of design, along with the possibility to target any type of protein molecule. As with other classes of drugs, the release of these RNA molecules, in sufficient concentration, in the target tissue constitutes a problem. Some types of small molecules and protein drugs interact with membrane proteins, while siRNA must be able to enter the cell and shRNA and miRNA must also reach the nucleus.

Unlike other small molecules, siRNAs are able to inhibit the target of any class, including proteins with particular conformations not exposed to bond with other molecules, historically labeled as *non-druggable*. Target classes so far successfully inhibited using siRNA include neurotransmitters, growth factors, receptors of growth factors and transcription factors [151-154]. In addition, siRNA allows allele-specific silencing. A work shows in fact that an siRNA can silence the mutated version of gene *torsinA*, but not its wild-type [155].

4.1.2 RNAi for cancer therapy

For the therapeutic use of RNAi, it is necessary to evaluate its effectiveness in cell cultures before any *in vivo* study. Firstly, the target genes are identified, such as oncogenes, antiapoptotic genes or genes capable of promoting the development of cancer such as growth factors or their receptors. Obviously cancer-specific genes that are mutated or translocated are often selected. Several studies have demonstrated the efficient silencing of a wide variety of oncogenes such as mutated K-Ras, p53, Her2/neu and Bcr-Abl [16,156-158].

After selecting the targets and having designed specific siRNAs for each, the efficacy of these latter must be evaluated by measuring levels of the target mRNA or protein *in vitro*. Moreover, functional analyses are necessary (cell morphology, proliferation, apoptosis, etc.) in order to assess the mechanism of the anti-tumor effect. In subsequent *in vivo* experiments, it is necessary to develop the appropriate cancer model for the evaluation of the effects of siRNA on the tumor.

To investigate the use of siRNA and other small regulatory RNAs for cancer therapy, several studies have been conducted on animal models. For an efficient and lasting silencing effect, several groups have used viral vectors as *delivery* systems. *Adenoviruses* are among the most used viruses for the transport and release of siRNA. A work some years ago showed the effectiveness of the intra-tumoral injection of an adenovirus encoding an siRNA able to silence gene HIF-1, in combination with ionizing radiation [159]. Another delivery method is provided instead by the use of antibodies specific to tumor cells. Song and colleagues showed that an antibody associated with an siRNA designed to silence ErbB2 is able to release the siRNA only into breast cancer cells that express ErbB2 [160].

Finally, several ongoing and already completed RNAi-based clinical trials suggest encouraging results [161]. siRNA-mediated cleavage of a target mRNA, with a consequent reduction of protein expression level, was obtained in the first in-human phase I clinical trial in which siRNA were administered systemically to solid cancer patients [148].

The use of small regulatory RNAs as therapeutic agents is therefore promising. The main challenges are the specificity, efficacy and safety of delivery systems, as well as the design of artificial RNA able to silence target molecules in a specific manner, thus reducing side effects as much as possible.

4.2 miR-Synth: a tool for the design of highly specific synthetic miRNAs

Many diseases, such as cancer and neurological pathologies, occur as the result of multiple alterations in genes which are part of crucial cellular pathways.

Up to the present day, drug development has generally been focused on therapeutical targeting of *individual* genes or gene products. This strategy, however, has proven to be limited because the inhibition of single molecules may not be sufficient to effectively counteract disease progression and often leads to drug resistance with consequent relapse. In light of this evidence, the focus of drug therapy, as already introduced in the previous section, may need to shift from single- to *multi-target* approaches [162]. This approach is further justified by the fact that most cancers reflect a dysfunctionality in *multiple* pathways, along with an accumulation of new oncogenic mutations as the disease progresses. Thus, a valid strategy can come from targeting multiple genes involved in altered pathways rather than single genes, potentially assuring greater and more durable

therapeutic benefits [163]. An important experiment in antiviral therapy research, for example, has shown that stable expression of a single shRNA target- ing the HIV-1 Nef gene strongly inhibits viral replication, but the shRNA does not maintain such inhibition due to mutation or deletion of the nef target sequence which allows the virus to escape. Nevertheless, a delay in virus escape is observed instead in HIV-1 infected cells that were previously transduced with a double shRNA viral vector [164].

Optimizations for co-expression of siRNA have also been proposed. In a recent work, dual-targeting siRNA with two active strands were specifically designed to target distinct mRNA transcripts with complete complementarity. This resulted in easier RISC entry since only two strands, instead of four, were competing for it [165].

An alternative approach for targeting multiple genes is suggested by the way endogenous miRNA carry out their silencing role [166]. miRNAs, indeed, are naturally intended to tar- get multiple genes, often in multiple sites, due to the partial complementarity they exhibit to their targets [167]. This strategy would also enjoy the advantage that comes from involving fewer number of molecules.

All these considerations have led to the development, within the context of this thesis work, of ***miR-Synth***, a novel bioinformatics web application for the design of synthetic miRNAs able to target *multiple genes* in *multiple sites*.

Further reasons which have lead to the development of miR-Synth have been, firstly, the lack of specific computational tools for the development of synthetic *miRNAs*, and, more importantly, works such as that of Tsuda and colleagues, who have designed a miRNA able to silence gene *Gli-1* (Glioma-Associated Antigen 1) in pancreatic cancer cells [168], demonstrating the greater efficacy of these molecules compared to classic siRNA, for which there is instead a wide range of database and design tools available on the web.

miR-Synth is available at *http://microrna.osumc.edu/mir-synth*.

4.2.1 The miR-Synth algorithm and the design features

miR-Synth is a tool for the design of a-miRs for the repression of single or multiple targets. The problem of designing effective a-miRs is strictly connected to the prediction of miRNA binding sites. The main issue is that target prediction tools yield many false positives [169]. Nevertheless, the remarkable progress made in recent years has identified key features to characterize miRNA functional target sites.

In miR-Synth, well-established knowledge on miRNA targeting together with siRNA design rules and empirical observations on validated miRNA/target interactions have been combined into a pipeline which consists of three steps: (i) identification and filtering of repeated patterns; (ii) design and filtering of a-miR sequences; and (iii) scoring and ranking of the designed a-miRs.

The first step mainly relies on the concept of miRNA seed, which is the 5' region of the miRNA, centered on nucleotides 2–7 (Figure 20a). The miRNA seed is the most conserved portion of metazoan miRNAs and allows the characterization of miRNA families. The seed generally matches complementary, often conserved, canonical sites on the 3' UTRs of regulated targets [167,170]. There is evidence that the lack of perfect seed pairing in functional binding sites is, at times, balanced by the presence of centered or 3' compensatory sites. However, these cases are much less abundant than canonical sites, which represent the predominant interaction model associated with greater target repression. Among canonical sites, *7mer-m8* and *8mer* sites yield the *strongest repression*, while 6mer sites are associated with mild to very mild efficacy. In order to

achieve a significant repression of the targets, it has been chosen to consider only canonical sites, especially favoring 7mer-m8 and 8mer matches.

In order to estimate the number of human 3′ UTR sequences that share at least a common 7nt pattern, gene expression data associated with distinct diseases was collected from the Gene Expression Atlas (GEA) [171] and focused on the upregulated genes, thus mimicking a plausible scenario for the employment of a-miRs. For each disease, all the possible combinations of two and three upregulated genes were calculated, counting how many of them share at least a 7mer 3′ UTR site. Then polyA-signal motifs, homopolymer motifs and sites matching the seeds of endogenous miRNAs were filtered out. This analysis was performed on all upregulated gene pairs and triplets as detected in 83 different diseases, revealing that 97.3% of pairs and 81.32% of triplets share at least one 7mer site. On average, pairs and triplets shared about 136 and 24 7mer sites, respectively. In light of this, and considering cases in which a set of highly similar sequences is chosen for targeting, it was decided to set a maximum threshold of eight target sequences that users can provide as input to the system. This limitation makes sense, because eight is already a considerable number of targets, unlikely to be practical in most applications. These sequences are screened for repeated patterns of six or seven nucleotides (depending on user choice), which will constitute the binding sites for a-miR seeds (Figure 20b). These sites are then filtered based on user-provided specifications, e.g. a site must appear in multiple copies on the same target and/or it must be present at least once in every target. Moreover, users can also provide a list of sequences that must not be targeted. In this case, the system will remove all the sites that appear at least once in any of the provided sequences.

The second step of the algorithm consists of the actual a-miR sequence design. For each repeated pattern identified in the previous phase, an anticomplementary a-miR seed is created. The rest of the sequence is constructed by aligning the seed's binding sites and maximizing the match outside the seed region through a sequence profile technique, as depicted in Figure 1c. The a-miR sequences thus obtained will be 22 nt long.

The designed a-miRs are then filtered based on their nucleotide composition, combining well-established siRNA design rules with endogenous miRNA features. In particular, sequences with GC content out of the user's specified range (23–78% by default), or containing stretches of six or more nucleotides of the same kind, are discarded [172,173]. These particular thresholds were chosen according to what has been observed in typical endogenous miRNA nucleotide composition.

In this phase users can also choose to discard a-miR sequences sharing a seed with any endogenous miRNA. Moreover, users can enable the prediction of potential off-target genes. A filter allows the removal of those a-miRs whose seed is predicted to bind more than a user-provided maximum number of off-target genes. Alternatively, the user can request the top 10 a-miRs with the smallest number of off-target hits. This is an important feature, since a single a-miR may target even thousands of different genes. This issue will be further discussed in the Validation and Discussion sections below.

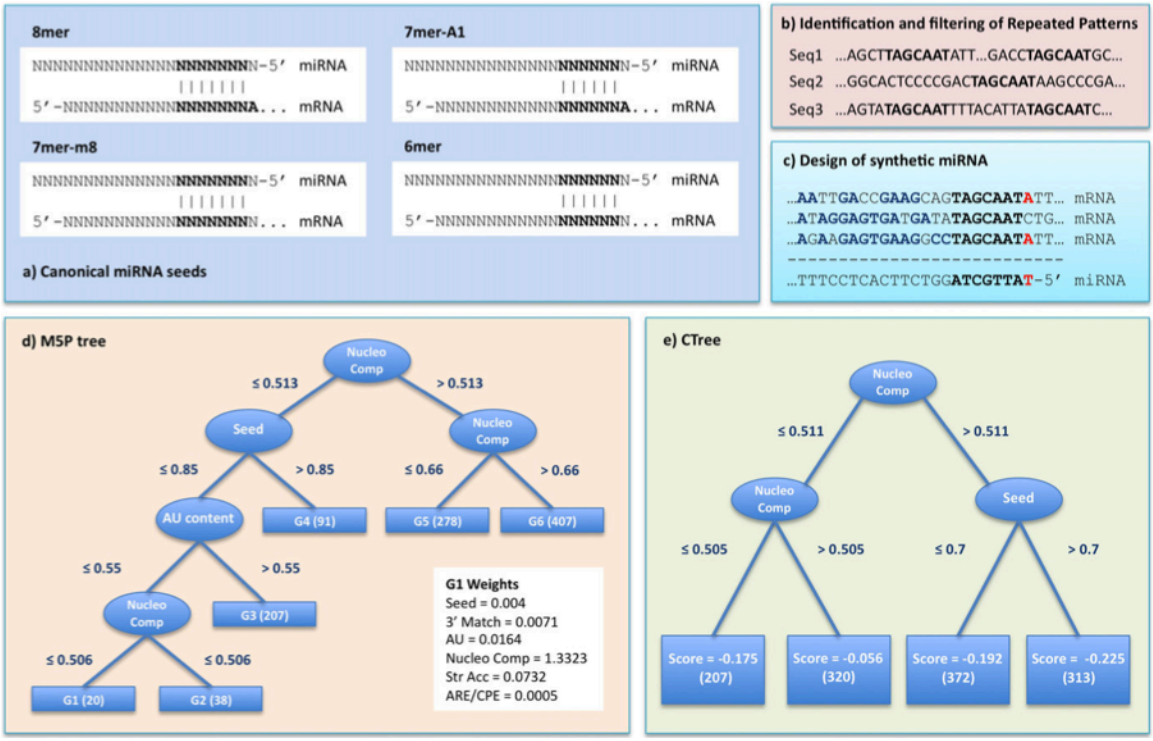

Figure 20: (a) The four different kinds of canonical miRNA seeds are depicted. They all share a 6mer core (bases 2–7). 7mer-A1 sites feature an A opposite of the first base of the miRNA, 7mer-m8 sites are full 7mers (bases 2–8) and 8mer sites are 7mer-m8 with an A opposite of the first base of the miRNA. **(b)** Input sequences are screened for repeated 6mer/7mer subsequences that will constitute the binding sites for the synthetic miRNA seeds. **(c)** Repeated patterns are used as anchors for the alignment of the binding sites of synthetic miRNAs. miRNAs are designed by maximizing complementarity to the consensus target sequence (see also Supplementary Figure S1). Target bases complementary to miRNA bases are indicated in blue and the seed match is indicated in red. **(d)** The tree generated by the learning system M5P. Six different sets of weights for the six considered features are calculated based on the values of the three discriminant features 'seed type', 'nucleotide composition' and 'AU content'. The white box contains the set of weight G1. See Supplementary Table S1 in the published paper for the complete list of weight's sets. **(e)** The tree generated by the learning system CTree. The system assigns each miRNA to one of four different score classes, based on the values of the discriminant features 'seed type' and 'nucleotide composition'.

Synthetic miRNAs for c-MET						
Rank	**ID**	**Sequence**	**Sites**	**Seed types**	**M5P score**	**CTree score**
1	60	UUUGAAACGGAGGCUGUCUAGA	3	8mer/8mer/8mer	−0.261	−0.225
2	118	UUUAUAAAGUCGAUACGUGUUU	3	8mer/8mer/8mer	−0.260	−0.225
3	181	UUCUUUCUAAGGACGGGGCCGU	2	8mer/8mer	−0.253	−0.225
4	176	UCAGUACAAAACCUUGUGGCUU	2	8mer/8mer	−0.246	−0.225
Synthetic miRNAs for EGFR						
Rank	**ID**	**Sequence**	**Sites**	**Seed types**	**M5P score**	**CTree score**
1	3	UGUGGCUUCACCUCCUGUAUCG	3	8mer/8mer/7mer-m8	−0.241	−0.225
2	106	UGUGUGACACUGCGUAAGGGGG	2	8mer/8mer	−0.238	−0.225
3	25	CAAAUGCUCGAGAGUCCGAUGU	2	8mer/7mer-m8	−0.229	−0.225
4	83	UAACAAUGCACUGGGGGCCCUG	2	8mer/7mer-m8	−0.228	−0.225
Synthetic miRNAs for c-MET and EGFR						
Rank	**ID**	**Sequence**	**Sites**	**Seed types**	**M5P score**	**CTree score**
1	141	UUCCAAUUCGAGGGGAGGUGGG	1+1	8mer/8mer	−0.262	−0.225
2	23	UCAAUUUCGGUCCCGAGUUCCA	1+1	8mer/8mer	−0.258	−0.225
3	140	UCCAAUUGGACGGGAGGUGGGU	1+1	8mer/8mer	−0.249	−0.225
4	106	UUUCAUGAGCCCUAGACUGGGG	1+1	8mer/8mer	−0.246	−0.225
5	196	UGAGUUUCUCAGCGACGGACCG	1+1	8mer/8mer	−0.241	−0.225
6	98	UUUCUUAAGCACGCCGUUGGGG	1+1	8mer/8mer	−0.239	−0.225

Figure 21: Details about the tested miRNAs

4.2.2 Scoring and ranking of a-miRs

The third step of the miR-Synth pipeline consists in the evaluation and ranking of the designed a-miRs. A scoring function based on *six* different features of validated endogenous miRNA/target interactions was developed: *seed type*, *pairing of the miRNA 3′ region*, *AU content* of the binding site and its surrounding regions, *miRNA nucleotide composition*, *structural accessibility of the binding site* and *presence of ARE* (*AU Rich Element*) and CPE (Cytoplasmic Polyadenylation Element) motifs upstream of the binding sites [144,170,174-176]. For any given a-miR, each feature is assigned a score ranging from 0 to 1 and a total repression score is calculated by combining the tree-based multiple linear regression learning system M5P with CTree [177,178].

The system has been trained on a set of publicly available gene expression profiles following the over-expression of nine individual endogenous miRNAs [170]. In particular, binding sites were predicted for each transfected miRNA on downregulated genes, then feature values were calculated. The gene expression fold change was used as a measure of the degree of repression induced by the miRNA. Thus, lower values mean stronger downregulation of the target. Only transcripts with single binding sites for the transfected miRNAs were considered, in order to reduce the chances of indirect effects.

According to the M5P tree (Figure 20d), the most *discriminant* features were the *nucleotide composition of the miRNA*, the *type of seed* and the *AU content of the binding site*.

Depending on the values of these three, six different sets of weights were assigned to all of the features. Only the seed type and the nucleotide composition of the miRNA were considered as discriminant features by CTree (Figure 20e).

These two methods are used to evaluate the designed a- miRs. In particular, a-miRs are first ranked according to the CTree score and subsequently by the M5P score. CTree splits the a-miRs into major classes, while M5P is used to rank a-miRs within each class.

We validated this scoring function by using a database of experimentally validated human miRNA/target interactions called *miRTarBase* as a test set [179]. This dataset contains 495 cases of proven direct interactions, 490 cases for which direct binding was not verified and 71 negative cases. One thousand randomly created groups with the same number of proven direct and proven negative cases were considered. For each group, the top 10 interactions, as ranked by our approach, always contained a higher number of true direct interactions compared to sets of 10 cases randomly chosen ($P < 0.0001$).

4.2.3 Validation of single-target multi-site a-miRs

The a-miR design system was validated on c-MET and EGFR, two well-known genes involved in lung cancer. This choice constitutes a good example of beneficial employment of multi-target a-miRs, given the reciprocal and complementary relationship between EGFR and c-MET in acquired resistance to kinase inhibitors in lung cancer, and the necessity of concurrent inhibition of both to further improve patient outcomes [180].

We designed two different sets of multi-site a-miRs exclusively targeting c-MET and EGFR, respectively. The system returned 111 a-miRs for c-MET and 59 a-miRs for EGFR. For each of the two genes, the focus was on the top four a-miRs as ranked by the designed scoring system. The eight a-miRs thus taken into consideration had at least two binding sites on their targets, with a predominant presence of 8mer matches. To verify direct targeting, the wild-type 3′ UTRs of c-MET and EGFR were cloned into pGL3 control vectors downstream of the luciferase open reading frame. a-miRs for c-MET and

EGFR were individually co-transfected with the c-MET and EGFR 3′ UTR constructs, respectively, in HEK-293A cells. This resulted in a significant inhibition of the luciferase activity induced by two c-MET a-miRs and three EGFR a-miRs, as compared to the negative control (Figures 22a and 23a). Moreover, western-blot and qRT-PCR assays showed that over-expression of a-miRs in HeLa cells strongly reduced the endogenous protein and mRNA levels of c-MET and EGFR as compared to control (Figures 22b and d and 23b and d), in agreement with the luciferase assay results. Expression of transfected a-miRs in HeLa transfected cells was confirmed by qRT-PCR (Figures 22c and 23c). Among the five functional a-miRs, a-miR-M-60 and a-miR-E-3 ranked first and yielded strong downregulation of c-MET and EGFR 3′ UTRs luciferase activity, respectively (Figures 22a and 23a). Hence, as further analysis, mutagenesis of a-miR-M-60 and a-miR-E-3 binding sites within the MET and EGFR 3′ UTRs was performed, abolishing the ability of these a-miRs to regulate luciferase expression, thus confirming that the binding sites are functional (Figures 22e and f and 23e and f).

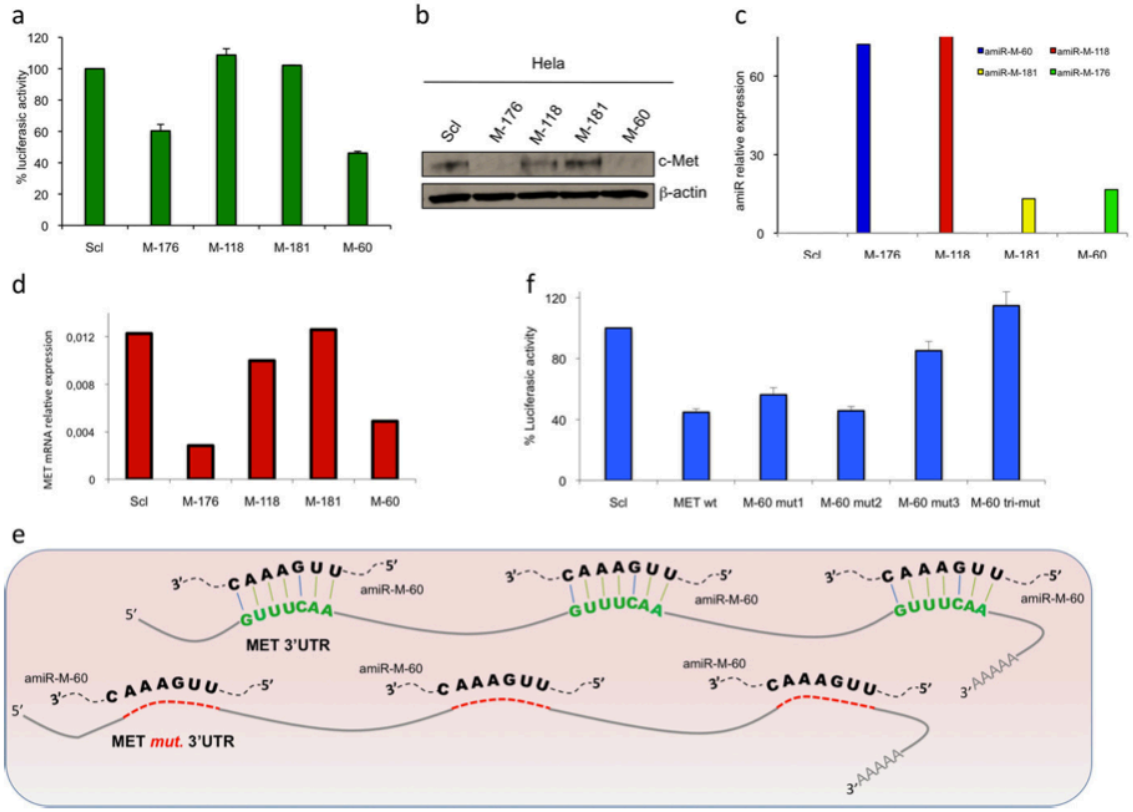

Figure 22: (a) pGL3-MET 3′ UTR construct was co-transfected with a-miRs or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, P < 0.05). (b) c-MET expression was assessed by western blot in HeLa cells transfected with a-miRs or negative control and harvested after 72 h. a-miR-M-60 and a-miR-M-176 enforced expression decreases endogenous levels of the c-MET protein. Loading control was obtained by using anti- -actin antibody. (c) qRT-PCR of the transfected a-miRs in HeLa cells. (d) qRT-PCR of the c-MET mRNA after a-miRs enforced expression in HeLa cells. (e) Representation of the c-MET 3′ UTRs binding sites for a-miR-M-60. In the figure, pairing of the seed region of a-miR-M-60 with the three c-MET binding sites is shown. The deleted binding sites are indicated in red. (f) c-MET 3′ UTR is a target of a-miR-M-60. pGL3-MET luciferase wild-type and mutated constructs were co-transfected with a-miR-M-60 or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, P < 0.05).

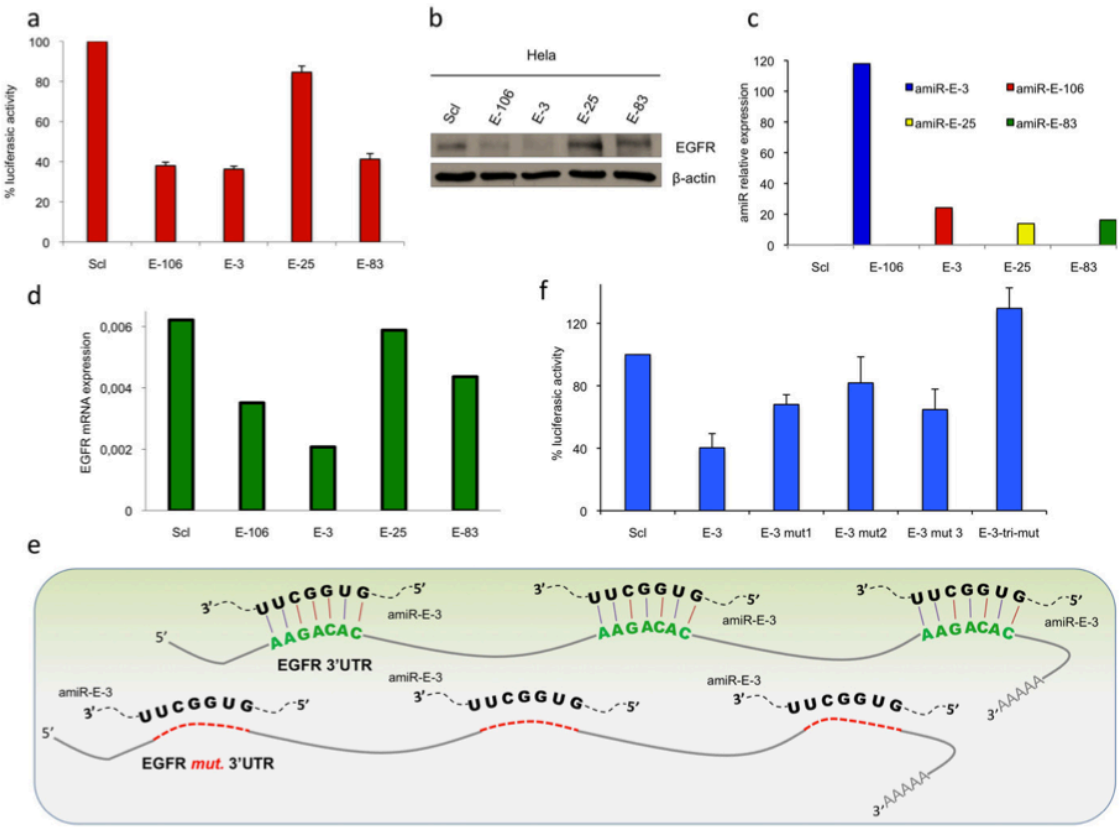

Figure 23: (a) pGL3-EGFR 3′ UTR construct was co-transfected with a-miRs or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, P < 0.05). (b) EGFR expression was assessed by western blot in HeLa cells transfected with a-miRs or negative control and harvested after 72 h. a-miR-E-3 and a-miR-E-106 enforced expression decreases endogenous levels of the EGFR protein. Loading control was obtained by using anti-actin antibody. (c) qRT-PCR of the transfected a-miRs in HeLa cells. (d) qRT-PCR of the EGFR mRNA after a-miRs enforced expression in HeLa cells. (e) Representation of the EGFR 3′ UTRs binding sites for a-miR-E-3. In the figure pairing of the seed region of a-miR-E-3 with the three EGFR binding sites is shown. The deleted binding sites are indicated in red. (f) EGFR 3′ UTR is target of a-miR-E-3. pGL3-EGFR luciferase wild-type and mutated constructs were co-transfected with a-miR-E-3 or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, P < 0.05).

4.2.4 Validation of multi-target synthetic a-miRs

A-miRs intended to target both c-MET and EGFR concurrently were subsequently designed. The algorithm returned a total of 125 a-miRs with 7mer-m8/8mer matches on the UTRs of both genes. The top six a-miRs as ranked by our scoring function were selected. All of them had one 8mer binding site on each gene.

To verify multiple direct targeting of c-MET and EGFR, the designed a-miRs were individually co-transfected with both wild-type c-MET and EGFR 3′ UTR constructs

into HEK-293A cells. a-miR-ME-196 and a-miR-ME-141 induced a significant inhibition of the luciferase activity for both constructs, while a-miR-ME-140 and a-miR-ME-106 yielded a significant repression of c-MET only, as com- pared to the negative control (Figure 24a). Moreover, over-expression of the a-miRs in HeLa cells induced a strong repression of the endogenous c-MET and EGFR proteins and mRNAs in three cases and a mild downregulation in the three remaining cases, as compared to the control (Figures 24b and d). Interestingly, although not all tested a-miRs were functional at the luciferase level, the effects on the endogenous proteins, whose repression represents our primary goal, was much stronger. This could be due to the intrinsic limitations of the luciferase assay, being based on an artificial construct. Nevertheless, out of the six tested a-miRs, a-miR-ME-196 was chosen for further investigation because of its greater downregulation at both the protein and the luciferase level (Figures 24a, b and d). The expression of a-miR-ME-196 in HeLa transfected cells was confirmed by qRT-PCR (Figure 24c). Mutagenesis of the a-miR binding site within the c-MET and EGFR 3′ UTRs eliminated its ability to regulate luciferase expression, thus confirming that the binding site is functional (Figures 24e and f). In order to further demonstrate the robustness of the miR-Synth scoring function and the additional benefits of incorporating features other than the seed match, the bottom six a-miRs designed for c-MET and EGFR were tested and it was found that three of these a-miRs yielded a mild repression of EGFR, lower than observed for the best top six a-miRs, and that none of them was able to significantly repress c-MET, despite their good seed matches (7mer/8mer).

On a final note, in order to assess the general applicability of the method, miR-Synth was additionally run on 14,325 pairs of upregulated genes in eight diseases retrieved from the GEA dataset mentioned above. miR-Synth was able to design at least an a-miR for 95%

of pairs and at least six a-miRs for 86.9% of pairs. The feature and global scores of the top six a-miRs from GEA were very comparable to the scores of the validated c-MET/EGFR top six a-miRs. In particular, this held true for features such as AU content and structural accessibility, which solely depend on the target sequence, thus confirming the results obtained with the 7mer analysis described above. However, when the off-target filter was applied, it was found that only 43% of gene pairs shared at least a 7mer with no more than 2000 off-target hits, and the percentage dropped to 5.6% when gene pairs sharing a 7mer with no more than 1000 off-targets were considered. This is an intrinsic factor of any a-miR, due to the short length of the seed region, which with no doubt requires proper consideration. The experiments showed, however, that a perfect seed match is not the only indicator of effectiveness and that other features must be taken into account. In light of all this, the off-target prediction analysis and filters provided by the system constitute a useful tool to help the user select the best a-miRs.

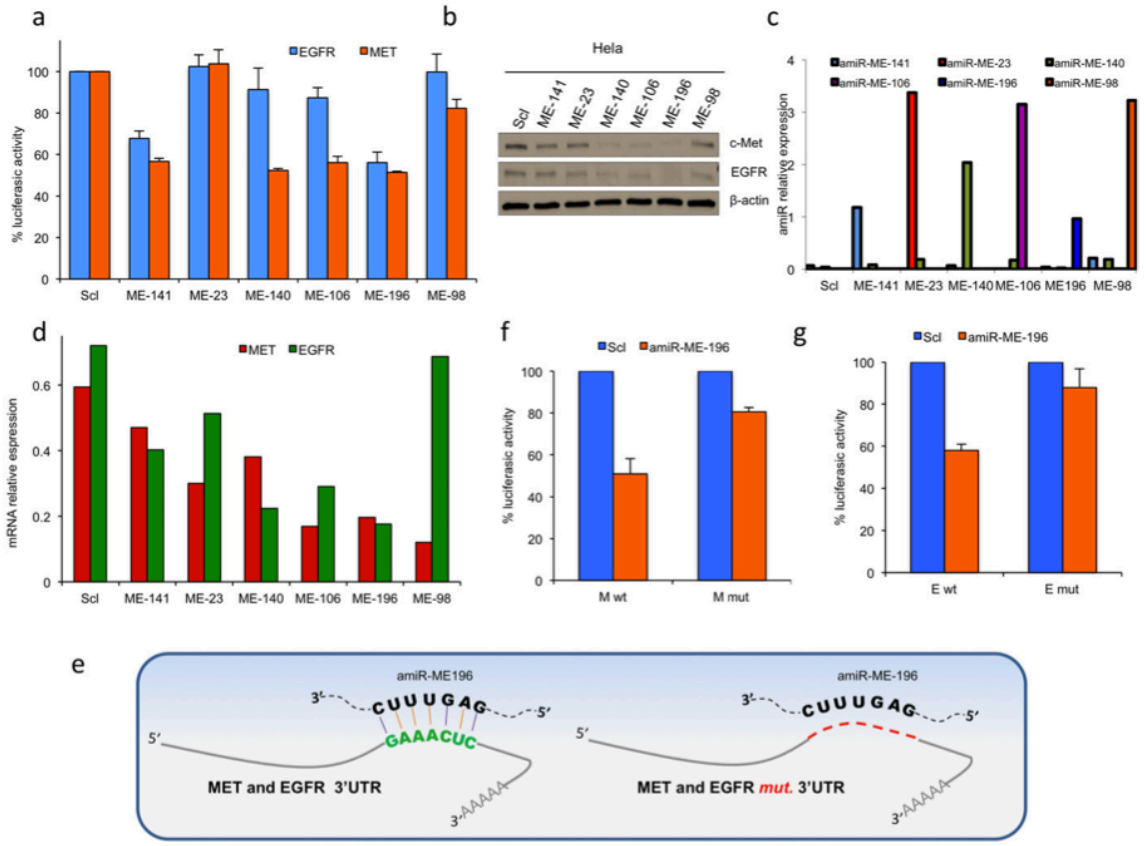

Figure 24: (a) pGL3-MET 3′ UTR and pGL3-EGFR 3′ UTR were co-transfected with a-miRs or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, P < 0.05). (b) EGFR and c-MET expression was assessed by western blot in HeLa cells transfected with a-miRs or negative control and harvested after 72 h. Loading control was obtained using anti- -actin antibody. (c) qRT-PCR of the transfected a-miRs in HeLa. (d) qRT-PCR of the c-MET and EGFR mRNA after a-miRs enforced expression in HeLa cells. (e) Representation of the c-MET and EGFR 3′ UTRs binding sites for a-miR-ME-196. In the figure, pairing of the seed region of a-miR-ME-196 with the c-MET/EGFR binding site is shown. The deleted binding site is indicated in red. (f) MET 3′ UTR is target of a-miR-ME-196. pGL3-MET luciferase wild-type and mutated constructs were co-transfected with a-miR-ME-196 or negative control in HEK-293A cells and luciferase assay was performed. (g) EGFR 3′ UTR is a target of a-miR-ME-196. pGL3-EGFR luciferase wild-type and mutated constructs were co-transfected with a-miR-ME-196 or negative control in HEK-293A cells and luciferase assay was performed (error bars: ± SEM, P < 0.05).

4.2.5 The miR-Synth web interface and output

miR-Synth is freely available for academic use through a web interface (*http://microrna.osumc.edu/mir-synth* - Figure 25). Users can provide up to eight UTR sequences or select them from a menu by their name, Refseq accession number or En-trez gene ID. Although the system was trained on human miRNAs, it allows selection of targets from other species as well, such as *mouse* and *rat*. Users can either request to

design a-miRs simultaneously targeting all of the provided sequences or to include a-miRs targeting subsets of them as well. A list of sequences (or their IDs) that must not be directly targeted by the designed a-miRs can also be provided.

In the available options users can specify the kind of seed matches allowed (6mer and/or 7mer-m8/8mer), the GC% content range (default is 23–78%) and whether the endogenous miRNA filter should be applied. Sequence masks can also be provided, in order to specify portions of the input sequences that should not be targeted. This can be a useful option when the presence of SNPs (Single Nucleotide Polymorphism) or other mutations in the targets could negatively affect a-miR binding [143,181].

Finally, users can choose to view the list of potential off-target genes, which is obtained through the computation of seed matches on the whole database of UTR sequences from the selected species.

The system is fast. For example, the design of a-miRs for a pair of targets with default parameters takes 30 seconds at most. However, given the variability in the number of input sequences and the different options that can be selected, which could substantially increase computation time, users are provided with the results page link and a results code by e-mail once the computation has completed.

The results page can be accessed either through the direct link provided in the results notification email or by inserting the unique results code. Once accessed, in the results page the user is provided with a web link to the downloadable results file in pdf format. For each individual a-miR, details about interaction features and their binding sites are given, including partial and global scores along with the list of off-target genes and the number of their potential binding sites, if requested (An example is illustrated in Figure 26).

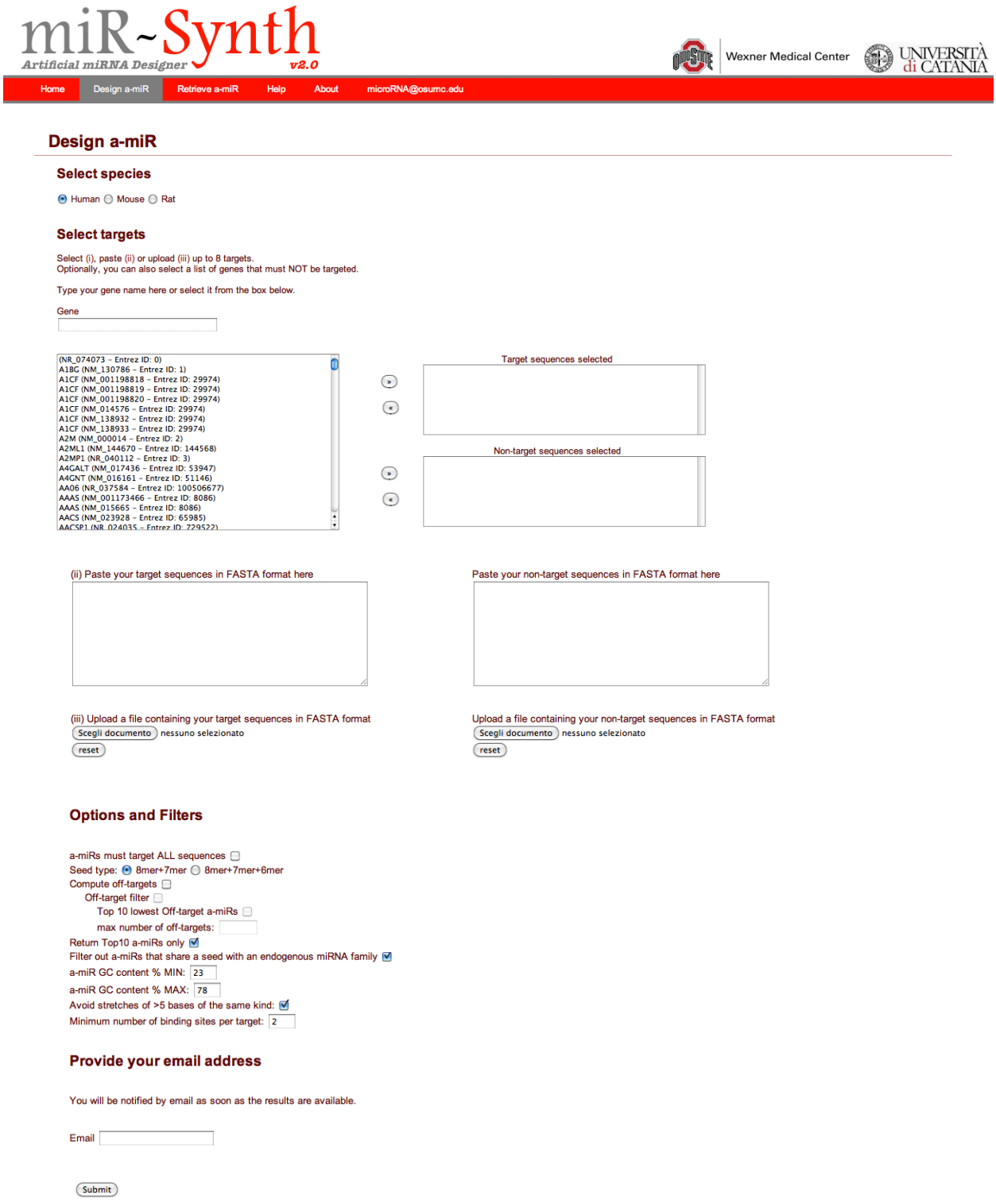

Figure 25: miR-Synth web interface.

Input parameters

Target sequence: MET – Accession number: NM_001127500 (Entrez Gene ID: 4233
Source organism: Human
Seed type: 7mer-m8 / 8mer
Minimum number of binding sites on each sequence: 2
Minimum GC content allowed: 23%
Maximum GC content allowed: 78%
Homologous miRNA filter enabled: YES
Repeats filter enabled: YES

miRNA list

1) miRNA ID: 60
Sequence: UUUGAAACGGAGGCUGUCUAGA
Total binding sites: 3

Scores

Seed score: 1.0
3' Match score: 0.4761904761904762
AU Content score: 0.5333333333333333
Nucleotide Composition score: 0.6739926739926739
Structural Accessibility score: 0.2727995679
ARE/CPE score: 0.2916666666666667

C-tree score: -0.22488171570287555
M5P score: -0.26149350969291935

Binding sites details

Position: 413
Type: 8mer
Alignment:

```
         GUC   G          U
3'-AGAUCU   GGA  GCAAAGUU -5' miRNA
   ||||||   |||  ||||||||
5'-UUUAGA   UCU  UGUUUCAA -3' mRNA
         AU    AG         A
```

Free energy: -13.1 Kcal/mol

Position: 700
Type: 8mer
Alignment:

```
   AGAUCUG       AGG        U
3'-          UCGG   CAAAGUU -5' miRNA
             ||||   |||||||
5'-          GGCU   GUUUCAA -3' mRNA
   UGUUGCCAA    G          A
```

Free energy: -11.6 Kcal/mol

Position: 848
Type: 8mer
Alignment:

```
   AGAUC       AGG        U
3'-      UGUCGG   CAAAGUU -5' miRNA
         ||||||   |||||||
5'-      ACAGUC   GUUUCAA -3' mRNA
   GCAAU       AAA        A
```

Free energy: -15.1 Kcal/mol

==========

Figure 26: Example of section of the results pdf provided by the miR-Synth system.

4.3 Discussion

Recent progress in the development of increasingly efficient carriers for the intracellular delivery of small RNAs, such as nanoparticles and viral systems, has made the establishment of therapeutics based on RNAi imaginable [182-184]. Moreover, new strategies for oral delivery of antisense nucleotides and recent findings suggesting that exogenous miRNA, such as those of plant origin, simply introduced through food in-take could be active and functional in recipient cells, opens a new scenario in which RNAi could constitute an appealing and concrete therapeutical tool for cancer, viral infections and other diseases caused or progressively maintained by the over-expression of multiple genes [185-187].

miR-Synth is a new computational method for the design of synthetic miRNA high specificity, developed as part of this thesis. Given one or more mRNA target sequences, miR-Synth designs and returns artificial miRNA sequences which can potentially bind and repress them, according to the mode of action of endogenous miRNA. Among the main requirements of an effective miRNA taken into account in the development of miR-Synth, special emphasis is on: the miRNA's ability to bind to the target in multiple structurally accessible sites; the satisfaction of empirical rules for pairing, inferred from experimentally validated miRNA/target pairs; the decrease of side effects due to interactions with unwanted off-target genes.

Although rules for efficient siRNA and shRNA design are nowadays well established, sequence design methodologies can nevertheless be further improved, especially to reduce off-target effects.

siRNAs are designed to regulate specific targets through perfect complementarity, but evidence shows that the presence of one or more perfect matches in 3' UTR sequences

with the siRNA seed region is associated to considerable off-target effects and represents a widespread and unintended consequence of siRNA-mediated silencing [188,189]. This phenomenon, which reflects the natural behavior of *miRNAs*, suggests a possible approach for designing *fewer* molecules that may reduce the expression of *many* targets. In fact, the experiments run to validate miR-Synth clearly show that a single a-miR may be able to repress at least two unrelated genes at the same time, while it may likely take a pool of different siRNAs/shRNAs to obtain the significant inhibition of even a single gene only. It is very important to point out that, in principle, *there is no difference between a single multi-target a-miR and a single-target siRNA in terms of basic seed-based off-target effects*. Any very short nucleotide sequence, such as a 7mer, is likely to appear in a substantial number of UTRs. Unsurprisingly therefore, an in silico test confirmed that double-targeting a-miRs are likely to have fewer off-targets than pairs of single-targeting siRNAs (see Figure 27). This indicates a substantial advantage in the employment of a-miRs in place of siRNAs/shRNAs.

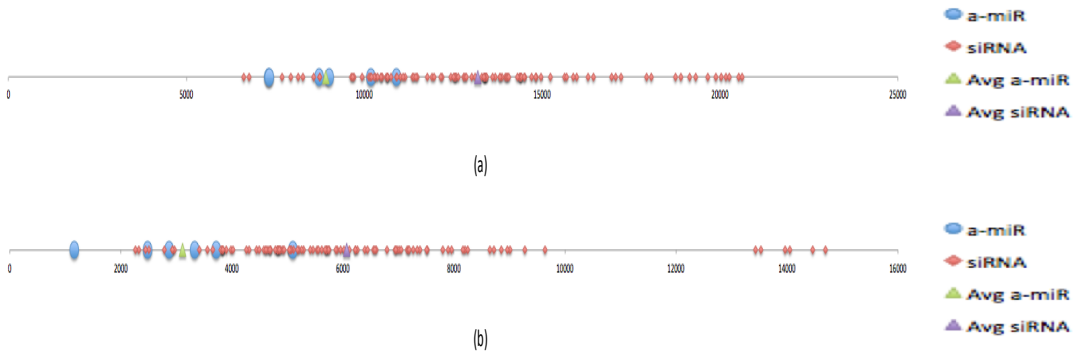

Figure 27: In silico analysis of off-target effects. Distribution of off-target numbers for double-target a-miRs and pairs of single-target siRNAs based on (a)6mers and (b)7mers.

4.4 Conclusions

This chapter introduces the concept of synthetic miRNA as a potential therapeutic agent and examples of applications that demonstrate the effectiveness of its employment were shown. The tool developed as part of this thesis, and described in the previous paragraph, miR-Synth, is currently the only tool of its kind. The miR-Synth pipeline allows the rational design of synthetic miRNAs by taking multiple factors into consideration. It integrates current knowledge regarding miRNA/target interaction and features with simple yet powerful options which allow, for example, to investigate off-target effects and design molecules virtually not affected by SNPs and other polymorphisms.

miR-Synth has been tested in silico on a set of genes whose overexpression is related to diseases, including BCL2, NRAS, NFKB, cIAP1 and AKT1. The corresponding miRNA/target duplexes satisfy the empirical rules inferred from endogenous duplexes, confirming the plausibility of the proposed method.

Finally, the biological validation of miR-Synth was performed at the Laboratory of Dr. Carlo M. Croce, at The Ohio State University, USA, where single- and multi-target synthetic miRNAs designed by the system were proven to effectively downregulate the expression of EGFR and c-MET, two well-known genes involved in lung cancer.

The work was published in 2014 in the Oxford Journal *Nucleic Acids Reasearch* [146].

Future developments include the refinement of the design process together with improved analysis of miRNA/target interactions in order to better understand the causal connection between the targeting features and the degree of downregulation, and thus improve the selection of effective molecules. Additionally, the integration of information from Gene Ontology may assist the user in selecting the most suitable a-miR for a particular application, in terms of biological functions and potential side effects.

Chapter 5

Design and development of an extended knowledge base for miRNA functional annotation

This chapter presents a web system for the functional annotation of miRNAs, developed in its second version as part of this thesis, called ***miRò***. The system is a real knowledge base, equipped with a web interface, which provides users with miRNA/phenotype associations in humans. miRò 2 integrates data from various online sources, such as databases of miRNAs, ontologies, diseases and genes, in a single environment with a flexible and intuitive query system with data mining capabilities. The main objective of miRò is to establish a *unified resource for miRNA research* through the implementation of a knowledge base that allows non-trivial analysis through sophisticated mining techniques, along with the introduction of a new level of association between genes and phenotypes, inferred from miRNA annotations. The knowledge base has been extended in terms of types of data and functionalities as compared to its first version published in 2009 [139], namely, the introduction of pathway, cellular component and genomic elements data, direct miRNA-disease associations integrated from several datasets, extracellular circulation info for miRNAs and a functional enrichment analysis for sets of miRNAs.

5.1 miRNA functional annotation

As already discussed in previous chapters, the abundance of miRNAs identified in many species and the large number of genes subject to post-transcriptional regulation imply a significant involvement of these molecules in many biological processes. The presence of binding sites for a single miRNA on muntiple transcripts, the co-regulation of the same transcript by multiple miRNAs and the existence of negative *feedback loops*, according to which a miRNA can be regulated by its own targets, reveal the great complexity of the mechanisms of miRNA-based regulation and their inevitable involvement in the pathogenesis of many diseases.

To date, several studies have documented correlations, more or less significant, between one or more miRNAs and physiological and pathological processes, and the precise association between miRNA and phenotype has been demonstrated for a growing, though yet still limited, number of cases. Much more, however, is known about genes. For example, the database *Gene Ontology* (**GO**) provides annotations on the processes and functions in which genes are involved. In addition, there is an extensive literature documenting the role of genes in disease. It is therefore possible to annotate miRNA with the information associated to their predicted or validated targets. For example, the correlation between the under-expression of miR-15a and miR-16 and over-expression of the anti-apoptotic gene BCL-2 in patients of B-cell chronic lymphocytic leukemia (CLL), allows to functionally associate these miRNAs with apoptosis and CLL [77].

A common approach in the study of diseases and biological processes that involve miRNAs, is that which requires the extraction of data from several independent sources, such as databases of miRNA/target predictions, functional annotations of genes,

expression profiles and biomedical literature. Hence, it is important to have systems that integrate data from heterogeneous sources in unique, extensible and updatable environments. Such systems should also be provided with data mining algorithms able to infer new knowledge from the data collected.

miRò is a web knowledge base that allows the execution of simple searches and sophisticated data mining queries. The main objective of the system is to provide users with a comprehensive environment of powerful tools for the discovery of non-trivial associations between heterogeneous data and therefore help the user generate new hypotheses that can further miRNA research.

5.2 The miRò2 knowledge base system

miRò 2 is a web based knowledge base for miRNA research. The original idea of developing a system so complex and complete as miRò came forth from the need to have a tool able to highlight relationships between genes, processes, functions and diseases at the miRNA level. Currently, there's a plethora of web resources which provide users (doctors, biologists, researchers, ...) with a growing amount of information regarding, for instance, miRNA-target interactions as well as relationships between genes, diseases, pathways and ontologies.

miRò 2 integrates heterogenenous miRNA-related data in order to provide a unique access point and, more importantly, to allow the inference of new knowledge regarding miRNA effects on the phenotype based on non-trivial associations between different types of data through the execution of simple searches as well as sophisticated data

mining queries.

Within the context of this thesis work, miRò 2 was developed with the goal to provide a substantial update and extension of the previous version, in terms of both data integration and query facilities. Figure 28 provides a schematic overview of the system. miRNAs are linked to pathways, diseases, processes, functions and cellular components through their experimentally validated or computationally predicted target genes. Each miRNA basically inherits the annotations of its target genes, as it may be functionally associated to the corresponding terms. This was the key concept of the first version of miRò, which has then been followed and reinterpreted by other tools as well (miRGator, FAME, DIANA-mirExTra, DIANA mirPath, miTEA, miRSystem,..). Gene annotations have been extended in order to include pathways and cellular components, which were not present in the first release. Statistical tests were implemented for the evaluation of miRNA functional enrichment, in order to identify the most significant associations, as a single miRNA-target connection is not always a sufficient indication of miRNA functional specificity.

Data on the genomic distribution of CpG islands, translocation breakpoints, fragile sites and repetitive sequences, was also integrated in order to implement two layers of miRNA annotation, one based on direct associations, i.e. miRNA genes overlapping fragile sites, breakpoints, etc., and one based on indirect associations, i.e. miRNAs targeting genes located on fragile sites, breakpoints, etc. Additionally, data on direct proven miRNA-disease associations and miRNA expression in different disease conditions from different sources was included and linked to mRNA expression, where possible, in order to provide support to target predictions. Finally, miRNAs were annotated with information

about their presence in extracellular body fluids.

In the next sections details about data integration, the update process, the query system and the web interface will be provided. The usefulness of miRo' will also be discussed through typical use cases and in silico validation tests.

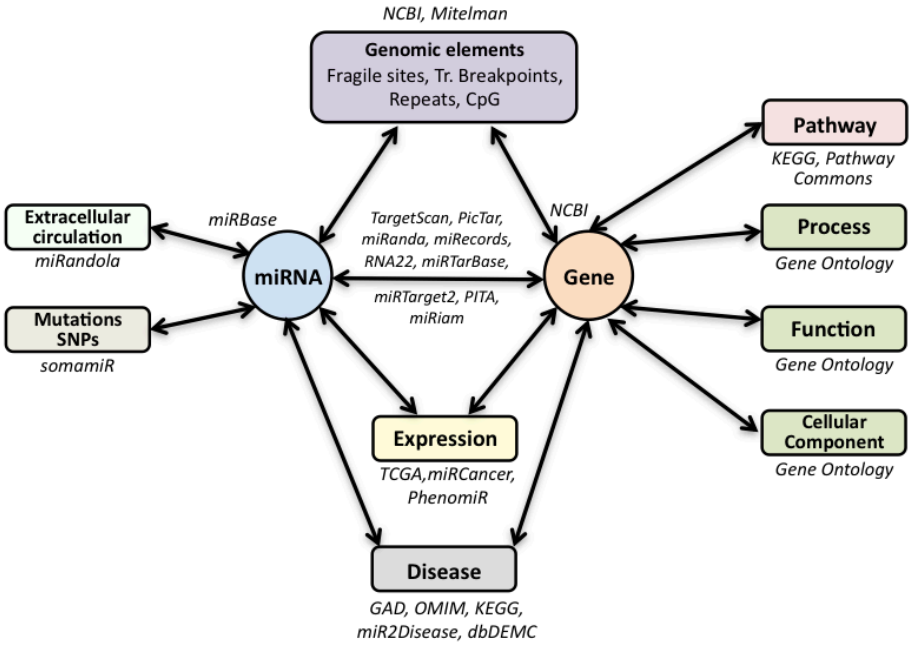

Figure 28: The miRò v.2 system schema

5.2.1 Data retrieval

All data are collected appropriately in a unified MySQL database through a semi-automated system within the context of another data repository application, *BioXml-Builder*, specifically developed for the remote retrieval of heterogeneous biological data which is in turn provided to users as converted in the standard XML format for easier management and exchange. BioXML-Builder's semi-automated data retrieval system

consists of a main module to perform administrative tasks, an update module, and a collection of parsers for importing data from heterogeneous file formats, such as csv (comma separated values), xls (microsoft excel), fasta (sequence data) and generic flat files (structured and un-structured text documents). Once the system has been configured with the URLs of the data sources and the grammar specifications, it automatically downloads and parses the data, then stores it in a sources repository (Figure 29). Then, for each source, the features of interests are extracted and imported into the miRò database in a specific order to ensure data integrity and coherence. The update module, which is run every six months, is responsible for the retrieval of new data from all the sources, should it have become available. Most sources are updated automatically, as opposed to a few of them which require manual retrieval, since they are not immediately/completely available and/or need prior processing (e.g. fragile sites or TCGA data).

In the next sections the different data sources will be described, providing details about their integration in miRò.

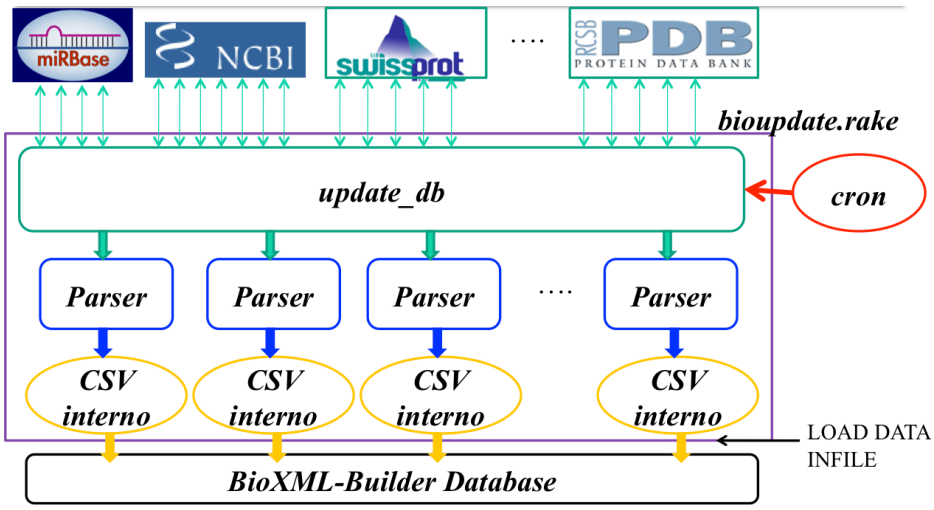

Figure 29: The BioXML-Builder remote data retrieval system.

5.2.2 Data sources and integration

The following paragraphs describe the different data sources by category.

miRNAs and their target: experimental validation and prediction consensus

Basic miRNA data was retrieved from miRBase, the primary online repository for miRNA sequences [190]. Genomic coordinates, sequence and structure of miRNA precursors were imported, as well as their mature sequences, families and cluster information. The current miRBase release (miRBase 21, June 2014) contains 1,881 human precursors encoding 2,588 mature sequences. Data on experimentally validated target genes were imported from two major archives, miRTarbase and miRecords. In particular, 2,492 miRNA-target pairs were retrieved exclusively from miRTarbase, 356 pairs from miRecords and 970 pairs reported from both, for a total of 3,818 miRNA-target interactions. A substantial fraction of targets, however, are from computational predictions, imported from six well-established tools which make their data easily available for download, in order to provide users with consensus prediction: Targetscan, miRanda, PicTar, miRTarget2, PITA and RNA22 [47,49,144,191-193]. Although it is not necessarily guarantee of a true interaction, the consensus prediction may be a useful tool for ranking and identifying the most probable targets for a miRNA, since the tools are based on different prediction strategies and features.

Unfortunately, miRNA target databases are not regularly updated along with the semi-annual releases of miRBase, thus a substantial number of miRNAs had no predicted targets. In order to overcome this issue, target prediction was performed for the complete current miRNA collection through *miRiam*, a target prediction tool published by Prof.

Alfredo Ferro's bioinformatics group in 2010, updated with the new scoring function described and applied for the miR-Synth system illustrated in the previous chapter [133]. For miRNA-target pairs predicted with miRiam, details on the interactions are given, including the predicted structure of the duplex, the binding features scores and the total repression scores. For the sake of completeness and reliability, predictions will be periodically computed as soon as new miRBase data is available, in order to always provide users with complete up-to-date miRNA annotations. Altogether, the database contains 12,970,809 miRNA-target interactions.

Finally, data such as official symbols and aliases, genomic coordinates and transcript sequences was imported for 20,092 genes that are targets of at least one miRNA, from the *NCBI Gene* and *Refseq* databases (National Center for Biotechnology Information - http://www.ncbi.nlm.nih.gov/gene/; http://www.ncbi.nlm.nih.gov/refseq/).

Extending gene annotations to miRNA: Pathways and Gene Ontology

We imported gene annotations for the three categories provided by *Gene Ontology* (***GO***): Biological Processes, Molecular Functions and Cellular Components [194]. A *biological process* is defined as a collection of molecular events with a defined beginning and end (e.g. Apoptosis), while *molecular function* refers to the specific "abilities" of a gene (e.g. DNA Binding). The *cellular component* ontology, instead, describes cellular locations in which gene products are found (e.g. Nucleus), or which gene products are subcomponents of (e.g. Ribosome). A total of 25,536 processes, 10,458 functions and 3,317 components was collected. Only GO terms associated to at least one gene were considered in the database. While broad terms such as "signal transduction", "protein binding" or "cytoplasm" were associated to thousands of genes, 4,166 terms associated to

a single gene each were found.

In the same way, *gene-pathway* associations were retrieved from two major resources, *KEGG* and *Pathway Commons*. Although many pathways may correspond to specific GO terms, they are not equivalent to biological processes. Signaling and metabolic pathways, indeed, consist of series of biochemical reactions that occur within a cell in order to carry out a specific cell function or modify an initial molecule to form another product. A total of 86 pathways from KEGG and 1,639 pathways from Pathway Commons were collected. As for Gene Ontology, a few large pathways containing thousands of genes, and 79 associated to one gene each were found.

miRNA-disease associations and extracellular circulation

We implemented two layers of *miRNA-disease* associations. The first layer is based on the experimentally verified deregulation of miRNAs in various human diseases. Data from five different manually curated sources was collected: *miR2Disease*, *PhenomiR*, *miRCancer*, the *Database of Differentially Expressed miRNAs in Human Cancers* (***dbDEMC***) and the *Human MicroRNA Disease Database* (***HMDD***) [195-199]. They all provide associations of miRNA and disease retrieved from the literature, with details about the miRNA deregulation (i.e. up or down) and the detection method (e.g. microarray, RT-PCR, …). While miR2Disease, PhenomiR and dbDEMC are general disease resources, miRCancer and dbDEMC are specific to cancer. A total of 32,324 different miRNA-disease associations were imported.

The second layer of miRNA-disease association is based on the annotation of miRNA targets. The miRò dataset was extended based on the *Genetic Association Database*

(***GAD***) [200] with data extracted from the *Online Mendelian Inheritance in Man* (***OMIM*** - http://www.ncbi.nlm.nih.gov/omim) and *KEGG Disease* (http://www.genome.jp/kegg/disease/). These three databases provide manually curated data on human complex diseases and genetic phenotypes. The disease entries retrieved from KEGG were 1,286, while 3,442 were retrieved from GAD and 21,395 from OMIM, for a total of 50,240 gene-disease associations. As with GO terms and pathways, miRNAs are linked to diseases through their validated and predicted targets.

The integration of gene-disease data was a rather challenging task. While OMIM and KEGG have implemented a more rigorous classification and identification system for disease terms, GAD presented more variety, redundancy and ambiguity in the definition of its associations. For example, six different entries correlating the gene HK2 with "type 2 diabetes" were found. Three of them described the same exact positive correlation between the *MeSH* (Medical Subject Heading) terms "Diabetes Mellitus, Type 2" and the gene HK2, supported by as many papers. One entry associated the gene to the terms "Precancerous Conditions" and "Diabetes Mellitus, Type 2", and another associated the gene to the terms "Diabetes Mellitus, Type 2" and "Glucose Intolerance". The sixth term described a negative association between HK2 and the terms "Diabetes Mellitus, Type 2" and "Insulin Resistance". Single entries connecting one gene to multiple related diseases were also found. For example, the gene GABRA6 was associated to "anxiety disorder; blood pressure, arterial; memory impairment". Moreover, many cases of synonyms and variant names of the same diseases were found (e.g. "chronic obstructive pulmonary disease" was also indicated by its acronym "COPD"; 11 different terms containing the word "anxiety", including "anxiety disorder", "anxiety symptoms", "anxiety traits" and "somatic anxiety"). Lastly, the data provided by GAD as far as disease terms were

concerned, presented redundancies due to typos, punctuation mistakes and other errors inevitably related to any human recording activity.

In light of all this, firstly, human error was manually corrected in the disease nomenclature, also clearly separating main terms from comments and alternative names. Then, a semi-automatic data cleaning procedure was implemented in order to split entries involving different diseases and connect separate entries referring to the same disease. This procedure is based on the *Damerau-Levenshtein* (***DL***) distance between two strings, given by the minimum number of operations needed to transform one string into the other. In particular, entries containing multiple disease terms, usually separated by semicolon ";" symbols, are split into multiple entries. Then, disease names are screened for high frequent words, such as "syndrome", "type" or "disorder". These words are removed from the subsequent phases as they are deemed not informative. Then, for each pair of strings, the DL distance is calculated. The two strings are considered similar if: (1) DL <= 3; (2) the DL is smaller than the size difference between the two strings and DL is smaller than the smaller of the two strings (i.e. one string contains the other as a substring); (3) the two strings contain at least 2 words in common. Similar strings are clustered together and displayed to the user who selects one of them, in order to provide them with all the close terms that they might be interested in.

We also included miRNA expression data from *The Cancer Genome Atlas* (***TCGA***). In particular, data on 13 different types of cancer, for which both tumor and control data were available, was retrieved. Differential expression analysis was performed using the *Linear Models for Microarray Analysis* (LIMMA) package from Bioconductor in R, as suggested in [201]. The p-values were adjusted for multiple testing using the Benjamini and

Hochberg method to control the *False Discovery Rate* (FDR).

Finally, circulating miRNA data from the *miRandola database* [202] was included. For each circulating miRNA, information about the type of biological fluid (e.g. plasma, serum, urine, …) and the associated molecule or vehicle (e.g. Ago2, exosome, …) was retrieved.

miRNA and genomic elements: CpG islands, Fragile Sites, Translocation Breakpoints and Repeats

In a work previously published by Prof. Alfredo Ferro's Bioinformatics group, the incidence of human miRNA genes on fragile sites and their association with cancer-specific translocation breakpoints, repetitive elements, and CpG islands was investigated [203]. It was found that fragile sites are, on average, denser in miRNAs and genes and that the distribution of miRNAs and genes in fragile versus non-fragile sites depends on chromosome. It was also found that there is a positive correlation between fragility and repeats and between miRNAs and CpG islands. Thus miRò was extended with this data, in order to create two layers of associations between miRNA and genomic elements. The first layer is based on the mapping of miRNA genes. Fragile sites, repeats and CpG islands coordinates were imported from NCBI, and translocation breakpoints data from the *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* (cgap.nci.nih.gov/Chromosomes/Mitelman). This is a novel feature that allows to easily retrieve all the miRNA genes located on unstable genomic regions and to combine this information with the other data contained in the database. The second layer of associations is based on the mapping of miRNA targets, and allows the identification of all the miRNAs *whose targets* are located in unstable genomic regions.

miRNAs, genes and genomic elements are referenced by both the absolute genomic coordinates (start, stop, strand) and the map location (e.g. 1p36). This facilitates the queries and offers more flexibility to users.

5.3 miRò 2 web interface and system inquiries

The power of miRò 2 is revealed in that it allows to combine all data in order to build powerful queries. miRò 2 is equipped with three different search options: simple query, advanced query and miRNA-set functional analysis. Results are displayed in specific tables per category, all downloadable in CSV format singularly as well as globally.

5.3.1 Simple search

The simple query option allows users to obtain information about a single term belonging to one of the following categories: miRNA, gene, disease, pathway, process, function, cellular component, genomic elements, miRNA expression in disease, TCGA and genomic locations. Once a category has been selected, the user can choose one item by either typing its name or selecting it from a terms list. The user must also specify which type of information associated to the selected item they wish to retrieve, by means of a multiple check-box mask whose options are, usually, the other categories as listed above (Figure 31). Such a query allows, for example, to obtain the list of processes and pathways associated to a miRNA through its validated and/or predicted targets, check whether the miRNA and/or its targets are in an unstable genomic area, get a list of diseases in which the miRNA is de-regulated or check whether the miRNA has been found circulating in any body fluid. Several other queries are also possible, such as

selecting a gene in order to retrieve all its annotations along with the list of associated miRNAs, selecting a pathway in order to retrieve all the associated genes and the miRNAs which regulate them, or specifying a chromosomal band to obtain the related genomic annotations as well as the list of miRNAs and genes which are located in it.

This query option also allows the user to choose the *support* for miRNA associations, i.e. through validated and/or predicted targets, and specify a consensus threshold for the predictions, i.e. the minimum number of tools predicting the interaction. The results are shown in tables, one per category, and consist of lists of terms (Figure 30). *Specificity score* and *p-value* are provided for miRNA-term associations. A detailed description of these features is given in a following sub-section.

The reported terms are at times provided with additional information that can be retrieved by simply clicking their name, along with external links to the original sources. For example, the list of miRNAs associated to a certain pathway will contain detailed information about the miRNAs and their precursors, their predicted and/or validated target genes involved in the pathway and various information about the interactions (e.g. prediction tool, transcript, binding sites, score).

In the simple search for miRNA, information on the *genomic elements* is provided through *both* target genes and miRNA precursors. This allows, for example, to distinguish between miRNAs located in a fragile site and miRNAs whose targets are located in a fragile site.

When the query category is a genomic element, such as a translocation breakpoint or a fragile site, the user can specify a genomic location through *absolute coordinates* (i.e.

chromosome, start, stop and strand) or *cytogenetic bands* (i.e. chromosome, arm, region, band, sub-band, sub-sub-band). The system will return genes and miRNAs whose genomic locus overlaps with that of the specified genomic element, along with all the terms associated to those same genes, such as pathways, diseases, processes, functions, cellular components and targeting miRNAs.

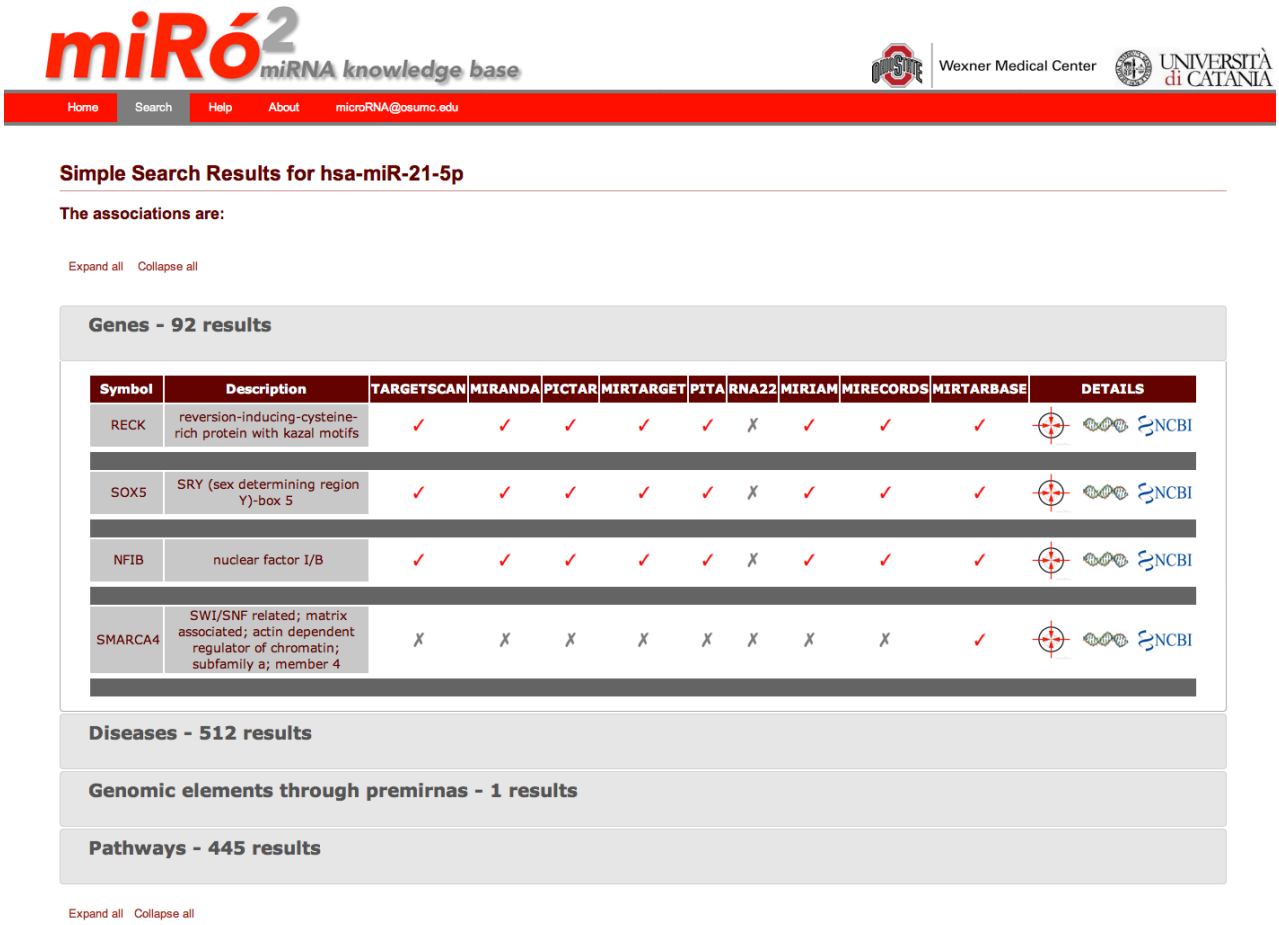

Figure 30: Simple Search results in miRò 2.

Figure 31: Simple search web interface in miRò 2

5.3.2 Advanced search

The advanced search allows users to express very sophisticated queries by selecting a category of interest (e.g. miRNA, gene, pathway, etc.), a list of terms belonging to the remaining categories as *constraints to be satisfied*, and an optional list of terms that must *not* be associated to the results (Figure 32). For example, one may retrieve all the circulating miRNAs encoded in fragile sites, which are also associated to certain pathways and processes but not to specific others, and whose over-expression has been reported in some specific disease conditions. Similarly, the user can request the list of pathways which share a subset of genes and miRNAs, as well as being potentially associated to certain processes and diseases through common genes.

As for the simple search, the user can specify the miRNA-target support, whether validated or predicted and, in the latter case, a threshold for the number of tools predicting the interaction.

When the query category is miRNA, a filter allows to specify details such as fold change, detection method and source when selecting miRNA expression in disease or from TCGA as constraints. The results are returned in tables with details on the target genes associated to the query terms (Figure 33).

This powerful type of search allows to identify miRNAs which are associated to the same terms albeit through different targets.

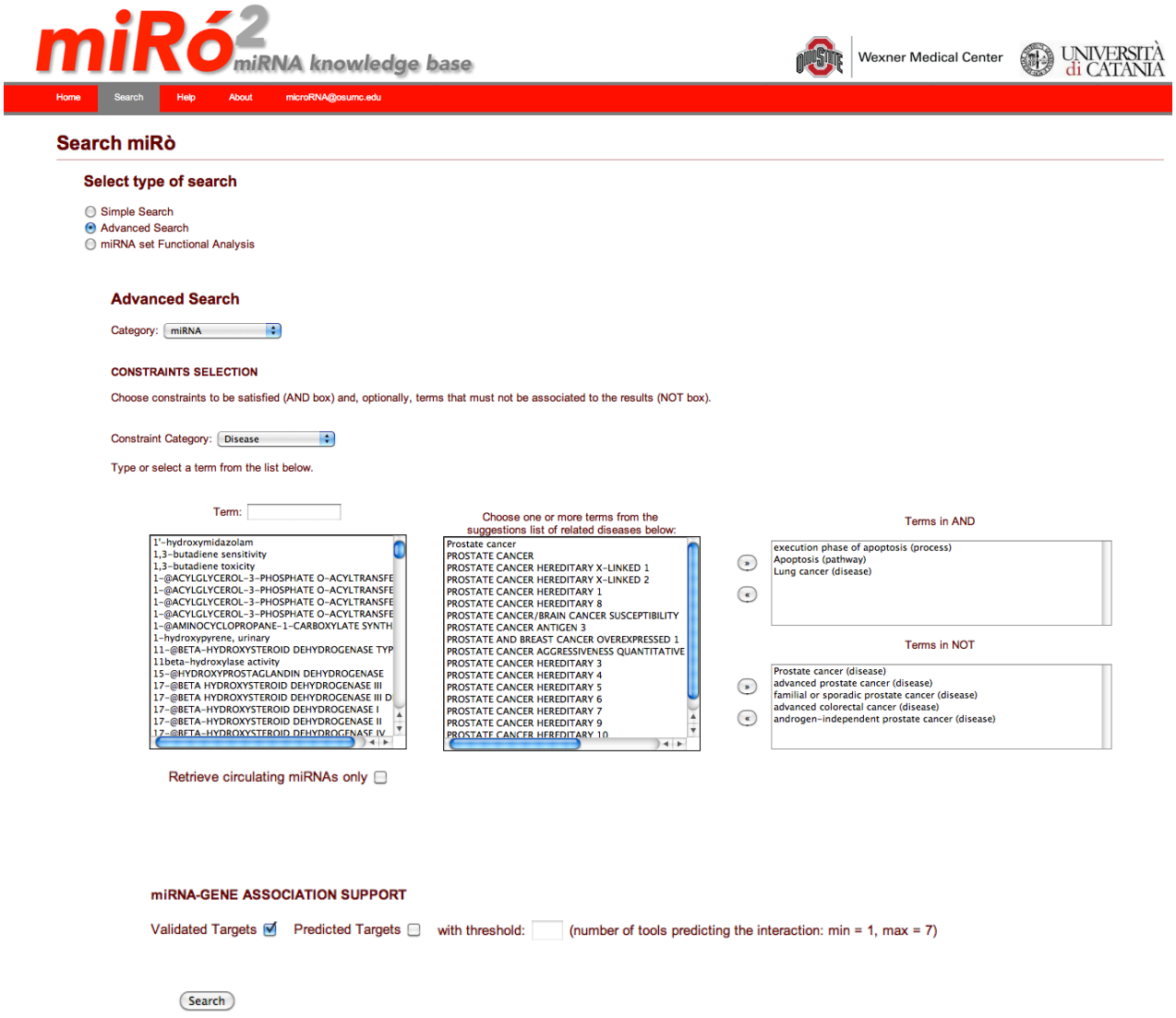

Figure 32: Advanced Search web interface in miRò 2

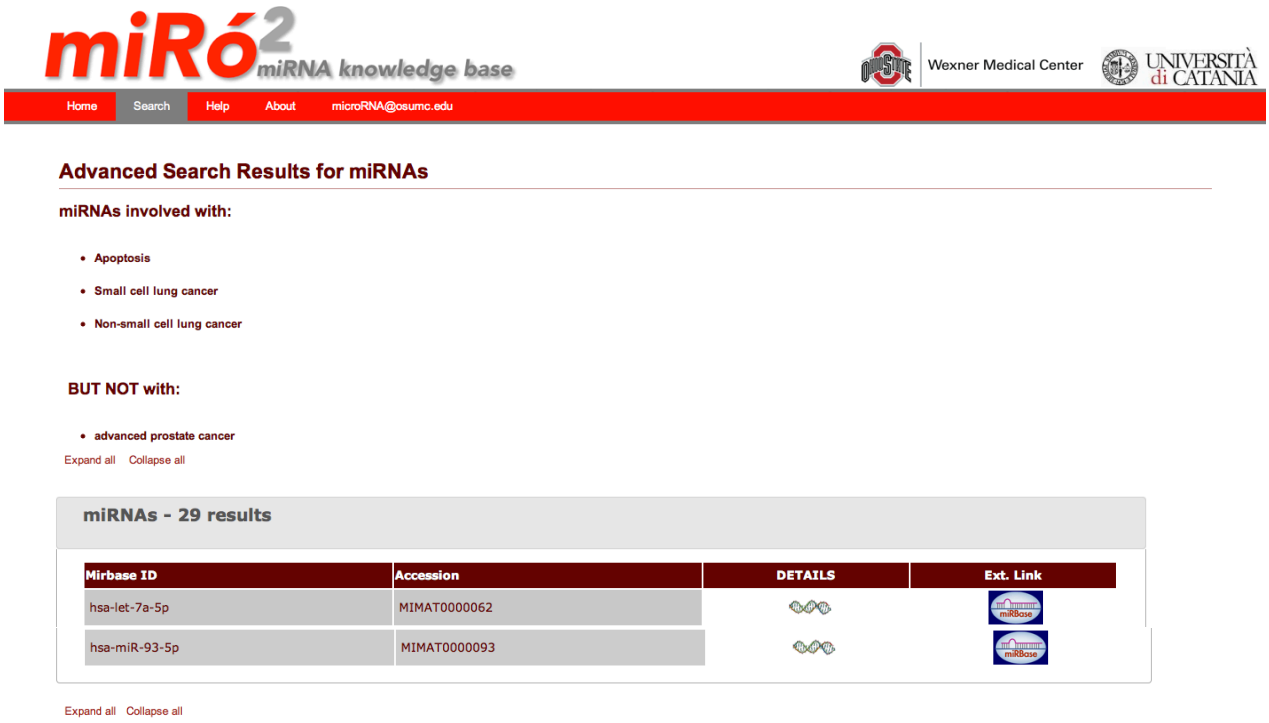

Figure 33: Advanced Search Results in miRò 2

5.3.3 Datamining in miRò2: miRNA-set Functional Analysis

With this type of search the user may perform a functional enrichment analysis for a subset of arbitrary miRNAs and a set of chosen categories (Figure 34). This can be very useful, for example, when one wants to investigate potential functional correlation shared by a group of miRNAs which are co-deregulated in specific conditions. Given a set of *n* miRNAs, the user may get the lists of diseases, GO terms, pathways and genomic elements shared by all or by any subset of them or find out, for example, if all or some of them have been found circulating in plasma or serum.

This query system also allows quick selection of miRNA clusters as input for the enrichment analysis. A *miRNA cluster* is defined as a group of precursors with an inter-miRNA distance of less than 10kb on the same genomic strand. When this option is enabled, the user can select a whole miRNA cluster by just choosing one of its members and perform the analysis on the whole cluster or on a subset. This unique feature of miRò 2 provides the user with a simple yet powerful tool for the detection of potential functions for clustered miRNAs. miRNAs from the same cluster, indeed, are mostly co-expressed and have been shown to share functions or be involved in the same diseases.

The results are displayed by category, each containing a table for each subset of the miRNA set selected by the user and corresponding to a group of miRNAs functionally associated to each other through the list of common terms for the given category specified for each group (Figure 35).

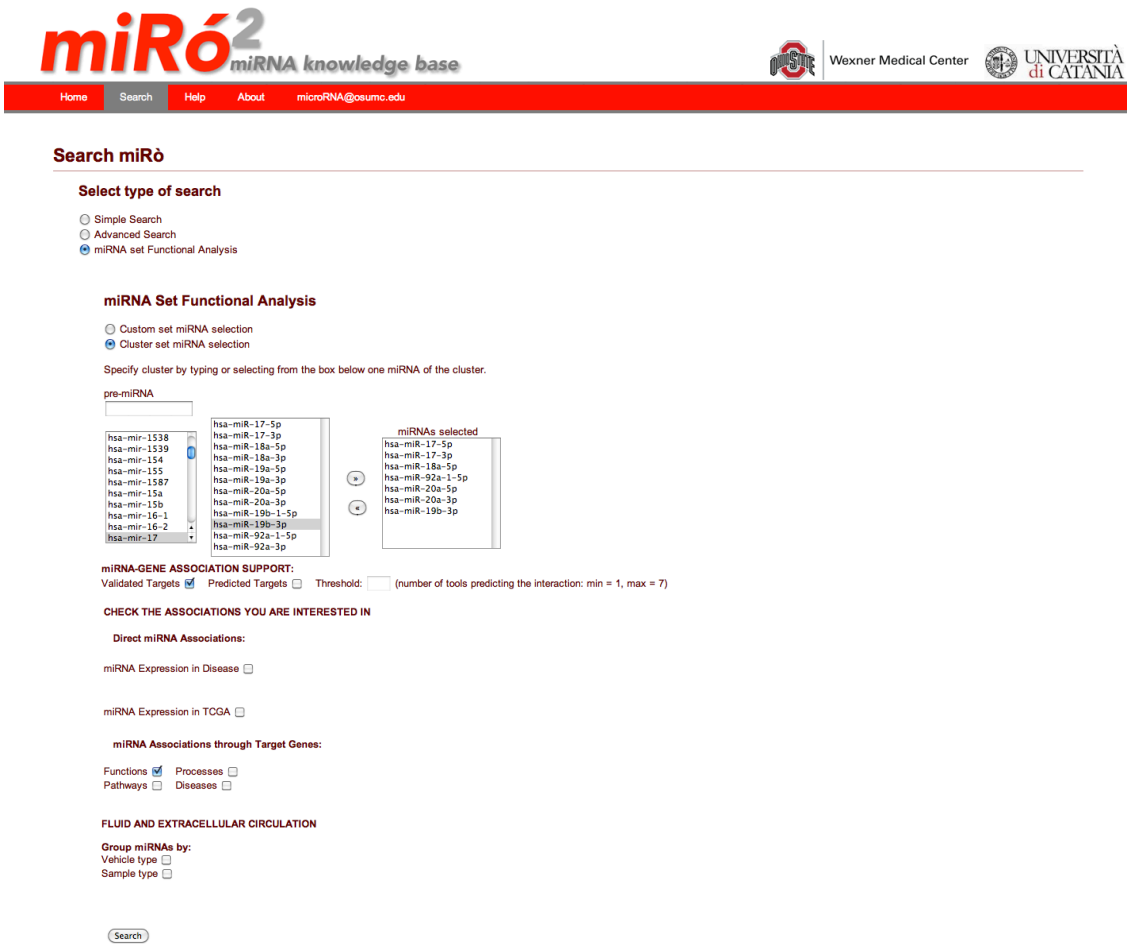

Figure 34: miRNA-set Functional Analysis web interface in miRò 2

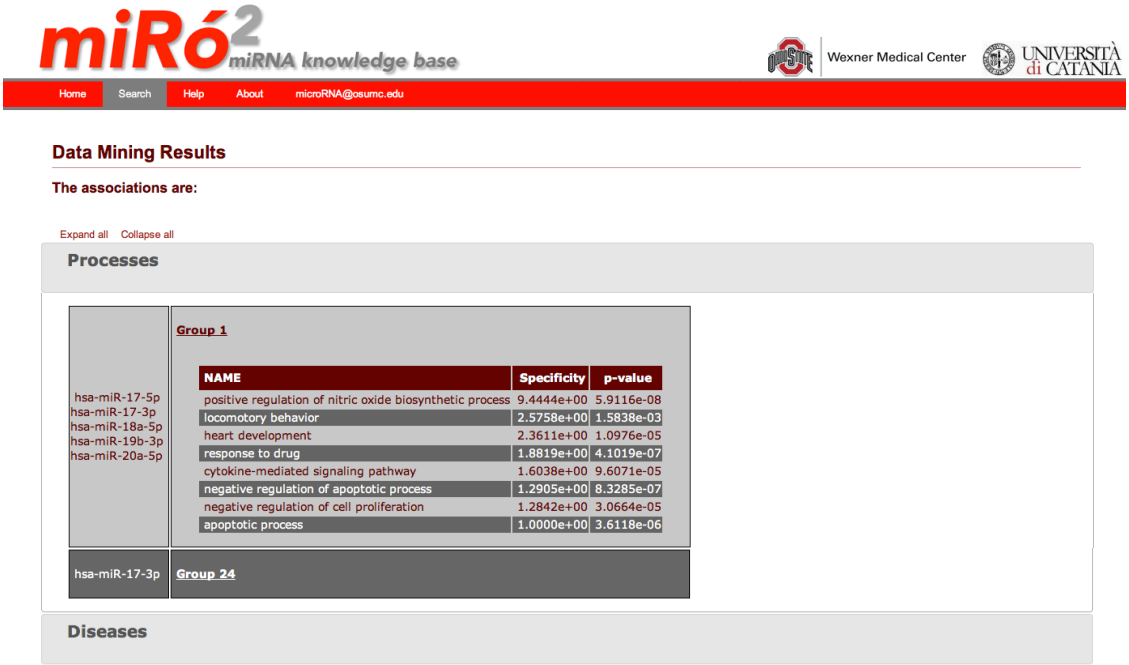

Figure 35: miRNA-set Functional Analysis results in miRò 2

5.4 Specificity of miRNA-phenotype associations

As in the first release of miRò, to each miRNA-term association a specificity score was assigned. Given a miRNA *mk* and a term *tj* (e.g. process), the ***specificity*** is calculated as follows:

$$S_{m_k,t_j} = \frac{\left|G_{m_k,t_j}\right|}{\left|G_{m_k}\right|} \cdot \frac{\sum_{g_i \in G_{m_k,t_j}} S_{g_i}}{\left|G_{m_k,t_j}\right|} = \frac{\sum_{g_i \in G_{m_k,t_j}} S_{g_i}}{\left|G_{m_k}\right|}$$

where $G_{mk,tj}$ is the set of the target genes of miRNA m_k involved in the process p_j, and G_{mk} is the set of all the target genes of m_k. The specificity of a gene S_{gi} is inversely proportional to the number of processes in which the gene is involved:

$$S_{g_i} = \frac{1}{\left|P_{g_i}\right|}$$

where P_{gi} is the set of the processes in which the gene g_i is involved.

The statistical significance of miRNA-term associations is assessed by means of the *Fisher's Exact Test*, which is a computationally efficient method of calculating a hypergeometric distribution. Given a miRNA-term pair, the p-value is calculated at run time, as the contingency table may vary depending on the chosen miRNA-target support and prediction threshold.

P-values of miRNA-term associations are returned for simple search queries and miRNA-

set functional analysis. In the latter case, the computation is extended to the association of a *set* of miRNAs to a single term.

The FDR is assessed by the Benjamini–Hochberg procedure.

All the associations with p-value >= 0.05 are automatically discarded and are not shown in the results. The user may also choose to lower the p-value or the adjusted p-value threshold in order to obtain more significant results.

5.5 Case studies and validation

In order to assess the reliability and robustness of miRò 2, specific queries designed to recover miRNA functions and associations reported in the literature were performed.

5.5.1 Cluster miR-17-92

The crucial role of miR-17-92 cluster in development and in various diseases has been amply demonstrated. The expression of these miRNAs promote cell proliferation, suppresses apoptosis of cancer cells and induces angiogenesis of tumors. In particular, these miRNAs are involved in lymphoma, melanoma and other types of cancer (breast, colon-rectum, lung, ovary, pancreas, prostate and stomach). The miR-17-92 cluster also plays an essential role in the normal development of the heart, lungs and immune system [57].

Performing a miRNA-set functional analysis in miRò 2 for diseases related to this miRNA cluster, it was found that all of the miRNAs in the cluster are associated with the

tumors mentioned above along with other diseases. Moreover, the same analysis for processes that involve the cluster, returns, among others, angiogenesis, apoptosis, cell cycle, growth and cell proliferation as well as the development of heart and lungs, confirming what was already reported in literature.

5.5.2 Other case studies

miRò has been successfully used to generate hypotheses on the involvement of miRNAs in diseases and biological processes, as reported in a number of papers [204-208].

For example, Kumar et al. [204] reported the identification of differential levels of circulating serum miR-132 and the miR-134 family of miRNAs, which could differentiate mild cognitive impairment from normal control population, after including in their analysis miRNAs reported by miRò to be involved in neurite- and synapse-associated processes. Furthermore, miRNA-set functional analysis confirmed the associations of miRNAs belonging to those two families to diseases such as *syndromic x-linked ZDHHC9-related mental retardation* as well as other types of mental retardation such as *autism* and *vascular dementia*. The miRNAs of these families are also reported by miRò as circulating in serum as well as other sample types, and associated to processes such as *cell morphogenesis involved in neuron differentiation*, *neuron maturation*, *regulation of neuron differentiation*, *neuron apoptotic process*, *neuron projection morphogenesis and development*, *neuron fate specification* and *neuron migration*, as well as being involved in pathways such as *Neurotrophic factor-mediated Trk receptor signaling* and *Neurotrophin signaling* and functions such as *neurotransmitter sodium symporter activity* and *neurotrophin binding and receptor activity*.

Palmieri et al. [207] also made use of miRò in identifying and validating new miRNAs targeting highly evolutionarily conserved sequences of the 3'-UTRs of HMGA1 and HMGA2 mRNAs. They were able to confirm a downregulated expression of these miRNAs, and other previously described HMGA-targeting miRNAs, in human pituitary adenomas in comparison with the normal pituitary gland. Among phenotype associations as reported by miRò for miRNAs miR-15ab, miR-16, miR-26ab, miR-196ab and Let-7a as potential HMGA-targeting miRNAs, cellular processes, including those in which HMGA proteins have been previously demonstrated to be involved, such as *cell proliferation*, *differentiation*, *DNA repair*, *chromatin modification* and *regulation of transcription* (Fusco and Fedele, 2007; Fedele and Fusco, 2010) were found. The direct targeting of HMGA1 by miR-16 (Kaddar et al., 2009), and of HMGA2 by Let-7a (Lee and Dutta, 2007) and miR196ab (De Martino et al., 2009a) were all previously reported. Moreover, previous studies showed that miR-26a regulates HMGA2 expression (Lee et al., 2011). Finally, among the many reports provided by miRò, it is interesting to point out that miRNA-set functional analysis on the miRNAs mentioned above reported *pituitary gland development* and *negative regulation of cell proliferation* among the top associated processes significantly shared by all of these miRNAs, a result confirmed by Palmieri whose functional studies have demonstrated that all the analyzed HMGA-targeting miRNAs inhibit the proliferation of a rat pituitary adenoma cell line, suggesting a critical role of their downregulation in pituitary tumorigenesis and thus attesting the reliability of miRò's reported associations.

Generally, though, the simple search can be very useful in a variety of circumstances as well. First of all, it provides users with a summary of miRNA functions. A search for miR-221-3p, which is one of the most extensively studied miRNAs, returns, among

others, a list of diseases for which its involvement has been experimentally determined, such as *thyroid cancer*, *stomach cancer*, *colon cancer*, *glioblastoma*, *melanoma* and *breast cancer*[209-214]. Results show also other experimentally validated associations such as a significant involvement of miR-221-3p in pathways such as *activation of PUMA*, *TRAIL signaling* and *p53 signaling* as well as processes such as *melanocyte migration* and *mast cell proliferation* [209,210,215-217].

Lastly, miRò has functionally associated miRNA-set miR-27a/23a/24-2/24-3p/30b/30d/145-3p/210/222-5p/4521 to *cellular response to hypoxia* and and to *breast cancer* with a high hit-ratio and significant p-value scores (~2.7/3.2 hit-ratio and p-value < 0.001; ~1.42/1.75 hit-ratio and p-value < 0.0001), as confirmed for the very first time by Camps et al. [218]. miR-4521 (a mostly non-elucidated miRNA) was also predicted to target FOXM1 as reported by Camps and significantly associated to *hypoxia-inducible factor-1alpha signaling* (through PDK3) and *hypoxic and oxygen homeostasis regulation of HIF-1-alpha* (through ENO1) pathways, as well as to breast cancer through KRAS, IGF1, MAPK1, ESPL1, ITGB3, DPYD, BTLA and CENPF.

5.5.3 Specificity function validation

The specificity function, introduced in section 5.4, has the objective of evaluating miRNA annotations, in order to enable the identification of the most significant ones.

Among the miRNA/disease and miRNA/process associations with the highest score, there are cases reported in the literature, as already mentioned in the previous paragraph. For example, among the strongest miRNA/disease association is the one that connects miR-433 to Parkinson's disease. This result is confirmed by a study in which it is shown

that the deletion of the binding site for miR-433 on the transcript of FGF20 increases the risk of disease onset. Indeed, over-expression of FGF20 is correlated to the over-expression of alpha-synuclein, for which a direct correlation with Parkinson's disease has been documented [219].

Similarly, the association between miR-224 and apoptosis is among the miRNA/process associations with greater score. This is supported by a study in which it shows the correlation between miR-224, overexpressed in hepatocellular carcinoma, and increased cell death through the targeting of apoptosis inhibitor API-5 [220].

5.6 Conclusions

This chapter presented miRò 2, a system for the functional annotation of miRNAs, based on the web. The system, previously published in its first version, has undergone extension and update and was here illustrated in its second version. The main innovation is the substantial contribution of additional data and powerful data mining capabilities in the endeavor to provide users with a unified resources for miRNA research. Thus, the ability to highlight relationships between genes, processes, functions and diseases at the miRNA level has been further extended to pathways, miRNA expression in diseases and genomic elements such as repeats, CpG islands, fragile sites and translocation breakpoints. An important and very actual aspect of miRNA research today such as circulation info on miRNAs has also been included in the second version of the system, allowing such

important aspect, as well as others, to be accounted for, especially through the novel data mining functionality offered by miRò 2 for miRNA sets and the function of specificity which allows the selection of the most significant associations.

Currently the system is in its final stage of development and is undergoing the integration of a few final categories of data, such as mutation and single nucleotide polymorphisms info through a source called *SomamiR*, a database containing somatic mutations that may create or disrupt miRNA target sites and which integrates these somatic mutations with germline mutations within the same target sites, genome-wide and candidate gene association studies of cancer and functional annotations that link genes containing mutations with cancer [221]. Other data regarding expression profiles of miRNA target genes, as well as information on inter-species conservation of miRNAs and targets will be considered as future improvements along with a new scoring function for miRNA annotation based on semantic associations between words, taken from the biomedical literature.

The work will soon be submitted for publication and can be currently found on the web at *http://microrna.osumc.edu/miro*.

Conclusions

This thesis has elucidated the research projects developed within the context of the doctorate program, having as objective the study of miRNA-mediated gene regulation through the analysis of RNA sequences and structures aided by heuristics and algorithms for data mining, with applications to the field of RNA interference, synthetic biology and RNA editing.

In particular, the development of a database for predicted RNA A-to-I edited miRNA binding sites, an algorithm for the design of synthetic miRNAs with high specificity and a knowledge base with Data Mining tools for miRNA functional annotation, have lead to results which show the importance of using efficient computational methods for the analysis of biological data in order to better search meaningful patterns and more appropriately predict biomolecular interactions and functions, providing a solid base for promising future developments.

Future directions will focus on expanding the tools here developed, specifically: adding important info on editing events occurring in the mature miRNA sequences that could further affect their gene-silencing role; the expansion of the miR-Synth system in order to introduce new synthetic RNA types, such as long non-coding RNA, along with multiple synthetic miRNAs, for a more comprehensive design and evaluation of artificial RNA molecules for the regulation of altered pathways in cancer; provide interspecies info and expression correlation data to the miRò knowledge base.

ACKNOWLEDGEMENTS

I would like to thank many, many people that I've crossed paths with during the experience that has been this PhD program. Unfortunately, I don't have enough space here but just know that I hold much more space in my heart where you're all accounted for! I am very grateful for all of you so even if I don't mention your name in these acknowledgements, just know you're in here too.

I couldn't go without explicitly acknowledging some specific people though, for you have been the direct reason I got where I am today.

In particular, I would like to express my profound gratitude to my coordinator, Prof. Alfredo Ferro, for believing in my abilities and encouraging and giving me the opportunity to accomplish this work and live this very edifying experience, both professionally and from a human stand point. I am truly grateful!

I'd also like to thank Dr. Carlo M. Croce and all the people in his lab for welcoming me and allowing me to grow professionally in such challenging and prestigious research environment at The Ohio State University, I feel very honored.

I couldn't go without acknowledging my supervisor Dr. Alessandro Lagana'. Ale, we've come a long way in this experience that has forged both our lives and put us in the paths we were meant to walk in. I'm just thankful you were there to accompany me and introduce me to a whole new world that I never imagined. I wish you the best to you and Emily in NYC!

My gratitude goes very much also to Dr. Gianni Nigita, co-worker and friend. Gianni, thank you for supporting me in my challenging professional moments and for teaming up with me in facing all the requests of the lab once Ale was gone and grow together. You've been a great help and may God reward you accordingly!

All the friends, the close and the far ones, in Italy and in the USA! So many! I'm very grateful in particular for Anna Nora Gurgone, you've been an inspiration and a constant support even from far away, God bless you; all my closest friends since forever and the more recent ones that have ALWAYS been there for me, supporting me, encouraging me, just giving me love. I'd like to thank Alex Holznienkemper for being a true PhD buddy!

All the dear friends at the residence Alcantara, Sicily, and Warwick House, Pittsburgh, as well as Father Antonio Carvalho of Holy Name Church in Columbus, Ohio, for the prayers and constant spiritual support and true friendship: you've made me grow and I love you guys!!

A most profound gratitude goes to my family: my grandparents, my uncles and aunts, my beautiful girl cousins in Michigan, as well as all my relatives in Italy. You've always been there for me and I know you support me always! Thank you so much for your love!

To my parents, my brother and my sister. Mom and Dad, you've been the two pillars on which my life has grown and you've always, ALWAYS, reassured me and have proven to be that great help and support any son dreams of. I will never be able to repay you for all your love. You made me who I am in every way! I dedicate this work to honor the both of you.

Denis, you've always been so close to me even from far far away. Your loving care, your prayers and encouragement have been precious and I'm so thankful this experience of mine has been able to give you the opportunity to embark on a new adventure and accompany me as always!

Vanessa, God knows how much I miss you! Thank you for your wonderful love and closeness, even from far away…I've always been too often rather geographically far

from you but with the heart I never leave your side. Thank you for your prayers and for your strength, you're wonderful!

The most intimate and cherished gratitude goes to the one that has walked by my side for the past two years, constantly loving me and making my life just SHINE! Thank you Nikki, for being that wonderful, solid and strong encouragement, always there for me with a most beautiful and heart-warming company. I am so grateful to Our Lord for putting you in my life and for your wonderful family, whom has welcomed me like a son!! You are a godsend and I love you with my whole heart!

Last, but certainly not least, my heart goes out to The One Who's the reason of my many thanks to so many wonderful people, each one of them reflecting His love for me! To my God and Lord Jesus Christ, through His Most Blessed Mother whom I love most dearly, I offer all the strides of this accomplished work, thanking Him for making me walk this distance and grow toward His Kingdom of Love: *Deo omnis gloria*!

References

1. Central dogma of molecular biology. 1970;227(5258):561–563. Available at: http://www.nature.com/nature/focus/crick/pdf/crick227.pdf.

2. Origin of life: The RNA world. 1986. Available at: http://www.nature.com/nature/journal/v319/n6055/pdf/319618a0.pdf.

3. Non-coding RNA genes and the modern RNA world. 2001;2(12):919–929. doi:10.1038/35103511.

4. RNA takes center stage. 2007;87(5-6):275–278. doi:10.1002/bip.20824.

5. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. 2007;447(7146):799–816. doi:10.1038/nature05874.

6. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. 2008;24(7):344–352. doi:10.1016/j.tig.2008.04.005.

7. Non-coding RNAs: the architects of eukaryotic complexity. 2001;2(11):986–991. doi:10.1093/embo-reports/kve230.

8. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. 1993;75(5):843–854. doi:10.1016/0092-8674(93)90529-Y.

9. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. 1998;391(6669):806–811. doi:10.1038/35888.

10. Perspective: machines for RNAi. 2005;19(5):517–529. doi:10.1101/gad.1284105.

11. Small RNAs: Regulators and guardians of the genome. 2007;213(2):412–419. doi:10.1002/jcp.21230.

12. MicroRNAs: genomics, biogenesis, mechanism, and function. 2004;116(2):281–297. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=14744438&retmode=ref&cmd=prlinks.

13. Mechanisms of gene silencing by double-stranded RNA. 2004;431(7006):343–349. doi:10.1038/nature02873.

14. miRNA, siRNA, piRNA: Knowns of the unknown. 2008;5(4):181–188. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19182524&retmode=ref&cmd=prlinks.

15. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. 2001;411(6836):494–498. doi:10.1038/35078107.

16. A system for stable expression of short interfering RNAs in mammalian cells. 2002. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=11910072&retmode=ref&cmd=prlinks.

17. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. 2002;16(8):948–958. doi:10.1101/gad.981002.

18. Role for a bidentate ribonuclease in the initiation step of RNA interference. 2001;409(6818):363–366. doi:10.1038/35053110.

19. Mammalian microRNAs predominantly act to decrease target mRNA levels. 2010;466(7308):835–840. doi:10.1038/nature09267.

20. The widespread regulation of microRNA biogenesis, function and decay. 2010;11(9):597–610. doi:10.1038/nrg2843.

21. MicroRNAs in Cancer: Small Molecules With a Huge Impact. 2009;27(34):5848–5856. doi:10.1200/JCO.2009.24.0317.

22. miRBase: microRNA sequences, targets and gene nomenclature. 2006;34(suppl 1):D140–D144. doi:10.1093/nar/gkj112.

23. miRBase: integrating microRNA annotation and deep-sequencing data. 2010. Available at: http://nar.oxfordjournals.org/content/early/2010/10/30/nar.gkq1027.short.

24. MicroRNA biogenesis: coordinated cropping and dicing. 2005;6(5):376–385. doi:10.1038/nrm1644.

25. Functional relevance of miRNA sequences in human disease. 2012;731(1-2):14–19. doi:10.1016/j.mrfmmm.2011.10.014.

26. Biogenesis of small RNAs in animals. 2009;10(2):126–139. doi:10.1038/nrm2632.

27. Mammalian Mirtron Genes. 2007;28(2):328–336. doi:10.1016/j.molcel.2007.09.028.

28. The Mirtron Pathway Generates microRNA-Class Regulatory RNAs

inDrosophila. 2007;130(1):89–100. doi:10.1016/j.cell.2007.06.028.

29. Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. 2007;450(7167):219–232. doi:10.1038/nature06340.

30. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. 2003;17(24):3011–3016. doi:10.1101/gad.1158803.

31. Selective Blockade of MicroRNA Processing by Lin28. 2008;320(5872):97–100. doi:10.1126/science.1154040.

32. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. 2007;14(7):591–596. doi:10.1038/nsmb1250.

33. A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. 2001;293(5531):834–838. doi:10.1126/science.1062961.

34. Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. 2012;151(3):521–532. doi:10.1016/j.cell.2012.09.022.

35. TUT4 in Concert with Lin28 Suppresses MicroRNA Biogenesis through Pre-MicroRNA Uridylation. 2009;138(4):696–708. doi:10.1016/j.cell.2009.08.002.

36. Dicer partner proteins tune the length of mature miRNAs in flies and mammals. 2012;151(3):533–546. doi:10.1016/j.cell.2012.09.027.

37. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. 2005;436(7051):740–744. doi:10.1038/nature03868.

38. The long and short of microRNA. 2013;153(3):516–519. doi:10.1016/j.cell.2013.04.003.

39. c-Myc-regulated microRNAs modulate E2F1 expression. 2005;435(7043):839–843. doi:10.1038/nature03677.

40. Reciprocal actions of REST and a microRNA promote neuronal identity. 2006;103(7):2422–2427. doi:10.1073/pnas.0511041103.

41. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. 2009;139(1):112–122. doi:10.1016/j.cell.2009.06.044.

42. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. 2009;23(3):304–317. doi:10.1101/gad.1749809.

43. RNA-Binding Protein Dnd1 Inhibits MicroRNA Access to Target mRNA. 2007;131(7):1273–1286. doi:10.1016/j.cell.2007.11.034.

44. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5“ UTR as in the 3” UTR. 2007;104(23):9667–9672. doi:10.1073/pnas.0703820104.

45. Mechanism of translational regulation by miR-2 from sites in the 5' untranslated region or the open reading frame. 2010;16(12):2493–2502. doi:10.1261/rna.2384610.

46. Switching from Repression to Activation: MicroRNAs Can Up-Regulate Translation. 2007;318(5858):1931–1934. doi:10.1126/science.1149460.

47. Combinatorial microRNA target predictions. 2005;37(5):495–500. doi:10.1038/ng1536.

48. The microRNA.org resource: targets and expression. 2008;36(Database issue):D149–D153. doi:10.1093/nar/gkm995.

49. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. 2005;120(1):15–20. doi:10.1016/j.cell.2004.12.035.

50. Prediction of mammalian microRNA targets. 2003;115(7):787–798. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=14697198&retmode=ref&cmd=prlinks.

51. Therapeutic targeting of microRNAs: current status and future challenges. 2014;13(8):622–638. doi:10.1038/nrd4359.

52. Dicer is essential for mouse development. 2003;35(3):215–217. doi:10.1038/ng1253.

53. Dicer function is essential for lung epithelium morphogenesis. 2006;103(7):2208–2213. doi:10.1073/pnas.0510839103.

54. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. 2005;102(31):10898–10903. doi:10.1073/pnas.0504834102.

55. Essential role for Dicer during skeletal muscle development. 2007;311(2):359–368. doi:10.1016/j.ydbio.2007.08.032.

56. Aberrant T cell differentiation in the absence of Dicer. 2005;202(2):261–269. doi:10.1084/jem.20050678.

57. Targeted Deletion Reveals Essential and Overlapping Functions of the miR-17~92 Family of miRNA Clusters. 2008;132(5):875–886. doi:10.1016/j.cell.2008.02.019.

58. MicroRNAs Modulate Hematopoietic Lineage Differentiation. 2004;303(5654):83–86. doi:10.1126/science.1091903.

59. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. 2008;14(6):843–853. doi:10.1016/j.devcel.2008.03.012.

60. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. 2004;5(3):R13. doi:10.1186/gb-2004-5-3-r13.

61. Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. 2005;19(19):2343–2354. doi:10.1101/gad.1356105.

62. MicroRNAs in Skeletal and Cardiac Muscle Development. 2007;26(4):219–225. doi:10.1089/dna.2006.0556.

63. MicroRNA-26a Targets the Histone Methyltransferase Enhancer of Zeste homolog 2 during Myogenesis. 2008;283(15):9836–9843. doi:10.1074/jbc.M709614200.

64. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. 2007;21(1):71–84. doi:10.1101/gad.1500707.

65. The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. 2007;27(3):435–448. doi:10.1016/j.molcel.2007.07.015.

66. The Expression of MicroRNA miR-107 Decreases Early in Alzheimer's Disease and May Accelerate Disease Progression through Regulation of β-Site Amyloid Precursor Protein-Cleaving Enzyme 1. 2008;28(5):1213–1223. doi:10.1523/JNEUROSCI.5065-07.2008.

67. A MicroRNA Feedback Circuit in Midbrain Dopamine Neurons. 2007;317(5842):1220–1224. doi:10.1126/science.1140481.

68. A New Role for MicroRNA Pathways: Modulation of Degeneration Induced by Pathogenic Human Disease Proteins. 2006;5(24):2835–2838. doi:10.4161/cc.5.24.3579.

69. First in vivo evidence of microRNA-induced fragile X mental retardation syndrome. 2006;11(7):616–617. doi:10.1038/sj.mp.4001828.

70. microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. 2007;8(2):R27–R27. doi:10.1186/gb-2007-8-2-r27.

71. miR-122, a Mammalian Liver-Specific microRNA, is Processed from hcr mRNA and MayDownregulate the High Affinity Cationic Amino Acid

Transporter CAT-1. 2004;1(2):106–113. doi:10.4161/rna.1.2.1066.

72. Interferon modulation of cellular microRNAs as an antiviral mechanism. 2007;449(7164):919–922. doi:10.1038/nature06205.

73. Distinctive patterns of microRNA expression in primary muscular disorders. 2007;104(43):17016–17021. doi:10.1073/pnas.0708115104.

74. MicroRNA-133 controls cardiac hypertrophy. 2007;13(5):613–618. doi:10.1038/nm1582.

75. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. 2007;13(4):486–491. doi:10.1038/nm1569.

76. The Hallmarks of Cancer. 2000;100(1):57–70. doi:10.1016/S0092-8674(00)81683-9.

77. miR-15 and miR-16 induce apoptosis by targeting BCL2. 2005;102(39):13944–13949. doi:10.1073/pnas.0506654102.

78. RASIs Regulated by thelet-7MicroRNA Family. 2005;120(5):635–647. doi:10.1016/j.cell.2005.01.014.

79. MicroRNA Let-7a Down-regulates MYC and Reverts MYC-Induced Growth in Burkitt Lymphoma Cells. 2007;67(20):9762–9770. doi:10.1158/0008-5472.CAN-07-2462.

80. MicroRNA 29b functions in acute myeloid leukemia. 2009;114(26):5331–5341. doi:10.1182/blood-2009-03-211938.

81. A microRNA polycistron as a potential human oncogene. 2005;435(7043):828–833. doi:10.1038/nature03552.

82. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. 2008;105(10):3945–3950. doi:10.1073/pnas.0800135105.

83. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. 2008;9(4):405–414. doi:10.1038/ni1575.

84. Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. 2009;23(24):2806–2811. doi:10.1101/gad.1872909.

85. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. 2006;103(18):7024–7029. doi:10.1073/pnas.0602266103.

86. CRISPR interference: RNA-directed adaptive immunity in bacteria and

archaea. 2010;11(3):181–190. doi:10.1038/nrg2749.

87. RNA-editing in trypanosome mitochondria. 1989;1007(2):131–139. doi:10.1016/0167-4781(89)90031-6.

88. Rna Editing. 1993:196. Available at: http://books.google.com/books?id=E4XwAAAAMAAJ&q=intitle:RNA+Editing+The+Alteration+of+Protein+Coding+Sequences+of+RNA&dq=intitle:RNA+Editing+The+Alteration+of+Protein+Coding+Sequences+of+RNA&hl=&cd=2&source=gbs_api.

89. Modification and Editing of Rna. 1998:596. Available at: http://books.google.com/books?id=s6Vk4y0g8MgC&dq=intitle:Modification+and+Editing+of+RNA&hl=&cd=1&source=gbs_api.

90. RNA editing in plant mitochondria. 1989;246(4937):1632–1634. doi:10.1126/science.2480644.

91. Identification of critical nucleotide positions for plastid RNA editing site recognition. 1997;3(10):1194–1200. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9326494&retmode=ref&cmd=prlinks.

92. Regenerating good sense: RNA editing and trans splicing in plant mitochondria. 1992;8(9):322–328. doi:10.1016/0168-9525(92)90265-6.

93. RNA editing gives a new meaning to the genetic information in mitochondria and chloroplasts. 1993;21(6):1163–1170. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=8490134&retmode=ref&cmd=prlinks.

94. Editing of a chloroplast mRNA by creation of an initiation codon. 1991;353(6340):178–180. doi:10.1038/353178a0.

95. The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. 1991;10(10):3079–3085. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=1655410&retmode=ref&cmd=prlinks.

96. Cloning and Expression of Human G/T Mismatch-specific Thymine-DNA Glycosylase. 1996;271(22):12767–12774. doi:10.1074/jbc.271.22.12767.

97. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. 1991;67(1):11–19. doi:10.1016/0092-8674(91)90568-J.

98. Adenosine deaminase deficiency: clinical expression, molecular basis, and therapy. 1998;35(4):291–298. Available at:

http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9801258&retmode=ref&cmd=prlinks.

99. A developmentally regulated activity that unwinds RNA duplexes. 1987;48(4):607–613. doi:10.1016/0092-8674(87)90239-X.

100. Cell cycle expression of RNA duplex unwindase activity in mammalian cells. 1988;8(2):770–777. doi:10.1128/MCB.8.2.770.

101. The interferon-inducible, double-stranded RNA-specific adenosine deaminase gene (DSRAD) maps to human chromosome 1q21.1-21.2. 1995;30(2):372–375. doi:10.1006/geno.1995.0034.

102. Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extracts. 1994;269(18):13480–13489. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=8175781&retmode=ref&cmd=prlinks.

103. Purification and properties of double-stranded RNA-specific adenosine deaminase from calf thymus. 1994;91(22):10596–10600. Available at: http://www.pnas.org/content/91/22/10596.short.

104. Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. 1998;275(4):695–714. doi:10.1006/jmbi.1997.1506.

105. Predicted editing of additional transfer RNAs in Acathamoeba castellanii mitochondria. 1993. Available at: http://nar.oxfordjournals.org/content/21/18/4402.full.pdf.

106. On the evolution of RNA editing. 1993;9(8):265–268. doi:10.1016/0168-9525(93)90011-6.

107. Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. 2005;15(4):384–391. doi:10.1016/j.cub.2005.01.046.

108. RNA editing of a miRNA precursor. 2004;10(8):1174–1177. doi:10.1261/rna.7350304.

109. Identification of microRNAs of the herpesvirus family. 2005;2(4):269–276. doi:10.1038/nmeth746.

110. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. 2006;13(1):13–21. doi:10.1038/nsmb1041.

111. RNA editing of human microRNAs. 2006;7(4):R27–R27. doi:10.1186/gb-2006-7-4-r27.

112. Adenosine deamination in human transcripts generates novel microRNA binding sites. 2009;18(24):4801–4807. doi:10.1093/hmg/ddp443.

113. A-to-I RNA editing: the “ADAR” side of human cancer. 2012;23(3):244–250. doi:10.1016/j.semcdb.2011.09.003.

114. The RNA editing enzymes ADARs: mechanism of action and human disease. 2014;356(3):527–532. doi:10.1007/s00441-014-1863-3.

115. The noncoding RNA revolution-trashing old rules to forge new ones. 2014;157(1):77–94. doi:10.1016/j.cell.2014.03.008.

116. DARNED: a DAtabase of RNa EDiting in humans. 2010;26(14):1772–1776. doi:10.1093/bioinformatics/btq285.

117. Darned in 2013: inclusion of model organisms and linking with Wikipedia. 2013;41(D1):D258–D261. doi:10.1093/nar/gks961.

118. RADAR: a rigorously annotated database of A-to-I RNA editing. 2014;42(D1):D109–D113. doi:10.1093/nar/gkt996.

119. Adenosine-to-inosine RNA editing meets cancer. 2011;32(11):1569–1577. doi:10.1093/carcin/bgr124.

120. ADARs: allies or enemies? The importance of A-to-I RNA editing in human disease: from cancer to HIV-1. 2012;87(1):95–110. doi:10.1111/j.1469-185X.2011.00186.x.

121. Functions and Regulation of RNA Editing by ADAR Deaminases. 2010;79(1):321–349. doi:10.1146/annurev-biochem-060208-105251.

122. Systematic identification of edited microRNAs in the human brain. 2012;22(8):1533–1540. doi:10.1101/gr.131573.111.

123. Redirection of Silencing Targets by Adenosine-to-Inosine Editing of miRNAs. 2007;315(5815):1137–1140. doi:10.1126/science.1138050.

124. Systematic identification of abundant A-to-I editing sites in the human transcriptome. 2004;22(8):1001–1005. doi:10.1038/nbt996.

125. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. 2004;2(12):e391. doi:10.1371/journal.pbio.0020391.

126. Hypothesis: RNA editing of microRNA target sites in humans? 2007;13(4):463–467. doi:10.1261/rna.296407.

127. Combinatorial microRNA target predictions. 2005;37(5):495–500. doi:10.1038/ng1536.

128. Accurate microRNA target prediction correlates with protein repression levels. 2009;10(1):295–295. doi:10.1186/1471-2105-10-295.

129. Specificity of microRNA target selection in translational repression. 2004;18(5):504–511. doi:10.1101/gad.1184404.

130. A combined computational-experimental approach predicts human microRNA targets. 2004;18(10):1165–1178. doi:10.1101/gad.1184704.

131. Fast and effective prediction of microRNA/target duplexes. 2004;10(10):1507–1517. doi:10.1261/rna.5248604.

132. miR-EdiTar: a database of predicted A-to-I edited miRNA target sites. 2012;28(23):3166–3168. doi:10.1093/bioinformatics/bts589.

133. Prediction of human targets for viral-encoded microRNAs by thermodynamics and empirical constraints. 2010;6(1):379–385. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20628498&retmode=ref&cmd=prlinks.

134. Local RNA base pairing probabilities in large sequences. 2006;22(5):614–615. doi:10.1093/bioinformatics/btk014.

135. Efficient use of accessibility in microRNA target prediction. 2011;39(1):19–29. doi:10.1093/nar/gkq768.

136. ViennaRNA Package 2.0. 2011. Available at: http://www.biomedcentral.com/content/pdf/1748-7188-6-26.pdf.

137. miRBase: integrating microRNA annotation and deep-sequencing data. 2011;39(Database issue):D152–7. doi:10.1093/nar/gkq1027.

138. Most mammalian mRNAs are conserved targets of microRNAs. 2009;19(1):92–105. doi:10.1101/gr.082701.108.

139. miRò: a miRNA knowledge base. 2009;2009(0):bap008–bap008. doi:10.1093/database/bap008.

140. Principles of microRNA-target recognition. 2005;3(3):e85–e85. doi:10.1371/journal.pbio.0030085.

141. Disease-Associated Mutations That Alter the RNA Structural Ensemble. 2010;6(8):e1001074–e1001074. doi:10.1371/journal.pgen.1001074.

142. Single-nucleotide polymorphisms can cause different structural folds of mRNA. 1999;96(14):7871–7876. doi:10.1073/pnas.96.14.7871.

143. MicroRNA-mediated regulation of gene expression is affected by disease-

associated SNPs within the 3′-UTR via altered RNA structure. 2012;9(6):924–937. doi:10.4161/rna.20497.

144. The role of site accessibility in microRNA target recognition. 2007;39(10):1278–1284. doi:10.1038/ng2135.

145. Regulation of MDM4. 2011;16:1144–1156. doi:10.2741/3780.

146. miR-Synth: a computational resource for the design of multi-site multi-target synthetic miRNAs. 2014;42(9):5416–5425. doi:10.1093/nar/gku202.

147. Interest resparks in RNAi. 2012;30(11):1012–1012. doi:10.1038/nbt1112-1012.

148. Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. 2010;464(7291):1067–1070. doi:10.1038/nature08956.

149. siRNA vs. shRNA: Similarities and differences. 2009;61(9):746–759. doi:10.1016/j.addr.2009.04.004.

150. Multitarget therapy of malignant cancers by the head-to-tail tandem array multiple shRNAs expression system. 2009;16(6):516–531. doi:10.1038/cgt.2008.102.

151. RNAi of neuropeptide Y (NPY) for neuropathic pain. 2005;(Neuroscience 2005 Abstract). Available at: http://www.sfn.org/Annual-Meeting/Past-and-Future-Annual-Meetings/Abstract-Archive/Abstract-Archive-Search.

152. SiRNA-mediated Inhibition of Vascular Endothelial Growth Factor Severely Limits Tumor Resistance to Antiangiogenic Thrombospondin-1 and Slows Tumor Vascularization and Growth. 2003;63(14):3919–3922. doi:10.1016/0092-8674(89)90238-9.

153. Intravenous RNA Interference Gene Therapy Targeting the Human Epidermal Growth Factor Receptor Prolongs Survival in Intracranial Brain Cancer. 2004;10(11):3667–3677. doi:10.1158/1078-0432.CCR-03-0740.

154. RNA Interference Reveals a Requirement for Myocyte Enhancer Factor 2A in Activity-dependent Neuronal Survival. 2002;277(48):46442–46446. doi:10.1074/jbc.M206653200.

155. Silencing Primary Dystonia: Lentiviral-Mediated RNA Interference Therapy for DYT1 Dystonia. 2005;25(45):10502–10509. doi:10.1523/JNEUROSCI.3016-05.2005.

156. Synthetic small inhibiting RNAs: Efficient tools to inactivate oncogenic mutations and restore p53 pathways. 2002;99(23):14849–14854. doi:10.1073/pnas.222406899.

157. Small interfering RNA (siRNA) inhibits the expression of the Her2/neu gene, upregulates HLA class I and induces apoptosis of Her2/neu positive tumor cell lines. 2004;108(1):71–77. doi:10.1002/ijc.11497.

158. Specific inhibition of bcr-abl gene expression by small interfering RNA. 2003;101(4):1566–1569. doi:10.1182/blood-2002-06-1685.

159. Enhancement of Hypoxia-Induced Tumor Cell Death In vitro and Radiation Therapy In vivo by Use of Small Interfering RNA Targeted to Hypoxia-Inducible Factor-1 . 2004;64(22):8139–8142. doi:10.1158/0008-5472.CAN-03-2301.

160. Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. 2005;23(6):709–717. doi:10.1038/nbt1101.

161. Current prospects for RNA interference-based therapies. 2011;12(5):329–340. doi:10.1038/nrg2968.

162. Multitarget drugs: the present and the future of cancer therapy. 2009;10(4):589–600. doi:10.1517/14656560902781907.

163. Targeting Multiple Signaling Pathways as a Strategy for Managing Prostate Cancer: Multifocal Signal Modulation Therapy. 2004;3(4):349–380. doi:10.1177/1534735404270757.

164. Silencing of HIV-1 with RNA Interference: A Multiple shRNA Approach. 2004;14(6):883–892. doi:10.1016/j.ymthe.2006.07.007.

165. Dual-targeting siRNAs. 2010;16(6):1275–1284. doi:10.1261/rna.2005710.

166. Interfering ribonucleic acids that suppress expression of multiple unrelated genes. 2009;9(1):57–57. doi:10.1186/1472-6750-9-57.

167. MicroRNAs: Target Recognition and Regulatory Functions. 2009;136(2):215–233. doi:10.1016/j.cell.2009.01.002.

168. Synthetic MicroRNA Designed to Target Glioma-Associated Antigen 1 Transcription Factor Inhibits Division and Induces Late Apoptosis in Pancreatic Tumor Cells. 2006;12(21):6557–6564. doi:10.1158/1078-0432.CCR-06-0588.

169. Desperately seeking microRNA targets. 2010. Available at: http://www.nature.com/nsmb/journal/v17/n10/abs/nsmb.1921.html.

170. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. 2007;27(1):91–105. doi:10.1016/j.molcel.2007.06.017.

171. Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments.

2014;42(D1):D926–D932. doi:10.1093/nar/gkt1270.

172. Rational siRNA design for RNA interference. 2004;22(3):326–330. doi:10.1038/nbt936.

173. Killing the messenger: short RNAs that silence gene expression. 2003;4(6):457–467. doi:10.1038/nrm1129.

174. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. 2010;11(8):R90–R90. doi:10.1186/gb-2010-11-8-r90.

175. Effect of target secondary structure on RNAi efficiency. 2007;13(10):1631–1640. doi:10.1261/rna.546207.

176. Sequence context outside the target region influences the effectiveness of miR-223 target sites in the RhoB 3'UTR. 2010;38(1):239–252. doi:10.1093/nar/gkp870.

177. Inducing Model Trees for Continuous Classes. 1997. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.9768.

178. Unbiased recursive partitioning: A conditional inference framework. 2006. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.92.9930.

179. miRTarBase: a database curates experimentally validated microRNA–target interactions. 2010;39(Database):gkq1107–D169. doi:10.1093/nar/gkq1107.

180. Conversion from the "oncogene addiction" to “drug addiction” by intensive inhibition of the EGFR and MET in lung cancer with activating EGFR mutation. 2012;76(3):292–299. doi:10.1016/j.lungcan.2011.11.007.

181. Human polymorphism at microRNAs and microRNA target sites. 2007;104(9):3300–3305. doi:10.1073/pnas.0611347104.

182. RNAi-based nanomedicines for targeted personalized therapy. 2012;64(13):1508–1521. doi:10.1016/j.addr.2012.08.014.

183. A review of nanocarriers for the delivery of small interfering RNA. 2012;33(29):7138–7150. doi:10.1016/j.biomaterials.2012.06.068.

184. Intravenous delivery of a multi-mechanistic cancer-targeted oncolytic poxvirus in humans. 2011. Available at: http://www.nature.com/nature/journal/v477/n7362/abs/nature10358.html.

185. Oral delivery of siRNA and antisense oligonucleotides. 2009;17(7):491–495.

doi:10.1080/10611860903057674.

186. Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation. 2009;458(7242):1180–1184. doi:10.1038/nature07774.

187. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. 2012;22(1):107–126. doi:10.1038/cr.2011.158.

188. Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. 2006;12(7):1179–1187. doi:10.1261/rna.25706.

189. 3|[prime]| UTR seed matches, but not overall identity, are associated with RNAi off-targets. 2006;3(3):199–204. doi:10.1038/nmeth854.

190. miRBase: annotating high confidence microRNAs using deep sequencing data. 2014;42(D1):D68–D73. doi:10.1093/nar/gkt1181.

191. Human microRNA targets. 2004. Available at: http://dx.plos.org/10.1371/journal.pbio.0020363.

192. miRDB: A microRNA target prediction and functional annotation database with a wiki interface. 2008;14(6):1012–1017. doi:10.1261/rna.965408.

193. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. 2006;126(6):1203–1217. doi:10.1016/j.cell.2006.07.031.

194. Gene Ontology: tool for the unification of biology. 2000;25(1):25–29. doi:10.1038/75556.

195. miR2Disease: a manually curated database for microRNA deregulation in human disease. 2009;37(suppl 1):D98–D104. doi:10.1093/nar/gkn714.

196. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. 2010;11(1):R6–R6. doi:10.1186/gb-2010-11-1-r6.

197. miRCancer: a microRNA–cancer association database constructed by text mining on literature. 2013;29(5):638–644. doi:10.1093/bioinformatics/btt014.

198. dbDEMC: a database of differentially expressed miRNAs in human cancers. 2010;11 Suppl 4(Suppl 4):S5–S5. doi:10.1186/1471-2164-11-S4-S5.

199. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. 2014;42(D1):D1070–D1074. doi:10.1093/nar/gkt1023.

200. The genetic association database. 2004;36(5):431–432. doi:10.1038/ng0504-431.

201. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. 2004;3:Article3–Article3. doi:10.2202/1544-6115.1027.

202. miRandola: Extracellular Circulating MicroRNAs Database. 2012;7(10):e47786–e47786. doi:10.1371/journal.pone.0047786.

203. Variability in the Incidence of miRNAs and Genes in Fragile Sites and the Role of Repeats and CpG Islands in the Distribution of Genetic Material. 2010;5(6):e11166–e11166. doi:10.1371/journal.pone.0011166.

204. Circulating miRNA Biomarkers for Alzheimer's Disease. 2013;8(7):e69807–e69807. doi:10.1371/journal.pone.0069807.

205. In vivo NCL targeting affects breast cancer aggressiveness through miRNA regulation. 2013;210(5):951–968. doi:10.1084/jem.20120950.

206. Genome-wide multi-omics profiling of colorectal cancer identifies immune determinants strongly associated with relapse. 2013;4:236–236. doi:10.3389/fgene.2013.00236.

207. Downregulation of HMGA-targeting microRNAs has a critical role in human pituitary tumorigenesis. 2012;31(34):3857–3865. doi:10.1038/onc.2011.557.

208. Gender-Specific Reduction of Estrogen-Sensitive Small RNA, miR-30b, in Subjects With Schizophrenia. 2012;38(3):433–443. doi:10.1093/schbul/sbq091.

209. miR in melanoma development: miRNAs and acquired hallmarks of cancer in melanoma. 2013;45(22):1049–1059. doi:10.1152/physiolgenomics.00116.2013.

210. The investigation of miR-221-3p and PAK1 gene expressions in breast cancer cell lines. 2014. doi:10.1016/j.gene.2014.11.036.

211. Prognostic value of miR-221-3p, miR-342-3p and miR-491-5p expression in colon cancer. 2014;6(4):391–401. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4113501/.

212. In-Depth Characterization of the MicroRNA Transcriptome in Normal Thyroid and Papillary Thyroid Carcinoma. 2013;98(8):E1401–E1409. doi:10.1210/jc.2013-1214.

213. Reductions in the expression of miR-124-3p, miR-128-1, and miR-221-3p in pediatric astrocytomas are related to high-grade supratentorial, and

recurrent tumors in Mexican children. 2014;30(7):1173–1181. doi:10.1007/s00381-014-2416-5.

214. Expression Profiles of Six Circulating MicroRNAs Critical to Atherosclerosis in Patients With Subclinical Hypothyroidism: A Clinical Study. 2014;99(5):E766–E774. doi:10.1210/jc.2013-1629.

215. miR-130a targets MET and induces TRAIL-sensitivity in NSCLC by downregulating miR-221 and 222. 2012;31(5):634–642. doi:10.1038/onc.2011.260.

216. La voie du TGFβ. In: Paris: Springer Paris; 2010:83–90. doi:10.1007/978-2-8178-0028-8_6.

217. MiR-221 influences effector functions and actin cytoskeleton in mast cells. 2011;6(10):e26133–e26133. doi:10.1371/journal.pone.0026133.

218. Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. 2014;13(1):28–28. doi:10.1186/1476-4598-13-28.

219. Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. 2008;82(2):283–289. doi:10.1016/j.ajhg.2007.09.021.

220. Profiling microRNA expression in hepatocellular carcinoma reveals microRNA-224 up-regulation and apoptosis inhibitor-5 as a microRNA-224-specific target. 2008;283(19):13205–13215. doi:10.1074/jbc.M707629200.

221. SomamiR: a database for somatic mutations impacting microRNA function in cancer. 2013;41(D1):D977–D982. doi:10.1093/nar/gks1138.