A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in

*Complex Systems for Life Sciences*

*XXXII ciclo*

February 20, 2020

# Investigating the interplay between epigenetic components and micro-RNAs: an approach to identify epi-miRNAs

**Elisa Reale**

Tutor: Michele Caselle

Co-Tutors: Matteo Osella, Francesca Orso


Physics Department
Università degli Studi di Torino
Italy

Università degli studi di torino

*Complex Systems for Life Sciences*

## Abstract

# Investigating the interplay between epigenetic components and micro-RNAs: an approach to identify epi-miRNAs

**Elisa Reale**

MicroRNAs (miRNAs) are small non-coding RNAs, involved in different physiological and pathological pathways, able to post-transcriptionally regulate gene expression. Beside their classical mechanism of action, miRNAs crosstalk with other regulatory genes such as the epigenetic factors, thus amplifying their regulatory potential. miRNAs involved in epigenetic regulation are denoted as epi-miRNAs. Growing evidence shows a strong interplay between post-transcriptional regulation, mediated by miRNAs, and epigenetic regulation. We propose a pipeline useful both to prioritize candidate epi-miRNAs and to identify candidate epigenetic interactors of any given miRNA starting from preexisting miRNA transfection experiments. Using our approach, we identified 34/59 candidate epi-miRNAs, 19 of them, as expected, were known epi-miRNAs; while 15 were new epi-miRNAs. Moreover, we identified EZH2, a component of the Polycomb repressive complex 2, as an interactor of miR-214, a well-known prometastatic miRNA in melanoma and breast cancer suggesting a miR-214-EZH2 regulatory axis potentially relevant in tumor progression. Furthermore, the interactions between the epigenetic components and the miRNAs that resulted candidate epi-miRNA from our pipeline were confirmed in the data from TCGA tumor samples.

# Acknowledgements

I would like to take this small space to express my sincere appreciation to all the people that made possible for me to follow the ambitious goal of surviving these three years. I would like to thank Professor Michele Caselle for the guidance, the knowledge and the wisdom with which he followed me during these years, and my co-tutors, Matteo Osella and Francesca Orso, for the support and suggestions both during the development of the project and the writing of the manuscript.

I would like to say thank you to my "long term" colleagues, il Nonno, Mattia, Mottes, Eleonora and David, for their help and patience and for their benevolent insistence on trying to make me a sociable person, and to the "short term" colleagues too, namely Mazzolini, Antonio Rosanova and Marta. Thanks to my PhD colleagues too, for their help (and for the beers). For what concerns my loved ones, well, they deserve to be thanked for a lot. And I will do it, in random order. To the Care Bears: they have been part of everything since forever, being supportive, "immoral and creepy" during every step of the way (since 1985, ed.). They need a full "power slide" (cit.) for this. An imaginary one, we are getting old. I would like to thank Amanda, Marika and Silvia for managing to stick together despite the "European scattering" we have suffered, and for sharing this fatigue with me. Thanks to my family, that has continued to be full of love (and loud) through every step of the way; facing every problem head on; supporting and motivating me.

Least, but not last, none (literally, nothing) of this would have been possible without my annoying, teasing, fantastic and lovely life partner, that kept me strong during every storm with a surreal patience and a Supernatural ("Sammyyy!!") ability to keep up with my many, many, many oddities. You must be a wizard, Flavio!

I have been cheesy enough, so let's move on. Thank you all.

# Publication acknowledgements

The 4th chapter, describing the project, has been adapted from Epigenomics (2019), 1750-1911, doi: 10.2217/epi-2019-0050 with permission of Future Medicine Ltd.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Summary and purpose of the work

The following text describes the research work carried out during the PhD program in Complex Systems for Life Sciences. Specifically, my project focused on improving the knowledge on a field that has gained increasing interest in the last years: the field of the miRNA - epigenetic component interactions. This field of research has grown in the recent years due to a growing amount of evidence showing the importance of the interplay between miRNAs and the epigenetic components' layers of regulation. During the years this interplay resulted so deep that it has been recently suggested that it is only by viewing epigenetic regulators (from here on, epi-R) and miRNAs as associated in a combined regulatory network that it would be possible to truly understand epigenetic regulation (Gruber and Zavolan 2013). Several miRNAs have been identified until now as regulators of epigenetic components belonging to several different pathways, and many of these miRNAs are also considered to be under epigenetic control.

Despite the increasing importance of the topic, epi-miRNAs are still sporadically discovered, and no attempt has been made since now to perform a systematic search. Our project has been developed to tackle this issue and start to shed light on the true magnitude of the epi-miRNA interactions. Therefore, we developed a pipeline able to both detect candidate epi-miRNAs and, given a miRNA of interest, to identify its putative epigenetic interactors.

This pipeline is designed to investigate epi-miRNAs starting from gene expression data derived from miRNA transfection experiments. The overall main goal of this work is to fill the gap about the importance of the epi-miRNA interplay and through this analysis we have been able to investigate:

- if it is possible to search for these interactions exploiting only gene expression data from miRNA transfection experiments;
- the degree and amount of the epi-miRNA interactions;
- if it is possible to identify putative circuits in which these two components, the miRNA and the epigenetic factors, are involved.

The pipeline presented here exploits data available online to search for interactions between miRNAs and epigenetic components, addressing four main epigenetic pathways: PRC1, PRC2, HDAC and DNMT. Through this pipeline we identified a large number of candidate epi-miRNAs, we investigated and detected the epigenetic interactors of a specific miRNA of intereset, miR-214, and highlighted epi-miRNA tendency to act in a cell-type specific manner.

A similar behaviour arbours also from the investigation of the interplay between the candidate epi-miRNA resulting from our pipeline and the epigenetic components in cancer tissue samples. Several miRNAs are indeed known to be dysregulated in cancer, where they can act either as oncogenes (oncomiRs) or as oncosoppressors. Therefore, investigating miRNAs' interactions with the epigenetic complexes in this pathological context may be important to identify key epi-miRNA - epigenetic components axes in cancer.

The following dissertation is therefore arranged as follows: the following chapter will introduce the two main topics addressed in this work, epigenetics and miRNAs, summarizing the relevant literature and the main pathways involved in these two classes of interest. This introduction is fundamental to understand the complexity underlying the interaction between miRNAs and epigenetic components, which will be described in the third chapter. In the fourth chapter, the analysis and the results of the project will be described, while we will list the conclusions, critics and future perspectives in the last chapter.

# Chapter 2

# Introduction

## 2.1 Epigenetics

### 2.1.1 Introduction to epigenetics

The term "epigenetics" was first introduced by Waddington (Waddington 2011) in 1942, describing the study of the network involved in development mechanisms that lies in the middle of genotype and phenotype (Waddington 2011). At the time it was clear that there must be some regulatory network leading to the different phenotypes observed starting from the same genotype, but the mechanism through which this happened was unclear. Epigenetics definition given by Robin Holliday in 1990 was slightly more precise, describing it as the study of the processes that regulate gene activity both temporally and spatially over the development of an organism (R. Holliday 1990), practically describing epigenetics as every phenomenon that can affect DNA transcription, and thus more broadly the development of an organism, without affecting the DNA sequence. This definition portrays epigenetics as a complex network of sophisticated machineries, able to adapt chromatin, stabilize gene expression and give directions to cell identity (C. David Allis and Jenuwein 2016).

Some other stricter definitions have been used further, describing epigenetics in different ways, until the advent of the "Roadmap epigenomics project" of NIH which established the current epigenetics definition including both those alterations of gene activity that could be inherited and modify gene activation and expression levels, and those that are durable and affect the transcriptional potential of a cell, but are not inheritable (*Roadmap Epigenomics Project* n.d.), thus broadening and widening the concept of epigenetics to a more open and general definition, under which many mechanisms and pathways could fall.

As for many other bio-molecular fields, the advancements in epigenetics followed the technical advancements brought by the development of efficient analysis tools that made possible to identify specific modifications at single base-pair resolution, both at the DNA and histone levels. Furthermore, the advent of the next-generation sequencing techniques broadened the possibil-

ities, leading to the development of top level technologies and allowing to study the genome and the epigenome at a wider degree.

With these technologies many steps forward have been made and, from the epigenetic point of view, these steps led to a better understanding of the components that coordinate epigenetic control, their interactions and how they fine-tune many aspects of the cell life, from differentiation to pathological processes such as cancer.

### 2.1.2 Chromatin and nucleosomes

At the very basis of epigenetics there is the chromatin. Chromatin is the structure resulting from the DNA sequence rolling up around cylindrical molecules called histones. Histones contain four different core histone proteins (H2A, H2B, H3 and H4): each of them is present in two copies which assemble in octamers (Luger et al. 1997). When a tetramer composed by the two copies of both H3 and H4 create a complex with two H2A/H2B dimers to form cylinders around which DNA is able to fold, the octamer is completed: this structure composed of histones and DNA is named nucleosome, the base unit of chromatin, and a graphic model can be found in Figure 2.1. There is a fifth histone, named histone H1: it binds to the nucleosomes and protects the free linker DNA (20 bp) between the nucleosomal core particles (Brockers and Schneider 2019). Histones can be found also in different variants: for example, H2A has four main variants (H2A.Z (Talbert and Steven Henikoff 2010), H2A.X (Morrison and Shen 2005), H2A.B (Ishibashi et al. 2010) and macroH2A (Chakravarthy et al. 2005), with two isoforms); histone H3 has one variant, H3.3 (Malik and Steven Henikoff 2003). Every variant can have epigenetic consequences, because the replacement of a canonical histone by a noncanonical variant can affect the composition of chromatin (S. Henikoff and M. M. Smith 2015). These structures can be packed at different levels and can have different histone composition, and can result in chromatin categorized as heterochromatin or euchromatin.

The nucleosome is the essential unit of a eukaryotic chromosome, and it is composed of 146 base pairs (bp) of DNA coiled around the histone octamer described before (Luger et al. 1997). The more these structures are packed, the more the chromatin become inaccessible, as depicted in Figure 2.2. This type of chromatin is referred to as heterochromatin: this is highly condensed, gene-poor, strongly characterized by large, contiguous and repetitive DNA domains, transcriptionally silent.

Heterochromatin can be either constitutive or facultative. The first one is always present at the centromeres and telomeres of the chromosomes, which are regions that need an higher stability because of their importance in genome integrity. It never changes through the cell cycle and it is identical in every cell of the organism. Despite the condensed chromatin state, there are hints suggesting that transcription is possible also at these genomic regions and in certain circumstances may be necessary for the establishment of the heterochromatin itself, as shown in Drosophila (C. D. Smith et al. 2007).

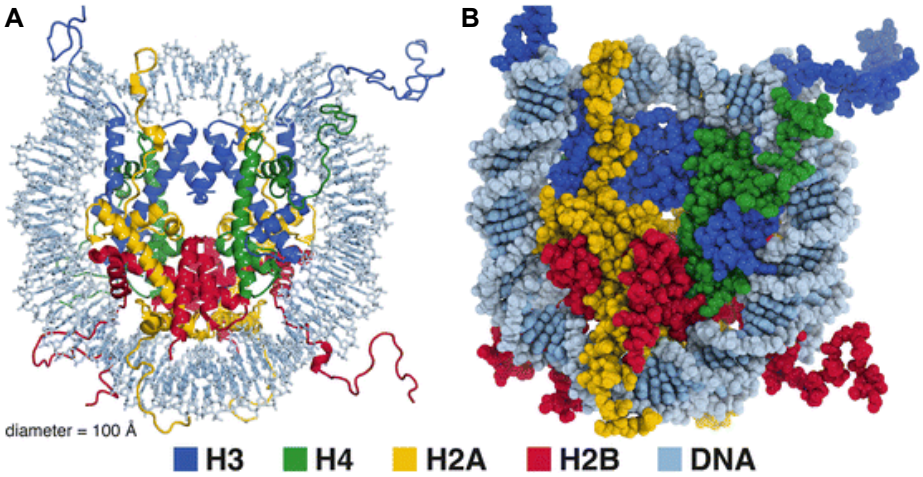



Figure 2.1: 3D model of a nucleosome. The color coding indicates the different molecules composing it. Panel **A** refers to the structure of the molecules, while panel **B** describes the bulk spacing. *Image from (McGinty and Tan 2014).*

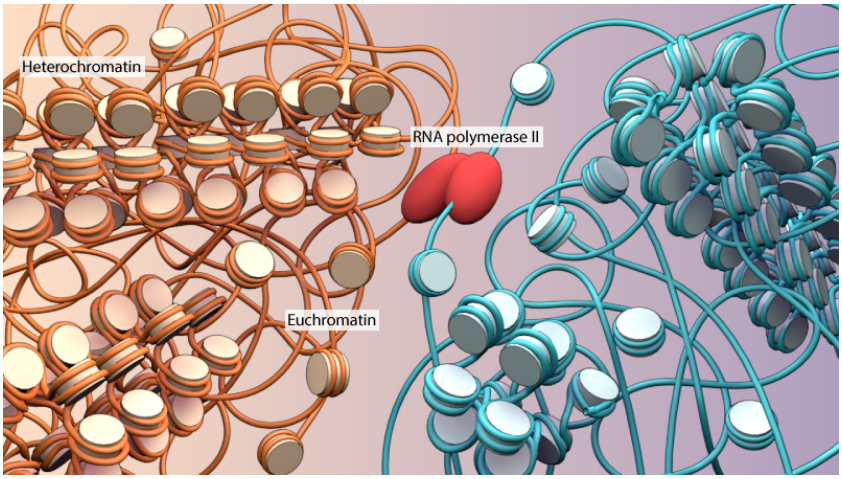



Figure 2.2: Heterochromatin vs. Euchromatin. Heterochromatin consists in more packed nucleosomes, while euchromatin has looser nucleosomes, allowing the access to RNA polymerases. *Image from (*What is chromatin, heterochromatin and euchromatin? *2018).*

However this happens at very low rates, not matching that of protein-coding genes in euchromatin (Trojer and Reinberg 2007).

The facultative heterochromatin is also transcriptionally silent, but it is kept in a bistable state, retaining the potential to interconvert between heterochromatin and euchromatin depending on three possible contexts: temporal (e.g. cell-cycle), spatial (e.g. nuclear organization changes) or parental/heritable (e.g. mono-allelic gene expression) (Trojer and Reinberg 2007). It is widely accepted that the establishment and/or maintenance of a facultative heterochromatin state is achieved by a combination of different processes. These events include incorporation of specific/alternate chromatin components, common and organized action of trans-acting factors, chromatin modulation and the specific sub-nuclear localization (Trojer and Reinberg 2007). It also has a major role in the X-chromosome inactivation in

mammals: the formation of facultative heterochromatin in this case is mediated by the long non-coding RNA Xist and promotes transcription of young LINEs (Long interspersed nuclear elements), which may therefore simplify the X chromosome inactivation at different levels (Chow et al. 2010).

While heterochromatin is higly packed and generally transcriptionally silenced, euchromatin comprises those sections of chromosomes, generally the greater part, that carry out normally their process of decondensation during the last phase of mitosis, and it is therefore in a decondensed form during interphase. Euchromatin is also characterized by a high number of genes, and it is more easily transcribed because of a greater accessibility to the genes given by its lower level of condensation (Tamaru 2010), and therefore higly present in cells undergoing active transcription. The accessible chromatin structure of euchromatin is achieved by specific histone modifications: high levels of histone acetylation, and methylation at H3K4 and H3K79. More details about these and other modifications will be listed in the next section.

### 2.1.3 Main epigenetic modifications

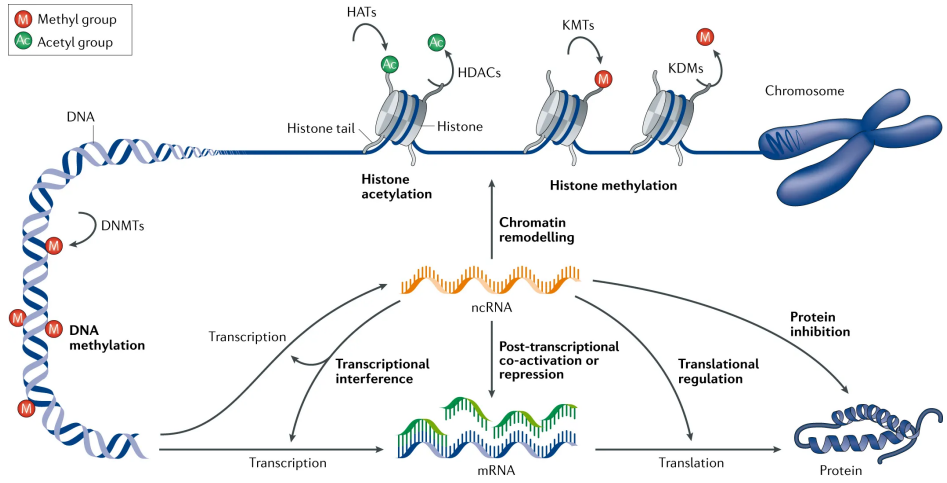



Figure 2.3: Overview of the main epigenetic pathways. *Image from (Joosten et al. 2018).*

Chromatin is not only characterized by its levels of condensation, but also by the interacting proteins and specific modifications. These modifications can act on the chromatin condensation state, leading to a more open or a more closed conformation, but they can also act as recruiters for specific epigenetic complexes that can affect the downstream pathways in many different ways. Thus far, many epigenetic modifications have been identified, like methylation, phosphorylation, acetylation, ubiquitylation, and sumoylation, that can affect the DNA (DNA methylation) or the histones; a representation of the main epigenetic pathways is shown in Figure 2.3. Each of these modifications is generally associated with a set of proteins which are usually categorized into three main groups:

- the writers are the enzymes capable of modifying nucleotides and specific amino acid residues on histones;

- the erasers are enzymes appointed to remove these marks;
- the readers, which belong to a category of proteins with specific domains that can recognize the different epigenetic marks.

Next, a brief description of the main epigenetic modifications will be provided, with particular attention to those that will be the focus of the project.

### DNA methylation

Among the known epigenetic modifications, the one that attracted most of the interest during the years is DNA methylation, and it is well known that it, cooperating with other regulators, results one of the primary epigenetic factors influencing gene activities (Portela and Esteller 2010). The first modified cytosine was discovered in 1948 by Rollin Hotchkiss (Hotchkiss 1948), while analyzing calf thymus using paper chromatography. He believed that the modified portion of cytosines that resulted from his analysis was 5-methylcytosine (5mC) because it separated from cytosine in a way that was similar to methyluracil (another name for thymine) separating from uracil, suggesting that this methylcytosine was a modification naturally present in DNA. Later on, many researches suggested that DNA methylation might take part in the regulation of gene expression, but only in the '80s a considerable number of studies finally proved that it exerted this role both in gene regulation and in cell differentiation (Portela and Esteller 2010). One of the scientist making this discovery was R. Holliday in 1975 (Holliday and Pugh 1975), one of the fathers of epigenetics mentioned before.

Now we know that DNA methylation consists in a covalent transfer of a methyl group from S-adenosyl methionine to the C-5 position in CpG dinucleotides in animals (Robertson 2005), in regions which are generally cytosine rich (Weinhold 2006).

Given the aforementioned roles of methylation, it is clear that its presence on CpG dinucleotides has a central role in mammals, both for development and differentiation; but even if because of this fundamental role the largest part of the CpG dinucleotides is usually methylated, there are exceptions involving specific DNA segments, usually 0.5–2 kb in length, which are, instead, generally unmethylated (Cooper et al. 1983). These cytosine rich areas are generally called "CpG islands". These are non-methylated DNA sequences which have an elevated G + C content and little CpG suppression, covering almost 60–70% of all human genes, and serving as major binding sites for activating histone modifiers (Illingworth and Bird 2009; S.-M. Lee et al. 2017).

DNA methylation can be sometimes associated to transcription activity too, and this happens when this modification happens inside the gene body, which is a characteristic of those genes that are expressed ubiquitously: one of the suggested hypothesis is that this phenomenon could be connected to elongation performance and could help preventing deceptive transcription initiation (Portela and Esteller 2010).

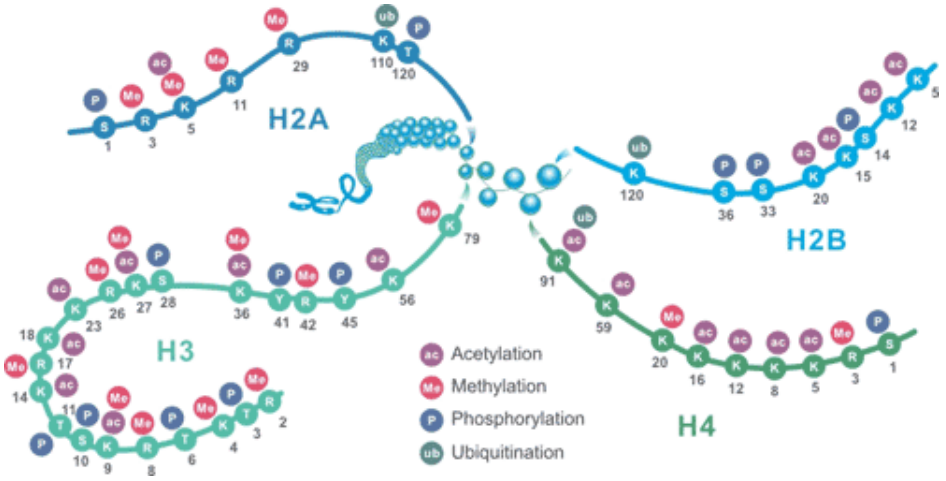



Figure 2.4: Principal histones modifications. Overview of the principal histones modifications, and their location on the histones' tails. *Image adapted from (*Four Common Histone Modifications *n.d.)*

The mammalian DNA methylation is always achieved through the DNA methyltransferases (DNMTs), whether it leads to gene silencing (more common) or to gene activation (more rarely). These enzymes are in charge for the settlement and preservation of the patterns of DNA methylation, while the Methyl-CpG Binding Proteins (MBDs) are those responsible of reading these modifications. DNA methylation inhibits transcription both in a direct way, by inhibiting transcription factors attachment on the DNA, and in an indirect way, through the recruitment of MBDs with repressive chromatin remodelling abilities (Robertson 2005).

DNMTs are generally divided into *de novo* DNMTs (DNMT3A and -3B) and maintenance DNMTs (DNMT1). The firsts are believed to be in charge of the establishment of the methylation patterns during development of the embryo, and in fact their expression levels are high in Embryonic Stem (ES) cells, while they are low in differentiated cells (Portela and Esteller 2010). DNMT1, instead, has a 30/40-fold propensity for hemi-methylated DNA, and it owns de novo DNMT activity too. DNMT1 is the DNA methyltransferase that is present in the highest quantity in the cell. Usually, its role consists in methylating hemi-methylated DNA generated after DNA replication (Portela and Esteller 2010).

Still, the division between *de novo* and maintenance methylation is usually fuzzy, and an update of the model proposes that DNMT3A and DNMT3B role is to methylate the sites neglected by DNMT1 at the replication fork. One other member of the DNMT group of enzymes is DNMT2, which contains all the catalytic signature motifs of standard DNMTs, but cannot be categorized as the other DNMTs because it has almost no DNA methyltransferase activity (Portela and Esteller 2010), but has been demonstrated to have tRNA methyl-transferases activity instead (Goll et al. 2006).

### Histone modifications

Histones are proteins that assemble in a complex that is composed by an heterotypic tetramer, composed of two copies of the H3–H4 dimer, that forms a complex with two H2A–H2B dimers (Mariño-Ramírez et al. 2005). The resulting complex is a cylindrical octamer, which is separated by its neighboring nucleosomes by a mean of 50 bp of loose DNA (Portela and Esteller 2010) as previously shown in Figure 2.1.

The core histones are mostly globular apart from their N-terminal tails, that result unstructured. The tails can undergo many post-transcriptional modifications like acetylation, methylation, phosphorylation, ubiquitination and sumoylation, among others, and an overview of the principal histones modifications is shown in Figure 2.4. These modifications have crucial roles in several processes like regulation of transcription, DNA repair and replication, alternative splicing and chromosome condensation. For instance, different chromatin states are characterized by different modifications: euchromatin is usually highly acetylated and presents the H3K4me3, H3K36me3 and H3K79me3 histone marks while heterochromatin is usually poorly acetylated and presents high degree of H3K9me, H3K27me and H4K20me (Portela and Esteller 2010). Analyzing the levels of these modifications it is possible to predict gene expression: for instance, studies have proved that genes under active transcription are defined by high levels of tri-methylated H3K4, acetylated H3K27 and H2BK5 and methylated H4K20 in the promoter and methylated H3K79 and H4K20 along the gene body (Karlic et al. 2010).

Each histone can undergo several modifications, that can also act in a combinatorial way, and it is common to spot a cross-talk between the different histone marks. This communication can happen within the same location or in the same histone tail, but also through different histone tails (Strahl and C. D. Allis 2000). In the end, the overall result of these modifications is due to the combined effect of all the marks on a specific nucleosome or genomic region, complessively acting as a proper "histone code" (Strahl and C. D. Allis 2000).

A huge number of enzymes catalyzing the covalent post-transcriptional modifications on the histones have been reported, together with enzymes that remove these dynamic post-transcriptional modifications. Nonetheless, the list of histone modifiers is far from being exhaustive. Among the enzymes capable of modifying histones we can mention histone methyltransferases and demethylases, kinases and ubiquitin ligases that are the most specific, and the histone acetyltransferases and deacetylases that are less specific and modify different residues (Portela and Esteller 2010). In the next paragraphs, a brief description of the best known histone modifications and of their writers will be provided.

- **Acetylation**. Levels of histone acetylation have a fundamental part in redesigning the chromatin structure and in gene transcription regulation. Histone acetylation happens mostly on the lysines. Lysine acetylation on histone tails causes the chromatin to be in a relaxed state

and therefore leads to gene-transcription activity. When lysines are deacetylated, the chromatin is more condensed, leading to deactivation of gene transcription (Ropero and Esteller 2007). Lysine acetylation is performed by HATs, Histone Acetyltransferases, that are therefore mainly linked to transcriptional activation, while Lysines deacetylation is performed by HDACs, the Histone Deacetylases, whose function is to give chromatin a fresh start by erasing acetylation at active genes. Using acetyl-CoA as co-factor, HATs are able to transfer an acetyl group to the $\epsilon$-amino group of lysine side chains. In this way the positive charge present on the lysines is neutralized and this may lead to the weakening of the interaction between histones and DNA (Andrew J. Bannister and Tony Kouzarides 2011). The targeted lysines are usually on conserved sites in histone H4 and H3, but they can also be found on less conserved sites in histones H2A and H2B (Kurdistani et al. 2004).

HATs are categorized into two classes, HAT A and HAT B. This classification depends on their mechanism of catalysis and on their localization in the cell. The ones that belong to the HAT A family are usually located in the nucleus, and are those acting on histone N-tails. The HAT A family comprises three subclasses, and the division depends on their homology with yeast proteins. The proteins belonging to the HAT B class are instead located in the cytoplasm and their role is to transfer the acetyl group from Acetyl-CoA to an $\epsilon$-NH2 group of free histones before they are deposited on the DNA (Peserico and Simone 2011).

HDAC enzymes have the opposite effect of HATs, reversing lysine acetylation and restoring its positive charge. This is the reasons why HDACs are generally transcriptional repressors: the increased ionic interactions between histones, that have a positive charge caused by histone deacetylation, and negatively charged DNA, leads to a more dense chromatin structure and this limits the possibilities for the transcription machinery of having access to genes (Ropero and Esteller 2007). So far, eighteen mammalian HDAC enzymes have been described. They can be classified in different groups: Class I consists of HDACs 1, 2, 3, and 8 and is located in the nucleus. This class can be ubiquitously found in human cell lines and tissues. Class II comprises HDACs 4, 6, 7, 9 and 10; it has tissue-specific expression and has the ability to shuttle among the nucleus and cytoplasm. The class III HDACs is that of the sirtuins (SIRT1-7). The organization inside the cell and the tissue-specific pattern of expression of this class have not yet been discovered. Differently from Class I and II, class III require the coenzyme NAD+ as a cofactor (Ropero and Esteller 2007). HDAC11 has been discovered a couple decades ago and is the only component of the class IV HDACs, homologous with both class I and class II (L. Gao et al. 2002).

Many HDACs are part of large complexes composed of multiple pro-

teins, like the transcriptional co-repressors mSin3, N-CoR, and SMRT. These complexes interact with DNA binding factors that recruit them to definite regions on the genome; the DNA binding factors consists of transcription factors, nuclear receptors, MBDs, DNMTs and HMTs above all (Glass and Rosenfeld 2000). The interaction that was best described is that of methyl-binding proteins recruiting HDACs to methylated DNA: one example of this kind of interaction is that of MeCP2. MeCP2 is a methyl-binding protein and its role is to recruit complexes containing HDAC to methylated gene promoters to induce gene-transcription repression (Jones et al. 1998).

- **Phosphorylation**: Histone phosphorylation is strongly dynamic, it can happen on serines, threonines and tyrosines, mostly in their N-terminal histone tails. The phosphorylation mechanism is involved in many pathways and they are not necessarily related to gene transcription regulation. It has instead been associated with many cellular processes, including cell death, cell cycle progression, DNA repair, chromosome condensation, and development. One example of histone phosphorylation that regulates gene transcription is the phosphorylation of Histone 3 at threonine 11 (H3T11) by PRK1 (protein kinase C-related kinase 1 ), that is one of the known epigenetic marks for gene activation, whose regulation may be achieved through interaction with other epigenetic modifiers (Metzger et al. 2008). These kind of modifications are controlled by kinases and phosphatases: kinases are responsible for phosphorylation, while phosphatases are in charge of reversing this modification. Histone kinases are enzymes able to transfer a phosphate group from ATP to the hydroxyl group of the aminoacid side chain: this adds significant negative charge to the histone. In most of the cases, it is not clear how the enzyme is specifically recruited to its target site. There is a small number of instances, represented by the MAPK1 enzyme, in which the kinase owns an intrinsic DNA-binding domain. Alternatively, recruiting a kinase may need the cooperation with a factor bound to chromatin before it directly interacts with the DNA. Very few is known about the activity of histone phosphatases, but since the turnover of these modifications is extremely rapid, it is clear that there must be a strong phosphatase activity inside the nucleus (Andrew J. Bannister and Tony Kouzarides 2011).

- **Ubiquitination**: The first protein to be described as modified by ubiquitin in cells is histone H2A. The most abundant ubiquitinated proteins in the nucleus to be identified since now are H2A and H2B. In addition to them, H3, H4 and the linker histone H1 have been later described as potentially ubiquitinated as well (Cao and Yan 2012).

  There are two main forms of ubiquitinated histones, which are monoubiquitinated H2A (H2Aub) and H2B (H2Bub). This modification consists in an individual molecule of ubiquitin linked to two specific lysine residues: Lys-119 for H2A, and Lys-120 for H2B in vertebrate. Exper-

iments of Chromatin immunoprecipitation (ChIP) proved that H2Aub is present at high levels in the satellite regions of genome; H2Bub, instead, can be specifically found in the gene body of genes under active transcription (Cao and Yan 2012). Ubiquitination is performed by a subset of proteins belonging to the family of Polycomb group proteins that form the Polycomb Repressive Complexes. The first Polycomb Repressive Complex has been isolated from Drosophila embryos and was termed PRC1. Mammalian cells contain similar complexes, except for the presence of additional subunits. The PRC1 complex owes its ubiquitin ligase ability to the proteins containing a RING domain. The first ubiquitin ligase (E3) to be described is RING1B and is held to monoubiquitinate H2A on lysine 119 (H. Wang et al. 2004); H2A mediated mono ubiquitination is required to achieve gene silencing through the action of polycomb (Cao and Yan 2012). RING1A, together with BMI1 (another RING domain containing protein) strongly stimulate the E3 ubiquitin ligase activity of RING1B, and together exert the ubiquitin ligase activity of the PRC1 (Buchwald et al. 2006). PRC1 main silencing mechanism consists in inhibiting RNA polymerase II (RNA Pol II)-activated transcription on chromatin templates. Impaired transcription depends on the preassociation of PRC1 complexes with the template, which also interferes with chromatin remodeling. Moreover, transcription inhibition may also result from the ability of PRC1 complexes to compact oligonucleosomes (Vidal and Starowicz 2017). H2A and H2B ubiquitination can be reversed by ubiquitin specific peptidases known as deubiquitinating enzymes (DUBs), of which many different versions have been described, and this suggests that they may display redundancy of functions or act according to their context (Cao and Yan 2012).

- **Methylation**: Histone methylation and its implication in transcription were first described in the sixties, but only 15 years ago the first histone methyltransferase was identified, starting the discovery of numerous histone methyltransferases.

  Histone methylation mainly takes place on the side chains of lysines and arginines, and does not lead to the alteration of the histone proteins' charge, but it is possible that methylation have effect on chromatin structure directly altering its ability to fold (Martin and Yi Zhang 2005). Lysines may be mono-, di- or tri-methylated, while methylation on the arginines may be single or double, symmetrical or asymmetrical (A. J. Bannister and T. Kouzarides 2011).

  - **Lysine methylation**. Lysine methylation is present on different sites on the histones and its levels are accurately controlled by the action of methyltransferases and demethylases. Histone lysine methylations can activate or repress transcription according to their locations and levels of methylation (mono-, di- or

tri-methylation), which can also influence recognition by effector proteins.

In general, H3K4, H3K36 and H3K79 methylations indicate active transcription, while H3K9, H3K27 and H4K20 methylations are repressive (Hyun et al. 2017). Gene expression is also influenced by the location of histone methyl-lysine residues inside the genes and their methylation levels. For example, H3K4me1/me2/me3 are known to mark actively transcribed genes in different ways: H3K4me1 levels are high at enhancers while H3K4me2 levels are higher near the 5' end of transcribed genes; H3K4me3 is a mark of the promoters of both actively transcribed and poised genes. The methylation of lysine residues does not lead to alteration of charge, so the direct effect of lysine methylation on the structure of chromatin may occur through a non-electrostatic process, such as the hydrophobic interactions (Martin and Yi Zhang 2005).

Lysine methylation is performed by the Histone Lysine Methyltransferases (HKMTs), that transfer the methyl group from the cofactor adenosylmethionine (AdoMet), which brings a highly reactive methylthiol group to the nitrogen of the substrate lysine residue, producing, as observed, mono-, di-, or tri-methylated final products (Q. Liu and M. W. Wang 2016).

**H3K4** methylation is an evolutionary conserved histone mark, which is generally associated to active gene transcription. In yeast, methylations on H3K4 are achieved by Set1 methyltransferase; Drosophila melanogaster contains three Set1 homologs, whereas mammals have six (Gu and M. G. Lee 2013). All the Set1 homologs operate as scaffold proteins within the methylation complexes, associating with four subunits common to all complexes, which are fundamental for the enzyme activity of methyltransferases, and varying unique subunits involved in distinct functions that may have important roles in recruitment and integrating additional histone-modifying abilities (Gu and M. G. Lee 2013).

**H3K9** methylation, instead, is an histone mark that is usually associated to silenced transcription and a closed heterochromatin structure. In mammalian cells H3K9 methylation is performed by different methyltransferasis, which have all distinct catalytic abilities and target genes, underlying their being involved in different phenomenons inside the cells (Hyun et al. 2017).

Both tangible and functional interactions have been reported several times among H3K9 methyltransferases and DNA methyltransferases (DNMTs), suggesting that these two pathways can interact (Hyun et al. 2017). For instance, SUV39H1 and -H2 and DNMT3A and -3B cooperate and can be engaged through their association with HP1 (heterochromatin protein 1) to the methylation of regions of constitutive heterochromatin with high levels

of H3K9, causing a reinforcement of the condensation state of chromatin.

**H3K27**, in its tri-methylated form, is one of the most famous hallmarks of transcriptional repression. Specifically, the methyltransferase activity is due to the EZH2 (enhancer-of-zest homolog 2) subunit of the PRC2 (polycomb repressive complex 2) complex, an evolutionarily conserved class of polycomb group proteins, which is responsible for the different levels (mono-, bi-, trimethylated) of H3K27 methylation. The mammalian PRC2 complex comprises four fundamental subunits: EZH1 and -2, SUZ12, EED and RbAp46 and -48. EZH1/2 is the catalytic subunit, but alone it has no enzyme activity, which is gained through the incorporation within a PRC2 complex where the association with other subunits triggers it to H3K27 methylation. SUZ12 has instead the role of binding the histone H3 N-terminal tail, while EED binds to tri-methylated H3K27. Therefore, a positive feedback mechanism is used to extend H3K27me3-repressive marks to neighboring gene loci (Mierlo et al. 2019).

– **Arginine methylation.** Arginine methylation happens on proteins both in the nucleus and in the cytoplasm, and it is highly present on those proteins that function as shuttles. In mammals, arginine methylation is a very common modification, that appears at levels similar to those of phosphorylation and ubiquitination. It is carried out by the family of Protein Arginine Methyltransferases, or PRMT family, which is composed of nine members, but many others molecules may be involved in the establishment of this modification (Blanc and Richard 2017). PRMTs mechanism of action is based on transferring a methyl group from S-adenosylmethionine (AdoMet) to a guanidino nitrogen of arginine: this results in S-adenosylhomocysteine (AdoHcy) and methylarginine (Di Lorenzo and Bedford 2011). Differently from other histone writers, arginine methyltransferases are known to interact with many other substrates involved in different biological processes including transcription, cell signaling, mRNA translation, receptor trafficking, protein stability, DNA damage signaling, and pre-mRNA splicing (Blanc and Richard 2017). Arginine methylation role in gene regulation is related to the ability of the PRMTs to deposit either active or repressive histone marks. For example PRMT1, which is responsible for about 85% of total protein arginine methylation activity, methylates histone H4 at arginine 3, generating H4R3me2a. This modification acts as a transcriptional activation mark, which can recruit MBPs and affect the writing and erasing of the epigenetic modifications in the proximity. On the other hand, PRMT6, that is predominantly localized in the nucleus, is the primary enzyme in charge for H3R2 methylation in mammals and its modifications coun-

teract the H3K4me3 activation mark, making it a transcriptional repressor. PRMT5 and PRMT7 modify the arginines on histone 3 and 4, respectively, forming H3R8 and H4R3, resulting in transcriptional repression (J. Zhang et al. 2019). Contrarily from the other histone modifications, arginine methylation is a relatively new topic, whose importance is rapidly increasing, framing it as a growing area that is very promising regarding the possibility of expanding our understanding in many biological and pathological mechanisms.

The first ever identified histone demethylase is LSD1 (lysine-specific histone demethylase 1, also known as KDM1A), a FAD (flavin adenine dinucleotide)-dependent nuclear amine oxidase, which is responsible for H3K4 demethylation. Its identification ended the belief of histone methylation being stable and inheritable, clarifying that it doesn't turn over more slowly than other histone modifications, as it was thought. The identification of histone demethylases made clear that, similarly to acetylation, histone methylation is also a dynamic process, capable of being written, read and erased too (Shi et al. 2004)

## 2.2 miRNAs

### 2.2.1 Introduction to non-coding RNAs

RNA is mainly known as the messenger (mRNA) of the genetic information in the cell. Beside this role, though, RNA can be found in many other forms. Part of the RNA molecules that do not have a messenger role have instead non-coding roles, and can be categorized in the field of the non-coding RNAs.

Non-coding RNAs (ncRNAs) are known to take part in a number of different processes: these processes include regulation of transcription, DNA replication, RNA processing and modification, mRNA stability and translation, and protein degradation and translocation (Storz 2002). There are many types of ncRNAs and reviewing all of them is way beyond the purpose of this dissertation, but a quick overview is depicted in Figure 2.5.

Given the high number of different ncRNAs, it is not difficult to imagine that they cover plenty of functions, which are summarized in Fig. 2.6. Some ncRNAs affect transcription and chromosome structure and have a demonstrated role in mechanisms of gene silencing and of chromatin structure modification over large sections of the chromosomes. One archetypal example is represented by the human Xist RNA, a non-coding RNA which is necessary for the process of inactivation of the X chromosome (Brockdorff et al. 1991). In this process, the Xist is generated by the inactive X chromosome and then expands in *Cis* over the chromosome. In this case, it has been hypothesized that the association between the RNA and the chromosome is able to engage proteins that have effect on the chromatin structure, as for the case of PRC1 (Schoeftner et al. 2006) and PRC2 complexes (Rocha et al.

| Type | Percent of total RNA by mass | Molecules per cell | Average size (kb) | Total weight picograms/cell | Notes | Reference |
|---|---|---|---|---|---|---|
| rRNAs | 80 to 90 | 3–10 × $10^6$ (ribosomes) | 6.9 | 10 to 30 | | Blobel and Potter (1967), Wolf and Schlessinger (1977), Duncan and Hershey (1983) |
| tRNA | 10 to 15 | 3–10 × $10^7$ | <0.1 | 1.5 to 5 | About 10 tRNA molecules /ribosome | Waldron and Lacroute (1975) |
| mRNA | 3 to 7 | 3–10 × $10^5$ | 1.7 | 0.25 to 0.9 | | Hastie and Bishop (1976), Carter et al. (2005) |
| hnRNA (pre-mRNA) | 0.06 to 0.2 | 1–10 × $10^3$ | 10* | 0.004 to 0.03 | Estimated at 2–4% of mRNA by weight | Mortazavi et al. (2008), Menet et al. (2012) |
| Circular RNA | 0.002 to 0.03 | 3–20 × $10^3$ | ~0.5 | 0.0007 to 0.005 | Estimated at 0.1–0.2% of mRNA** | Salzman et al. (2012), Guo et al. (2014) |
| snRNA | 0.02 to 0.3 | 1–5 × $10^5$ | 0.1–0.2 | 0.008 to 0.04 | | Kiss and Filipowicz (1992), Castle et al. (2010) |
| snoRNA | 0.04 to 0.2 | 2–3 × $10^5$ | 0.2 | 0.02 to 0.03 | | Kiss and Filipowicz (1992), Cooper (2000), Castle et al. (2010) |
| miRNA | 0.003 to 0.02 | 1–3 × $10^5$ | 0.02 | 0.001 to 0.003 | About $10^5$ molecules per 10 pg total RNA | Bissels et al. (2009) |
| 7SL | 0.01 to 0.2 | 3–20 × $10^4$ | 0.3 | 0.005 to 0.03 | About 1–2 SRP molecules/100 ribosomes | Raue et al. (2007), Castle et al. (2010) |
| Xist | 0.0003 to 0.02 | 0.1–2 × $10^3$ | 2.8 | 0.0001 to 0.003 | | Buzin et al. (1994), Castle et al. (2010) |
| Other lncRNA | 0.03 to 0.2 | 3–50 × $10^3$ | 1 | 0.002 to 0.03 | Estimated at 1–4% of mRNA by weight | Mortazavi et al. (2008), Ramsköld et al. (2009), Menet et al. (2012) |

*The size for the average unspliced pre-mRNA is 17 kb; however, most pre-mRNAs are partially spliced at any given time, and the average size of hnRNA is estimated at 10 kb (Salditt-Georgieff et al., 1976).
**Based on the finding that 1–2% of all mRNA species generate circular RNA, which is present at 10% of the level of the parental mRNA.

Figure 2.5: List of the different types of ncRNAs from (Palazzo and E. S. Lee 2015).

| | | Mean size | Function |
|---|---|---|---|
| Long ncRNA | Ribosomal RNA | ~1.9 kb | Essential for protein synthesis |
| | XIST RNA | ~17 kb | Chromosome X inactivation |
| | Other lncRNA | > 200 nt | Involved in epigenetic modification, post-transcriptional processing, modulation of chromatin structure, etc. |
| Small ncRNA | miRNAs | 18-21 nt | Gene regulation |
| | siRNA | ~21 nt | Gene regulation; defense against viruses and transposon activity |
| | rasiRNA | 24-27 nt | Orientation of heterochromatin in the formation of centromeres |
| | snoRNA | 60-300 nt | Methylation and pseudo uridylation of other RNAs |
| | snRNA | 100-300 nt | Involved in spliceosome complex |
| | piRNA | 26-30 nt | Regulation of transposon activity and chromatin state |

Figure 2.6: List of the different functions of the main ncRNAs from (Dogini et al. 2014).

2014). In some other cases, ncRNAs participate in the processing and modification of RNA. For instance, in higher level organisms, small nuclear RNAs (snRNAs), which are small non-coding RNAs usually localized in the nucleus, are fundamental for the splicing of pre-mRNAs; small nucleolar RNAs (snoRNAs), are peculiar small non-coding RNAs, which act predominantly at the rybosomial level and direct the 2'-O-ribose methylation (C/D-box type) and pseudouridylation (H/ACA-box type) of rRNA and tRNA by pairing with sequences in specific sites near the targets to be modified.

Furthermore, ncRNAs can affect the stability and transport of proteins too. Protein translocation across membranes requires the signal recognition particle (SRP), whose core is composed by a small cytoplasmic RNA, and this is conserved in organisms belonging to all the biological *regna* (Storz 2002).

ncRNAs also regulate the stability and translation of mRNA. The first

example to be studied is that of C. elegans let-7 miRNA (Reinhart et al. 2000), that represses translation by base-pairing with the 3' end of target mRNAs. As a matter of fact, many of the miRNAs characterized until now are expected to follow the same mechanism.

Summarizing, it is possible to categorize the mechanisms of action of the ncRNAs into four main general groups:

- Some ncRNAs use base-pairing with a molecule of RNA or DNA as the mechanism to carry out their job. Examples of this category are the snoRNAs, that guide RNA modification, the bacterial RNAs that regulate translation through base-pairing with specific target mRNAs and almost certainly the largest part of the miRNAs.

- Other ncRNAs have the ability to imitate the structure of different nucleic acids.

- Some ncRNAs have catalysis abilities, and an example is that of the RNase P RNA. This characteristic is very appealing for a variety of industries, and therefore many synthetic RNAs have been picked to exert many biochemical functions, but there is actually a narrow number of natural ncRNAs that have a demonstrated catalytic role.

- The largest part of ncRNAs are linked with proteins that provide them with a higher number of features; nonetheless, some ncRNAs, including the aforementioned snRNAs and the SRP RNA, have fundamental roles in the structure of RNA-protein complexes.

These categories are not univocal and many ncRNAs fit into different categories: for example the telomerase RNA exploits both the base-pairing and the structural mechanisms, providing the base-pairing template for synthesis of telomeres and serving as part of the structure of the telomerase ribonucleoprotein complex (Storz 2002). Furthermore, there are many other ncRNAs whose mechanism of action is still unknown and it is possible that many other mechanisms exists, but have not yet been detected.

Among the ncRNAs categories, one specific subset of small ncRNA is represented by the miRNAs: this sub-category is indeed the focus of this dissertation and will be described in the next sections.

### 2.2.2 miRNAs

As described so far, there are many categories of ncRNAs. The main subdivision is that between long non-coding RNAs and small non-coding RNAs. Among the subsets present in the last one, the one that has attracted most of the attention in the last years, particularly in human studies, is that of miRNAs.

miRNAs belong to the category of the short non-coding RNAs, they are ~19–24 nucleotides in length and they lead to gene silencing by acting as a

guide for the Argonaute (AGO) proteins to specific sites in the 3' untranslated region (UTR) of the target messenger RNAs. AGO proteins compose a huge family: they exploit single-filament small nucleic acids as guides to sequences with precise levels of complementarity on the RNA or DNA molecules targeted to be silenced. This assemble of miRNA and AGO constitutes the targeting unit of the final protein complex, the miRNA-induced silencing complex (miRISC), whose final objective is to repress translation and possibly lead to the degradation of specific mRNAs.

miRNAs have been discovered in 1993 by Lee and partners (R. C. Lee et al. 1993) in the nematode C. elegans. The first miRNA ever identified was a small RNA generated by the lin-4 gene, which is able to repress the lin-14 mRNA at the post-transcriptional level. It is interesting to notice that the resulting transcript of lin-4 wasn't translated into a protein with biological activity as expected, but it led to the production of two small RNAs, one of 21 nucleotides and the other of 61 nucleotides in length. The longer sequence bent into the typical stem-loop structure and resulted the precursor for the shorter RNA. Later it was found out that the shorter sequence of 21 nucleotides had antisense complementarity to different sites in the 3' UTR of lin-14 mRNA. Its binding to these complementary regions on the mRNA lowered the level of expression of its protein, LIN-14, even if no significant changes were recorded at the mRNA levels. The model brought by these studies described how base pairing took place between multiple lin-4 small RNAs and their matching sites on the 3' UTR of lin-14 mRNA, eventually leading to repression of the translation of its mRNA (Bhaskaran and Mohan 2013), practically describing the basic mechanism of action of the miRNA category.

In the beginning, it was generally believed that these small RNAs belonged specifically to nematodes, until their presence and abundance was demonstrated also in many more animal phyla, thanks to the discovery of various miRNAs in many diverse species of plants and animals, many of which were demonstrated to be conserved across the biological kingdoms and cell-type specific. These small non-coding RNAs were subsequently recognized as a new class of regulator ncRNA and, finally, termed miRNAs.

miRNAs involvement in almost all of the processes that happen inside the cells during their lifespan makes them fundamental for development, cell fate commitment and homeostasis; the removal of the enzymes responsible for their biogenesis, Dicer and Drosha, are lethal in mouse embryos (Gebert and MacRae 2019). Furthermore, dysregulation of miRNA behaviour is linked with various pathologies, especially cancer, in which they can act either as oncogenes (and in this case they are referred to as oncomiRs) or tumor suppressors. Also, global changes in miRNA expression are associated with differentiation, whose dysregulation is a hallmark of all human cancers (J. Lu et al. 2005). Some examples of miRNAs associated to dysregulation of the other cancer hallmarks are reported in Figure 2.7.

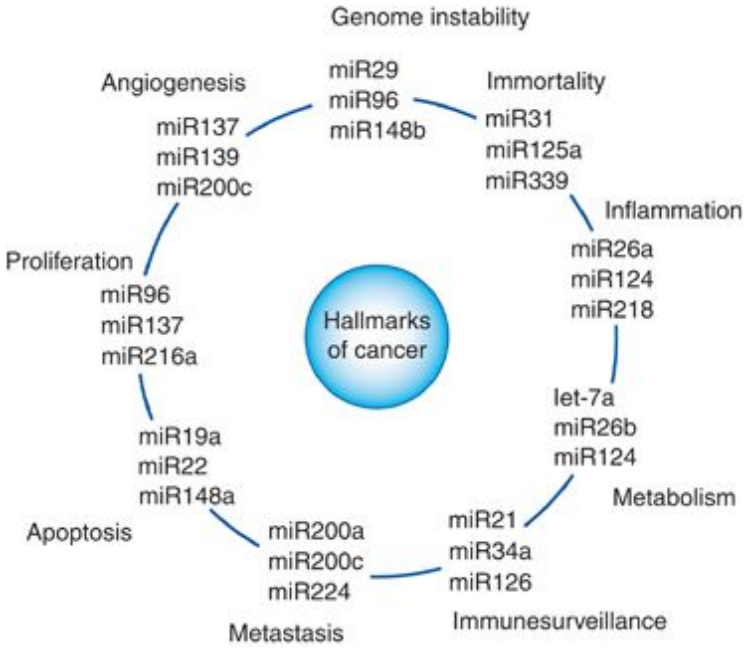



Figure 2.7: Schematic rappresentation of the interactions between miRNAs and the hallmarks of cancer. Every hallmark displays three of the possible miRNAs that can have effect on a specific cellular function related to a specific hallmark. *Image from (Pichler and G. A. Calin 2015)*

### 2.2.3 miRNA biogenesis and function

miRNAs maturation is a complex topic, starting from miRNA genes location. Their genes can be located both in overlapping introns belonging to protein coding transcripts and in the exons. miRNA genes can also be found in clusters of miRNA genes co-expressed poly-cistronically and potentially transcribed as a single unit. Almost half of all currently identified miRNAs are produced from transcripts with no protein-coding ability, while the remaining part can be usually found in the introns of coding genes, normally co-transcribed with their host genes, but undergoing independent processing and being regulated by their own promoters. At the present day, it is possible to state that genomic regions that are able to generate mature functional miRNAs can be located in different sections of the DNA. Their biogenesis takes place predominantly through one canonical pathway, even though to date, various non-canonical biogenesis pathways have been studied. Examples of some of them are described in Figure 2.8.

#### The canonical pathway

In the canonical pathway, pri-miRNAs are transcribed from their genes by RNA polymerase II in very log filaments, with a guanosine cap on the 5' and polyadenylation on the tail. This first form of miRNA is then processed into a precursor RNA of ~70-120 nucleotides in length by the microprocessor complex. This complex is composed by two main molecules: a protein that binds to the RNA named "DiGeorge Syndrome Critical Region 8" (DGCR8,

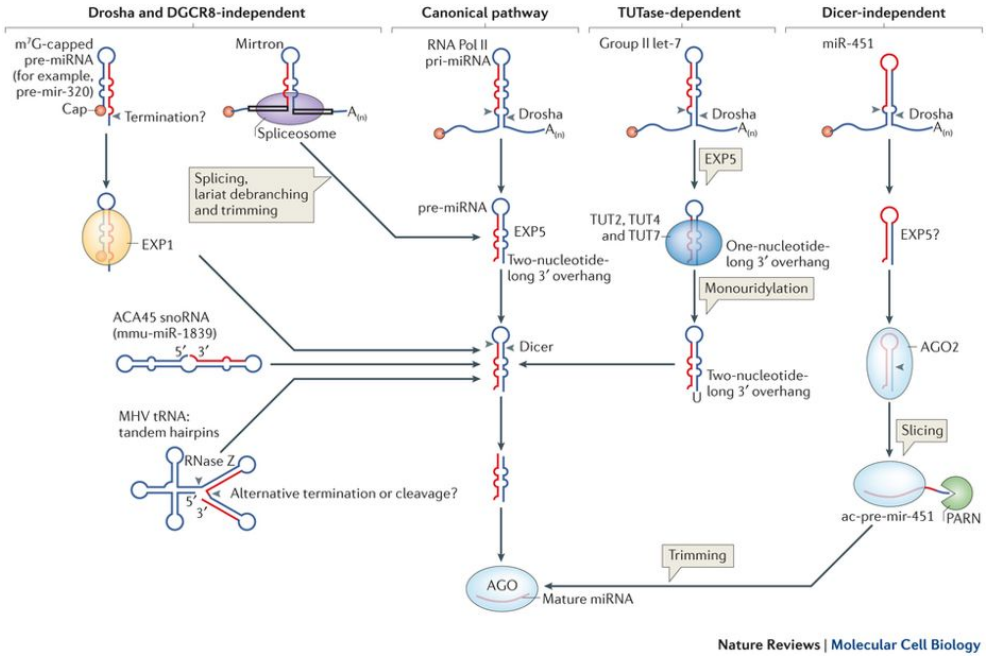



Figure 2.8: miRNA biogenesis pathways. In some cases, 7-methylguanosine (m7G)-capped pre-miRNAs, like pre-miR-320 in the example, bypass the Drosha processing. Mirtrons are able to produce pre-miRNAs directly via splicing and de-branching. Sometimes, pre-miRNAs generate from the cleavage of small nucleolar RNAs (snoRNAs). Some pri-miRNAs belong to the group of pre-miRNAs dependent from terminal uridylyl transferase (TUTase): these produce pre-miRNAs with a shorter 3 overhang that is below the standards required by Dicer processing. In the example, pre-miR-451 is produced by Drosha in a Dicer-independent way, is moved to the cytoplasm and loaded on Argonaute 2 (AGO2), where AGO2 performs the cleavage of its stem. The molecule is then further trimmed by the 3–5 exonuclease poly(A)-specific ribonuclease PARN. The question marks point out passages whose mechanisms are still not fully understood. MHV = murine -herpesvirus; mmu = Mus musculus; Pol II = polymerase II (Ha and Kim 2014).

or Pasha) which dimerizes with a ribonuclease III enzyme, named Drosha, which is highly conserved in animals (Denli et al. 2004). DGCR8 recognizes an N6-methyladenylate specific motif inside the pri-miRNA, while the role of Drosha is to cleave the pri-miRNA duplex typical hairpin structure (Han et al. 2004). The resulting pre-miRNA has a characteristic 5' phosphate and ~2-nucleotide 3' overhang. Next, the pre-miRNA is moved from the nucleus to the cytoplasm by a complex which is almost completely devoted to pre-miRNA export, the exportin 5 (XPO5)/RanGTP complex. In the cytoplasm, the pre-miRNA is modified by the RNase III endonuclease named Dicer (Denli et al. 2004). Through this process the terminal loop is removed and this results in a mature form of ~19- to 24-nucleotide-long duplexes.

The two miRNA strands are then divided: theoretically, each of the strands resulting from this separation can be loaded into the Argonaute (AGO) protein, spending the right amount of ATP. Nevertheless, it is the miRNA strand that possesses the most unstable base pairing at the 5' end or 5' uracil that usually is chosen as the guide strand and it is therefore loaded into AGO, while the strand that creates a stable base-pair at the 5' end, that is consequently named passenger or miRNA* strand, commonly undergoes degradation. Only occasionally it interacts with AGO proteins as well, and in this case the duplex generates two strands that can function as miRNA (O'Brien et al. 2018). In miRNA nomenclature, the miRNAs ending with a -5p denote those generated from the 5' end of the pre-miRNA hairpin; those ending with a -3p describe those with strand originated from the 3' end (Meijer et al. 2014). Each miRNA duplex generates very different percentages of AGO-loaded -5p or -3p strand according to the cell type and environment. The guide strand interacting with the Argonaute (AGO) proteins form the micro-ribonuclear protein complex (miRNP) named RNA-induced silencing complex, or RISC.

The complex is directed by the guide strand to its target mRNA, which is able to recognize thank to the sequence complementarity, leading to repression of its translation. AGO2 proteins are often located to specific structures inside the cytoplasm named GW/P-bodies (processing bodies), where miRNAs-mRNA complexes are supposed to be stored waiting to be degraded or to undergo translational repression (Patel et al. 2016).

### Non-canonical pathways

There is evidence that demonstrates that miRNA biogenesis can be achieved in different ways other than the canonical one. These non-canonical pathways are characterized by the usage of diverse combinations of those proteins that are part of the canonical one, above all Drosha, Dicer, XPO5, and AGO2 (O'Brien et al. 2018). Among the non-canonical miRNA biogenesis pathways it is possible to identify two main groups: the Drosha/DGCR8-independent and Dicer-independent pathways. For instance, the 7-methylguanosine (m7G)-capped pre-miRNAs are straightly exported to the cytoplasm without being cleaved by Drosha (Xie et al. 2013). Those miRNAs that are processed independently from Dicer, instead, undergo Drosha processing from endoge-

nous short hairpin RNA (shRNA) transcripts. These pre-miRNAs have a length that is suboptimal for Dicer processing, and therefore they conclude their maturation through AGO2 directly in the cytoplasm (Yang et al. 2010). Examples include hairpin structures similar to pre-miRNAs that are named Mirtrons, that generate from the introns belonging to spliced mRNA, a few small nucleolar RNAs (snoRNAs), and the short hairpin RNAs (shRNAs) produced endogenously (O'Brien et al. 2018). These and other examples of non-canonical pathways can be found depicted in Figure 2.8.

### 2.2.4 miRNA action

At the end of the biogenesis process, the mature miRNAs are ready to perform their function. At this point the product resulting from Dicer processing is loaded on the AGO protein, forming the RISC effector complex described earlier. These RNA duplexes sometimes show a specific affinity to some specific AGO protein. This happens for instance in flies, where miRNA duplexes preferentially load to AGO1, while siRNAs are sorted into AGO2, and the choice depends on their intrinsic structural properties (Ghildiyal et al. 2010). This is in opposition to humans, where a sorting system of this kind doesn't exist, and this leaves to the human AGO proteins (AGO1–4) the possibility to associate with almost equivalent sets of miRNAs.

The loading of the duplex is the first step of the RISC assembly and the subsequent unwinding is the second. After the loading of the miRNA duplex, the pre-RISC (pre- because the AGO protein is still associated with the double stranded RNA) rapidly dismisses the passenger strand, generating the mature RISC (Ha and Kim 2014).

Once the mature miRNA-RISC complex, the mi-RISC, is formed, the miRNA sequence dictates which mRNAs will be the target of the complex among those present in the cell, while the functional units of the miRISC will repress the translation of the resulting targets. Although AGO proteins are necessary for the silencing (Schmitter et al. 2006), there is evidence suggesting that association with GW182 is necessary for AGO proteins to achieve the silencing of their target mRNAs (Behm-Ansmant et al. 2006).

It is demonstrated that GW182 proteins facilitate miRNA repression serving as molecular scaffolding structure that connects to a variety of silencing effectors. One of these interactors is the PABP protein, that is binded by GW182 through its PAM2 domain. PABP binds at the same time the mRNA polyadenylated tail and multiple proteins involved in both mRNA translation and mRNA metabolism. Several mechanisms have been proposed about how the GW182-PABP complex is able to cause the mRNA silencing: one is that the association of GW182-PABP may improve miRNA-mediated deadenylation by physically approaching the deadenylation machineries recruited by miRISC next to the poly(A) tail; another describes GW182-PABP inhibiting translation of mRNA by preventing PABP-eIF4G interaction and the following mRNA circularization (Fabian and Sonenberg 2012). Even though these models are all demonstrated, they are still not completely understood and

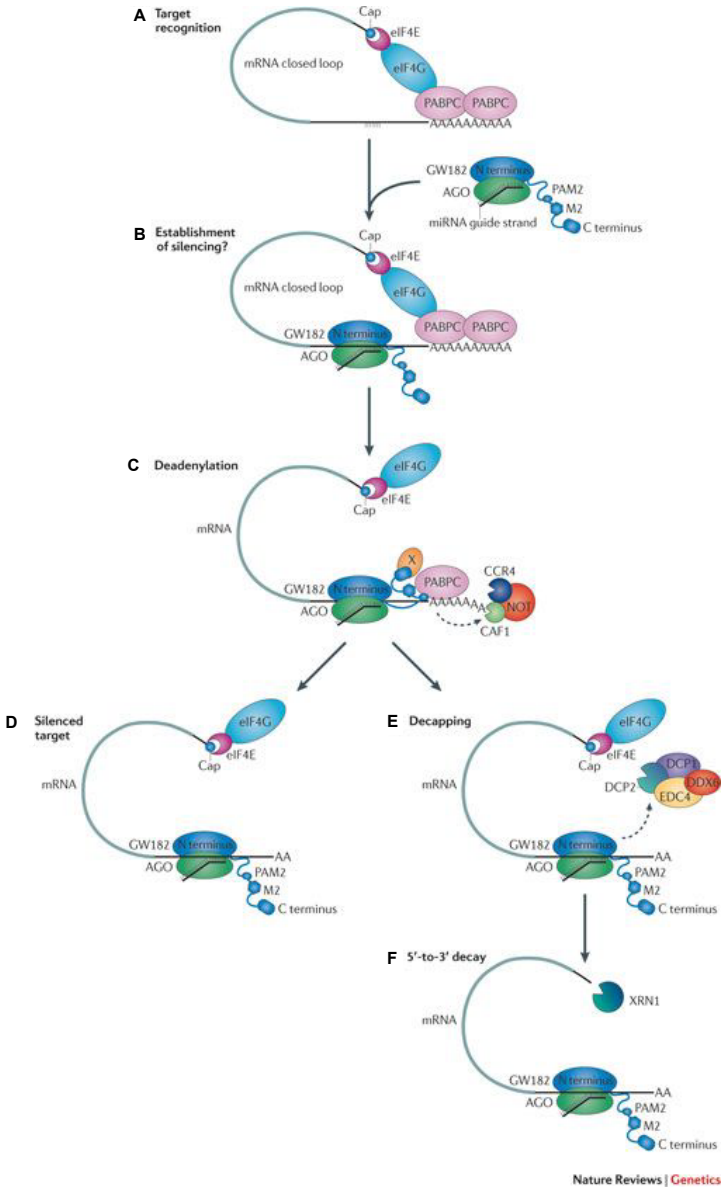



Figure 2.9: Summary of miRNA silencing mechanisms. **A** The mRNA target is pictured in a closed loop conformation. B,C. Animal miRNAs in complex with AGO identify the target mRNAs through base-pairing specific binding sites, mainly on the mRNA 3' UTR. AGO interacting with GW182 (**B**) interacting with PABPC, that is linked to the mRNA poly(A) tail (**C**). The AGO–GW182 complex leads to mRNA deadenylation (**C**). **D** Deadenylated mRNAs can be stored in a state of translational repression according to the cell type and target. **E,F** In animals, deadenylated mRNAs undergo decapping and are subsequently quickly degraded by the major 5'-to-3' exonuclease XRN1 (Huntzinger and Izaurralde 2011).

remain controversial. Another interactor of GW182 is EDD. This is probably a GW182-assisting unit, whose role is to engage miRNA effectors similar to RCK/p54 to target mRNAs, enhancing mRNA decapping and repressing cap-dependent translation (Su et al. 2011).

In general, after the binding of the miRISC complex on the 3' UTR of the target mRNA, there are different mechanisms through which it could lead to mRNA translation repression and silencing and an overview of the main pathways is reported in Figure 2.9. One of the mechanisms of miRNA-mediated silencing is mRNA deadenylation, that causes its instability and subsequent degradation. miRNA-mediated deadenylation requires the carbon catabolite repressor protein 4 (CCR4–NOT) deadenylation complex and, in part, the poly(A)-nuclease deadenylation complex subunit 2 (PAN2–PAN3). Several studies have demonstrated that GW182 proteins act also in this process as scaffolds for both the CCR4–NOT and PAN2–PAN3, therefore facilitating the procedure approaching the machinery to its substrate (Fabian and Sonenberg 2012).

It is now known that both PABP and EDD can be associated to the PAM2 motif of GW182, implying that these interactions may be mutually exclusive. One interesting hypothesis could be that the GW182 in association with EDD leads to repression of translation without having to resort to deadenylation, with the consequent mRNA destabilization, while the complex composed of GW182 and CCR4-NOT may instead lead to deadenylation and following mRNA decay (Fabian and Sonenberg 2012). Even though it is clear that miRNAs evoque translation repression and mRNA decay probably through the aforementioned processes, there are still open questions about the specific mechanisms causing these events.

Another way for the miRISC complex to achieve gene silencing is through inhibiting the beginning of translation, achieved by opposing to the function of the initiation factors. Many studies describe how the miRISC leads to their detachment from target mRNAs, inhibiting ribosome scanning and assembly of the eIF4F translation initiation complex (Gebert and MacRae 2019), but the overall miRISC-mediated translation inhibition mechanism is still not completely understood.

Furthermore, even if it is by now recognized that miRNAs carry out their silencing effect by binding to a precise sequence at the 3' UTR of their target mRNAs, causing inhibition of translation, there is remarking evidence that in some specific cases miRNA binding sites can be located in different sites on the mRNA, like the 5' UTR and coding sequence, and also within promoter sections, thus locating miRNAs in the nucleus, where interactions with miRNAs have been reported to induce transcription (Yijun Zhang et al. 2014). There are data implicating miRNAs in the regulation of the stability of mRNA in nucleoli and in alternative splicing. miRNAs may therefore be involved in gene expression regulation on a higher level, meaning the transcriptional level, where they may mediate both activation and inhibition of transcription of a target gene (Catalanotto et al. 2016).

miRNA understanding is increasing every day, with many studies trying

to understand the complexity of their dynamics in many different contexts. It is just by considering these characteristics and by considering miRNAs and the molecules they interact with as part of a sole network able to give rise to emergent unexpected phenotypes that we could be able to deal with their complexity.

### 2.2.5 Role of miRNAs in cancer

One of the first evidence of miRNA involvement in human cancer was found in B-cell chronic lymphocytic leukemia, where the genes transcribing for miR-15a and miR-16-1 resulted removed or downregulated in most of the clinical chronic lymphocytic leukemia cases (George Adrian Calin et al. 2002). Further analysis revealed that miR-15a and miR-16-1 have a tumor-suppressive effect, achieved by inducing cell death by repressing Bcl-2, which is a protein which provides resistance to cell death and whose overexpression leads to survival in malignant non-dividing B cells and many solid malignancies (Cimmino et al. 2005; George A. Calin et al. 2008).

More studies have highlighted how miRNAs can affect cancer, either in a repressive way, as oncosoppressors, or in an oncogenic way, as oncomiRs. Some miRNAs, for example, have a role into various pathways thar are fundamental for cell proliferation, and therefore the deregulation of these miRNAs have an important part in avoiding growth suppressors and keeping up proliferation signals in tumor cells (Peng and Croce 2016). One way through which they could interact with cell-cycle it is through the regulation of E2F proteins, a family of transcription factors that have an important cell-cycle-dependent regulating role in cell proliferation. It has been shown that there is a feedback loop between miR-17–92 cluster and E2F which keeps under control the cell cycle, making it take place in a normal way. However, several tumors show overexpression of miR-17–92, leading to the disruption of the feedback loop to promote cell proliferation (He et al. 2005).

Moreover, cell-cycle needs to be precise and functional, and it is therefore regulated by diverse cyclins, cyclin-dependent kinases (Cdks) and their inhibiting proteins and they are all generally regulated by miRNAs. One known oncogenic pathway, for instance, is related to miR-221/222 regulation of p27Kip1, a Cdk inhibitor (Peng and Croce 2016).

Another significant hallmark of tumor progression is preventing apoptosis, and it is believed to be regulated by miRNAs. For example, many miRNAs regulated by p53 are demonstrated to be implicated in p53 processes. One case is that of multiple myeloma, where there are three specific miRNAs (miR-192, miR-194 and miR-215) whose transcription is triggered by p53: after activation, the effect of the three miRNAs is the suppression of Mdm2 expression through the direct binding of its mRNA, eventually leading to p53's protection from degradation (Pichiorri et al. 2010).

Epithelial to Mesenchymal Transition (EMT) is known to be one of the earliest and fundamental stages in the metastatic process: the inhibition of E-cadherin induces loss of cell adhesion and at the same time genes linked

to motility and invasiveness are activated. In cancer, EMT is promoted in different ways: one of these is through miR-155, that induces EMT by downregulating RhoA GTPase, a key regulator of cellular polarity, tight junction formation and stability (Kong et al. 2008).

There are also miRNAs important for the control of metastatization. One example is miR-203, whose promoter is hypermethylated in highly metastatic breast cancer cell, which results in its significant downregulation. There, its restoration suppresses invasiveness *in vitro* and the presence of lung metastasis *in vivo* through the downregulation of SNAI2, and this suggests SNAI2 and miR-203 to be involved in an important regulatory loop in EMT and metastatization (Z. Zhang et al. 2011).

Angiogenesis is the process that leads to the development of new blood vessels in order to provide the tumor food and oxygen for growth and invasion. The two most studied angiogenesis-related transcrption factors are the hypoxia inducible factor (HIF), which is an important hypoxia-response transcription factor, and the vascular endothelial growth factor (VEGF), a fundamental angiogenic factor that directs the generation of new vessels from the endothelium. The most consistently and significantly hypoxia-induced miRNA is miR-210, whose overexpression in human umbilical vein endothelial cells under normal levels of oxygen induces the creation of structures similar to capillaries and VEGF-dependent cell migration (Fasanaro et al. 2008).

Abnormal miRNA expression in cancer provides the ability to keep up proliferation signals, avoid growth suppression, resist apoptosis, induce invasiveness and metastatization and promote angiogenesis.

miRNAs may sometimes act as tumor suppressors and others as oncogenes, depending on the context (Peng and Croce 2016). miRNAs can regulate many different targets, but their role in tumor settlement is possibly related to regulation of some peculiar targets, like for example the epigenetic components.

# Chapter 3

# miRNAs dynamics and complexity

We have already mentioned how miRNAs are produced in a cell-type-specific manner. We have spoken about their heterogeneous location in the genome. We have listed the different mechanisms through which they are transcribed and their many different ways of being involved in regulation. Still, this is describing just the tip of the iceberg of the greater complexity surrounding miRNAs and their action. This complexity is needed because of the delicate role that this specific subset of ncRNA has inside of the cell. As a matter of fact, the presence of miRNAs in the group of gene-regulating molecules is necessary to buffer gene expression to stability, and therefore their regulation is highly dynamic. A comprehensive understanding of this dynamicity is necessary to clarify the certainly high robustness of miRNA-mediated gene regulation. Robustness is achieved by a multi-level management of their behaviour, through various mechanisms that include solutions like a functional compartmentalization, shuttling of the miRISC complex across the cells compartments and the presence and abundance of both miRNAs and their target mRNAs.

These dynamics often ensure that the effect of miRNA-induced translational repression does not always lead to the same result in different cellular types. One of the mechanisms by which this can happen is related to the fact that the minimal miRNA-induced silencing complex (miRISC) indeed owns its target specificity to the interaction with the target sequences on mRNA, the miRNA response elements (MREs) (Vidigal and Ventura 2015). There are cases, for instance, in which alternative splicing and alternative polyadenylation have effect on the 3' UTRs, or in which some RNA binding proteins that are specific to a certain cell-type somehow modify the mRNA secondary structures, and these occurences may lead to modifications in the availability of the aforementioned MREs, eventually leading to subgroups of mRNAs that acquire sensitivity to a specific miRNA or become instead insensitive to their regulation, and this usually happens according to the specific cell-type or context that is considered.

According to the complementarity levels of the miRNA seed sequence and

its responsive elements on the mRNA, it is possible to determine if the target mRNA will be sliced by AGO2 or if its translation will be inhibited by the miRISC complex with the consequent mRNA decay. A perfect complementarity between miRNA and MRE triggers AGO2 endonuclease activity, causing mRNA cleavage. Nonetheless, this strong complementarity undermines the interaction of AGO with the miRNA inducing its degradation (Vidigal and Ventura 2015) and this mechanisms can be found mainly in plants (Park and Shin 2014). An imperfect complementarity would instead lead to mRNA decay or translational repression.

As already mentioned, miRNAs target recognition is primarily determined by the seed-sequence, which will recognize the MRE on the target messenger RNA. The vast majority of 3' UTRs have one single conserved MRE responding to a particular seed, but at the same time they usually have more than four conserved miRNA binding sites in total, thus allowing for combinatorial and overlapping regulation (Bartel 2009). Moreover, the shortness of the sequence required for targeting causes individual miRNAs to have the potential to regulate myriad of targets, even if the result on the single genes is commonly mild. Furthermore, more than one miRNA can control one single gene and at the same time whole cellular pathways can be controlled by one miRNA or clusters (Gebert and MacRae 2019), causing an increasing complexity in miRNAs regulation understanding.

There is also evidence of the difficulties of translating miRNA *in vitro* experiments results to *in vivo* contexts. Even in those cases in which the miRNAs and their targets tend to be negatively correlated *in vitro* (Baek et al. 2008), the data extracted *in vivo* are not always this clear, rendering difficult to have a perfect pairing between miRNAs experiments and their effect, like knockout and ectopic expression studies and their resulting phenotypes.

It is also important to keep in mind that the ability of a miRNA to repress the targets is strictly linked to its expression levels. This makes way more difficult to frame the biological roles of miRNAs and find clear interactions. miRNAs role of maintaining a steady state across the cells make their inactivation resulting in a very moderate upregulation of the mRNAs they control, typically lower than two-fold even for highly abundant miRNA, even if this may still have severe phenotypical consequences, especially if the targets are functionally linked (Vidigal and Ventura 2015).

In part, this non-intuitive behaviour is due to the fact that miRNAs do not act alone. One fundamental characteristic of these small non-coding RNAs is their positioning within gene regulatory networks, in particular within feedback and feedforward loops.

To sum up, two of miRNAs characteristic features are redundancy and involvement in interactions with other regulatory elements, often establishing with them regulatory loops. These, though, are also two common strategies used to achieve the stabilization of biological outcomes against genetic, environmental, and stochastic perturbations. This has led to the idea that miRNAs play a crucial role in precisely tweaking the cell environment with the aim of avoiding the small changes arising from such perturbations to have

a detrimental impact on homeostasis and of maintaining the steady state. In summary, the complexity surrounding miRNAs regulation is achieved in different ways and can end up either switching off or precisely tweaking expression levels, consequently buffering to oppose to random fluctuations ('noise') in gene expression (Gebert and MacRae 2019).

### 3.0.1 Introduction to epigenetic interactions

A huge part of the fine tuning that characterizes miRNA action is due to their being positioned in the middle of regulatory networks. In the last years it has been seen that in many cases these regulatory networks belong to the epigenetic pathways. In general, it is possibile to highlight many different kind of interactions between the ncRNA field and the epigenetic field. Interactions between epigenetic components and RNA can be enclosed in two main contexts: the one of the epigenetic components directly targeting the RNA and epigenetically modifying it and the one of the epigenetic components and the RNA being two different actors in transcriptional regulation, interacting at different levels.

The first case is mainly related to long non-coding RNAs (lncRNAs) and messenger RNAs (mRNAs). The internal modifications on these subsets of RNAs have been neglected for a long time, even though they are long known. It was in the 1970s that these modifications were firstly discovered, identifying the most abundant internal mRNA/lncRNA modification, the N6-methyl adenosine (m6A), which is present on average in over 3 sites per mRNA molecule (N. Liu and Pan 2015). Furthermore, many other types of modifications have been studied that can occur directly on the RNA molecules, such as m5C or 2'O-methylated nucleotides, and many m5C modification sites have until now been identified (N. Liu and Pan 2015).

The interest around these modifications was brought back to life by the discovery of the first RNA de-modification enzyme, FTO, which catalyzes oxidative reversal of methylated DNA and RNA and which is highly associated with diabetes and obesity in the human population. This enzyme uses m6A in mRNA/lncRNA modification as substrate and this discovery indicates that m6A modification is subject to a complex cellular control, also highlighting the idea that RNA modifications may act as epigenetic markers and control akin to DNA methylation and histone modification (N. Liu and Pan 2015). Although this field is very promising, it is still quite young and not well explored.

More is known about the interaction across epigenetic components and RNAs. In these relationships a huge role is covered by those RNAs that operate in a mechanism called 'Post Transcriptional Gene Silencing' (PTGS). This silencing mechanism is performed by three classes of small RNA that target transcripts in the cytoplasm, and the main pathways are described in Figure 3.1.

- miRNAs, are RNAs generated from hairpin structures, whose incomplete pairing to mRNAs leads to translational repression;

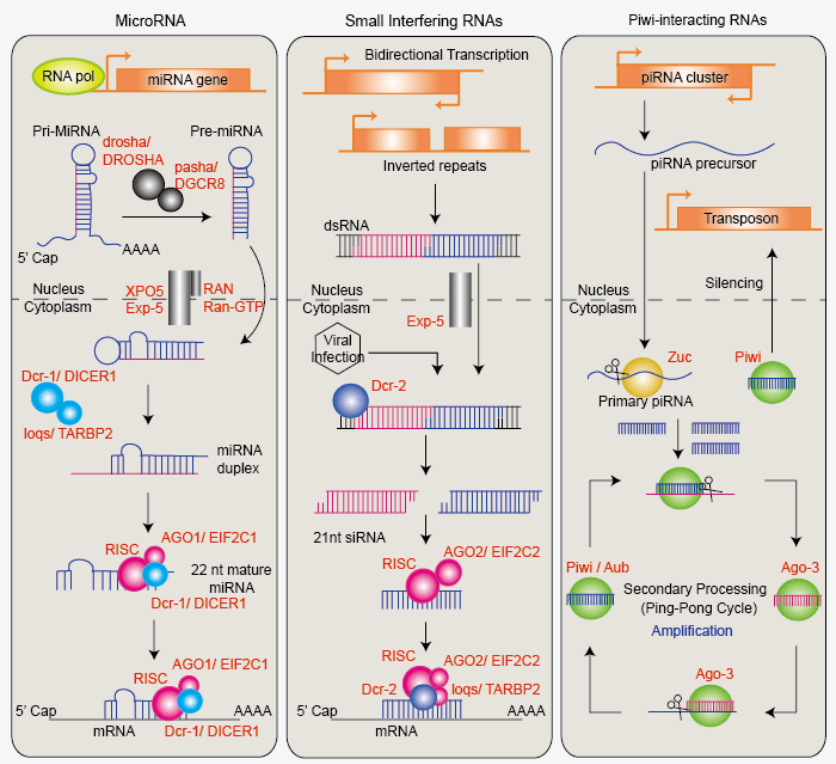



Figure 3.1: Overview of the small RNAs acting in the post-transcriptional gene silencing process. *Image from (*RNA interference *n.d.)*

- small interfering RNAs (siRNAs), pair perfectly to their targets, that are subsequently degraded;

- while PIWI-interacting RNAs (piRNAs) targets' are the transposon transcripts in animal germ lines (Castel and Martienssen 2013).

The interactions between these small ncRNA categories and epigenetics have been widely studied in S. pombe. In this model, there is evidence of these small RNAs interacting with members of the mechanism related to histone lysine methylation. From Arabidopsis thaliana *in vivo* analysis resulted that siRNAs can lead DNA methylation and its maintenance at specific unsymmetrical sites. This event shows an high number of physical associations among the molecules that are part of this pathway, establishing the ground of an epigenetic loop with self-reinforcement features. siRNA-mediated epigenetic regulation in Arabidopsis is led by AGO4, a specific member of the argonaute family, which is needed for siRNA accumulation and DNA and histone methylation (Holoch and Moazed 2015). In mice, piRNAs cause transposons silencing in the male germ line by causing *de novo* DNA methylation in two specific developmental stages: late development of the embryo and early neonatal (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008).

There are many other known interactions between epigenetics and ncRNAs, including also other subtypes of non coding RNAs. For example, there

is evidence that promoter-associated and nascent RNAs can regulate transcription via epigenetic mechanisms, apparently attracting other RNAs or acting as recruiters of regulatory and effector proteins (Mattick et al. 2009). It has been established that signal-induced ncRNAs can act as selective ligands for a specific RNA-binding protein, namely TLS (for translocated in liposarcoma). This association in turn ends up inhibiting the histone acetyltransferase activity of the CREB-binding protein (CBP) and p300 and in transcriptional repression (X. Wang et al. 2008).

These are just a few examples of RNA-epigenetics interactions, hints of how these two elements, together with their resulting modifications, cooperate on different levels to give rise to those specific chromatin states that are crucial for governing genomic functions.

### 3.0.2 miRNAs - epigenetic pathways interactions

The interactions among miRNAs and epigenetic components can act at different levels of the regulation and it is bilateral, meaning that miRNAs can affect epigenetic components as much as epigenetic components can affect miRNAs in return. About 50% of the miRNA genes surround CpG islands. For example in tumors, there are some miRNAs that are controlled by DNA methylation in a cancer-specific manner: this happens for instance for miR-31 in breast cancer and for miR-124a in colon cancer (Bianchi, Renzini, et al. 2017).

DNA methylation and histone modifications often associate to modulate the levels of miRNAs, and this is demonstrated by experiments combining the inhibition of HDAC with agents that induce DNA demethylation or inhibition of DNA methylation. It is interesting to see that there is a feedback system of regulation, because in turn some miRNAs resulted capable of regulating the expression levels of epigenetic factors. The presence of these feedback loops once more increases the amplitude of miRNA regulation, thus establishing a network of miRNAs-epigenetic machinery interactions that increases the strength and robustness of epigenetic regulation (Bianchi, Renzini, et al. 2017).

miRNAs are generally transcribed by Pol II, and consequently their expression can be controlled as any other RNA both spatially and temporally, and as for any other gene, their regulation can be either activating or repressive. For example, is demonstrated that c-myc is able to trigger the transcription of the miR-17-92 cluster (Dews et al. 2006; O'Donnell et al. 2005).

Furthermore, methylation on the DNA and histone modifications are able to influence the expressions of miRNAs. One example is that of miR-127, that resulted strongly up-regulated in tumor-derived cell lines as a result of the treatment with inhibitors of DNA methylases and histone deacetylase (Egger et al. 2004). This combined inhibition leads to a reduction of the levels of methylation on the DNA and a relaxed chromatin structure, thus restoring the expression of those genes that were silenced by epigenetic modifications.

These results suggest that epigenetic regulation have a role in controlling the expression levels of miRNAs.

On the other hand, miRNAs can affect epigenetics. For example, they can be implicated in the settlement of DNA methylation marks. In Arabidopsis, miR-165 and miR-166 are necessary to methylate the gene coding for the PHABULOSA (PHB) transcription factor. These two miRNAs associate with the new Phabulosa mRNA to modify the chromatin state of the PHB gene (Bao et al. 2004). This depicts an interesting alternative mechanism of miRNA-mediated gene expression regulation, different from the usual post transcriptional gene silencing. Furthermore, it is known that important enzymes in charge of DNA methylation (DNMT1, 3A, and 3B) result as miRNAs' predicted target in many cases (Dakhlallah et al. 2013; S. Wang et al. 2017).

miRNAs can also regulate chromatin structure, through the regulation of key histone modifiers. For example HDAC4 is a demonstrated target of miR-140 in mice (Tuddenham et al. 2006).

The processes mentioned thus far outline the interaction between miRNAs and epigenetic pathways, describing how miRNAs give their contribution and have an important role in the epigenetic regulation of gene expression. This interaction can take place in at least three different ways: with multiple epigenetic pathways regulating the expression of the miRNAs; with the miRNAs repressing the expression of epigenetic factors; with miRNAs and epigenetic regulators cooperating to control shared targets.

DNA methylation is indeed a mechanism of miRNA transcription regulation. As already mentioned, about 50% of miRNA genes contain CpG islands. DNA methylation on these CpGs is frequently present in a cancer-specific fashion. For example in a peculiar subtype of breast cancer, the triple-negative breast cancer of basal subtype, hypermethylation of the promoter is one of the principal mechanisms to induce silencing, causing the subsequent activation of its pro-metastatic target genes (Bianchi, Augoff, et al. 2012).

There are many histone modifications that resulted implicated in the regulation of miRNAs expression levels either in tumor or in the course of development. For instance a miRNA array analysis from Scott et al. (Scott et al. 2006) demonstrated a clear alteration of miRNA levels in response to a potent HDAC inhibitor in the SKBR3 breast cancer cell line, after which 27 miRNA species resulted differently regulated. As mentioned earlier in this text, DNA methylation and histone modifications often act in cooperative way to regulate miRNA expression (Z. Wang et al. 2013).

Investigating the role of miRNAs into the epigenetic world unveiled the strong intertwining of their complicated network of reciprocal interconnections. miRNAs main role of controlling gene expression at a post-transcriptional level goes hand in hand with their ability to interact with the epigenetic machinery, with which they are involved in many regulatory loops. The miRNAs involved in this kind of complex network are referred to as epi-miRNAs, as they are miRNAs that can directly or indirectly modulate the expression lev-

els of parts of the epigenetic layer of regulation and sometimes are regulated by them, creating feedback regulatory loops (Iorio et al. 2010). In the next paragraphs, a few examples of miRNAs-epigenetic interactions are reported.

In non-small-cell-lung cancer there is an inverse correlation between the expression levels of the members of the miR-29 family of miRNAs and those of DNMT3A and -3B and it is known that the same miRNAs down-regulate the expression of both DNMTs (Fabbri et al. 2007). miR-29 is also the first epi-miRNA to be ever studied. Its epigenetic function was first observed in lung cancer, and later in acute myeloid leukemia (or AML) (Garzon et al. 2009).

As miR-29 is known to directly downregulate DNMT3A and -3B expression, miR-290 is known to be involved in the negative regulation of three DNA methyltransferases, DNMT3A,-3B, and -1. This is obtained inducing the upregulation of their repressor, RBL2, that is a demonstrated target of miR-290 in mouse-derived embryonic stem cells (Benetti et al. 2008).

An experiment done in HeLa cells showed that miR-148 down-regulates DNMT3B expression levels and on the contrary that inducing miR-148 repression through shRNA causes an increasing of DNMT3B expression levels (Duursma et al. 2008). It has been demonstrated that DNMT1 is directly regulated also by miR-148a, along with miR-152 and miR-301, in human cholangiocarcinoma (Braconi et al. 2010).

miR-26a is a miRNA that takes part in many different signaling pathways, which has a tumor suppressor role in many tumor types, such as melanoma (Ryu et al. 2013), gallbladder cancer (Zhou et al. 2014), breast cancer (J. Gao et al. 2013), osteosarcoma (Song et al. 2014) and nasopharyngeal cancer (L. Yu et al. 2013). A research by Zhuang et al. (Zhuang et al. 2016) demonstrated that miR-26a is involved in a double negative feedback loop with EZH2, the catalytic component of PRC2, in hepatocellular carcinoma: miR-26a is epigenetically repressed by EZH2, and EZH2 is in turn repressed by miR-26a. In this case, the imbalance in the EZH2-miR-26a double negative feedback loop axis may contribute to miR-26a dysregulation and consequent tumor cell proliferation.

miR-1 is a well known epi-miRNA involved in myogenesis and related diseases. It resulted overexpressed in DNMT1-/- HCT 116 cells, indicating DNMT1 as the epigenetic component responsible of its silencing. In contrast, after treatment of HCT cells with 5-AzaC, a DNA hypomethylating agent, the downregulation of several miR-1 targets including FoxP1, MET, and HDAC4 was observed (Datta et al. 2008).

miR-140, which is down-regulated in different cancer types, has been proven to directly target HDAC4 too. There are several studies showing that histone-modifying enzymes are regulated by miRNAs, as well as miR-1 and miR-140 do with HDAC4.

HDAC1, is frequently overexpressed in many types of cancer and resulted to be a direct target of miR-449a (Noonan et al. 2009; Iorio et al. 2010) and miR-449a in turn negatively regulates the expression of HDAC1 in PC-3 prostate cancer cells, which makes it fall under the definition of epi-miRNA.

Polycomb group (PcG) proteins are a group of transcription repressors, and as such they also act on regulators of the development process in Embryonic Stem (ES) cells and in cell lineages already committed to differentiation, comprising Skeletal Muscle Cells (SMC). Polycomb proteins downregulate the expression of miR-214 in uncommitted SMC. Once the differentiation process starts, PcG disengage, activating miR-214 transcription. In turn, miR-214 negatively targets EZH2 (enhancer of zeste homolog 2) 3'UTR, an enzyme that represent the catalytic unit of the PRC2 complex, creating a negative feedback loop between these two elements. Therefore, miR-214 and EZH2 are involved in a regulatory feedback loop that controls the expression of those genes that depends on PcG regulation during differentiation in skeletal muscle and embryonic stem cells (Juan et al. 2009).

EZH2 overexpression in prostate cancer cell lines SKBR3 and DU145 is inhibited by miR-101. During prostate cancer progression miR-101 expression decreases and there is a parallel increase in EZH2 expression (Varambally et al. 2008).

## 3.1 epi-miRNA regulatory network

As it has been displayed until now, miRNA and epigenetic components expression and mutual interactions are often highly coordinated. This growing amount of evidence about the interactions between miRNA and epigenetic components suggests that epigenetic regulation can be completely understood only considering both elements as two constituents of the same regulatory network. The characteristics of this network have been described by Osella et al. (Osella, Riba, et al. 2014): in this epi-miRNA network it is possible to identify recurrent wiring patterns, and these recurring circuits in a network are usually called network motifs.

Particular attention should be made on a specific motif, that seems of peculiar relevance in the epi-miRNA network: the double negative feedback loop (DNFL), in which a miRNA (or, sometimes, a group of miRNAs that act cooperatively) targets an epi-R, which in turn regulates the expression of that same miRNA. This specific network structure is called "toggle switch", and this name refers to its ability to act as a genetic switch between different cell fates (Gardner et al. 2000), which makes it a fundamental gear in the complex engine of differentiation and developmental processes (Alon 2007).

This, though, is not the only function that it may exert. The same motif can be useful in many other contexts. The specific dynamics that this motif can arbour make it able to be used as a "memory unit", because it is able to fix a temporary stimulus into a stable expression pattern, persisting even when the triggering stimulus is over, making the whole system to "remember". With the aim of best performing this function, the DNFL must be regulated in a way such that it stays in the bistability region (e.g., it must allow two competing stationary states), it must have a fine tuned switching threshold, to avoid undesired accidental transitions between the two alternative states (e.g., it must be a "robust" switch), but it should also allow a switch-back

transition, if needed (Osella, Riba, et al. 2014).

In biology there are plenty of systems exploiting the toggle switch motif, and increasing evidence suggests that in several of these loops one of the two actors is a miRNA. One peculiar example is that involving ZEB1, a transcription factor and epithelial to mesenchymal (EMT) transition inducer, and miR-200c. In a work by Burk et al. (Burk et al. 2008) it has been proven via knockdown experimetns of ZEB1 in SW480, HCT116 colorectal and MDA-MB231 breast cancer cell clones that ZEB1 is able to downregulate several miRNAs. The miR-200 family underwent the strongest alterations: ZEB1 binded miR-141 and miR-200c to multiple highly conserved sites in their promoter, causing their direct suppression. Furthermore, is demonstrated that ZEB1 is a target of miR-200c, suggesting that invading cancer cell can establish a feed-forward loop to enhance EMT, and describing a clear case of toggle switch, which, in this case, is a fundamental element for the transition between the epithelial and the mesenchymal phenotypes. Another example of DNFL is the one involving miR-214 and EZH2 that has been discussed earlier in this text.

It is shown that in this kind of circuit miRNAs are better at providing robustness in cell fate commitment than transcription factors. This is mainly due to the mechanism through which miRNAs physically perform their regulatory function, that can naturally improve the stability of the circuit steady states reducing the random fluctuations due to gene expression noise. When the second partner of the toggle switch is an epi-R, the effect is even greater. Epigenetic regulation can be itself described as a switch between discrete states. Combining it with the post-transcriptional nature of miRNA regulation increments both the bistability and the general robustness of the switch when introduced in a complex network (Osella, Riba, et al. 2014).

The robustness conferred by this kind of switch is well described in a model by Osella et al. (Osella, Riba, et al. 2014), that will be the theme of the next section.

### 3.1.1 The toggle switch: the added value of miRNAs

During the years, several examples of toggle switches and bistable switches have been recorded, and these usually occur between two transcription factors (Tian and Burrage 2006). Nevertheless, in the last few years there have been many indications that suggested that in many of the biological examples of the switch, a miRNA takes the place of one of the two transcription factors. The typical example is the molecular switch at the basis of the EMT, the miR-200-ZEB toggle switch mentioned above, that therefore has a fundamental part in embryo development and in metastatization (Burk et al. 2008; M. Lu et al. 2013). In the paper by Osella et al. (Osella, Riba, et al. 2014) it is shown one of the potential benefits of having a miRNA as one of the two elements of this type of loop. The mechanism through which miRNAs regulate their targets is perfectly suited to boost the stability of the circuit steady states counteracting the random transitions implicit in

gene expression. This characteristic might be especially significant in the case in which there is an association between the steady state and a specific phenotype that needs to be preserved even without the specific triggering stimulus, as for the miR-200-ZEB example. In the simplest way to think about it, this happens because miRNAs act at the post-transcriptional level, decreasing the random fluctuations of target proteins (Levine et al. 2007; Osella, Bosia, et al. 2011), thus increasing the robustness of the switch.

It is interesting to underline that this behaviour is even more intense in the case in which one of the two elements of the toggle switch is an epi-R. As mentioned before, epigenetic regulation can be described as a switch between discrete transcription states of the target genes. Alterations in concentration of epi-R is therefore greatly non-linear and step-like, compared with the usually smoother dependence of target activity on the concentrations levels of its transcription factor. When this characteristic is coupled with the fact that miRNAs regulatory action takes place at the post-transcriptional level, it increments the range of bistability of the switch and its overall robustness when it is considered in a complex network with a wide range of possible inputs (Osella, Riba, et al. 2014).

These hints suggest that there may be an evolutionary-related reason behind the choice of miRNA regulation, with respect to transcriptional regulation, to create toggle switches that have an important role in cell fate commitment. Quantitatively speaking, this can be measured by a model in which transcription and translation will be considered explicitly for the epi-R, in order to entirely consider the stochastic effects that would arise from the burstiness of gene expression. Moreover, the model considers also the physical association of miRNAs with their target mRNAs, and the catalytic and stoichiometric nature of this coupling.

In this mathematical model designed in the work by Osella et al. (Osella, Riba, et al. 2014) *s*, *m* and *p* denote the number of miRNAs, mRNAs and proteins respectively. From this derives the description of the dynamics of the average amounts of the different molecules, in the form of the following three coupled equations:

$$\frac{ds}{dt} = \frac{k_s}{1+\left(\frac{p}{h}\right)^n} - \gamma_s s - \alpha k m s$$
$$\frac{dm}{dt} = k_m - \gamma_m m - k m s$$
$$\frac{dp}{dt} = k_p m - \gamma_p p.$$

Figure 3.2: Equation (1). Image from (Osella, Riba, et al. 2014)

The parameter *k* describes the rate of miRNA-mRNA coupling (and depends on the energy of the RNA-RNA connection), while $\alpha$ is the cataliticity parameter that describes the probability of miRNA degradation following a miRNA-mediated mRNA degradation event. The limit $\alpha \to 1$ corresponds

to a stoichiometric mode of action. The opposite case, the limit $\alpha \to 0$, portrays a perfect catalytic mode, in which the rate of mRNA degradation becomes simply a linear function of the number of miRNAs (Osella, Riba, et al. 2014).

The circuit randomly switches between the equilibrium of the two steady states. The switching rate between these two states is defined as the timing between these transitions, and the simulations performed with Equation (1) (Figure 3.2) that miRNA regulation can regulate gene expression fluctuations by lowering the target burst size (Figure 3.4). This can be derived from the prior equation as:

$$b = \frac{k_p}{\gamma_m + ks}$$

Figure 3.3: Equation (2). Image from (Osella, Riba, et al. 2014)

This describes the mean number of proteins that can be produced from a mRNA, which depends on two parameters that are fundamental in determining how much noise there is in the protein level, and consequently the probability of seeing a stochastic transition between the two different steady states: the average miRNA concentration *s* and the strength of repression *k*. In fact, the switching rate can vary of many orders of magnitude depending on the degree of miRNA regulation, and therefore on the effective target burst size (Figure 3.4 B). Finally, the degree of cataliticity of the miRNA-mRNA interaction apparently plays a relevant role in defining the stability of the motif. A low degree of cataliticity, portrayed for example from a high probability of a coupled miRNA-mRNA degradation after their physical interaction, allows a stronger reduction of the switching rate. It is important to underline that this model results a grainy portrait of the real dynamics of this kind of circuit, but it is still a good measure of the importance of having these specific players, the epi-R and the miRNA, involved in this kind of interaction.

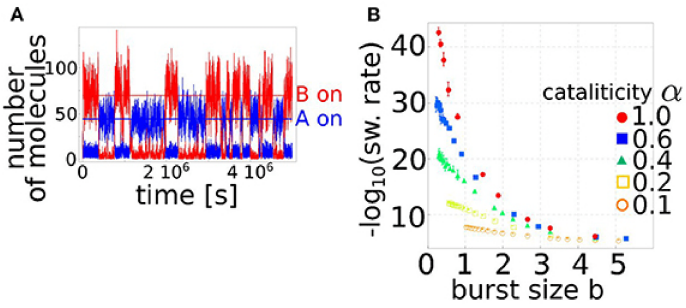



Figure 3.4: miRNA-mediated regulation increment the stability of the toggle switch. **A** Random gene expression noise can lead to transitions between the two stable steady states. In picture, simulation of the circuit switching between the case in which A is under active transcription while B is switched off and the opposite one. The timing between these transitions is set by the switching rate, which is a function of the circuit parameters. **B** The switching rate is displayed as a function of the effective burst size in (2), as set by the level of miRNA regulation. Each curve correspond to a different level of cataliticity $\alpha$ e. g. the ability of the miRNA to be recycled and not degraded with the targeted mRNA (see Equation (1)). For greater recycling ability ($\alpha \to 0$) there is lower circuit stability dependent on the burstiness of the process (Osella, Riba, et al. 2014).

Despite the robustness and remarkable importance of the toggle switch motif among the epi-miRNA network, it is obviously not the only circuit through which miRNAs can interact with the epigenetic mechanisms. Different kind of circuits can be highlighted, and will be reported later in the text.

### 3.1.2 Searching for new interactions

This quick overview about miRNAs and epigenetics should have clearly described the intrinsic complexity of the network in which these two elements cooperate. Even though miRNAs are historically known to act as post-transcriptional gene silencer, and that is indeed their main role, is impossible to solely categorize the regulatory function of miRNAs as "repressive", when more and more frequently evidence is found of their effect over a wider range of targets, not explainable by the simple repression of some common target mRNAs.

It is not possible to investigate the role of a miRNA in a particular organism, cell line or tissue, without taking into consideration its network and its molecular context of interactions. epi-miRNAs are a growing research field because thus far, it has been difficult to find a way to unwind this tangle of interconnections and identify those miRNAs that really act cooperatively with the epigenetic components: this is what this dissertation is all about. In the next chapter, a new way to prioritize candidate epi-miRNAs will be described, together with the obtained results.

# Chapter 4

# Project

## 4.1 Introduction to the project

In the last few years, the interplay between miRNAs and the epigenetic layers of regulation has been intensely investigated. As we stated in the earlier sections, growing amount of evidence supports the importance of a few selected miRNAs in regulating epigenetic factors (Gruber and Zavolan 2013). At the same time, several miRNAs (and among them most of known epi-miRNAs) have been shown to be under strict epigenetic control (Gruber and Zavolan 2013; Iorio et al. 2010; Kunej et al. 2011; Sato et al. 2011; Z. Wang et al. 2013). The accumulating evidence of their interaction clarifies how profound this should be and it has been recently suggested that a complete understanding of epigenetic regulation could be reached only analyzing epi-Rs and miRNAs as two associated elements of the regulatory circuitries, as previously discussed in the work by Osella et al. (Osella, Riba, et al. 2014).

Four are the main epigenetic pathways involved in this kind of networks and considered in our analysis. They have been addressed deeper in the previous sections and here a short summary is reported. Two out of four are pathways belonging to the category of Polycomb-group proteins. The Polycomb Repressive Complex 1 (PRC1) is a transcriptional repressor complex consisting of several proteins, among which an important role is played by the Polycomb group ring finger (PCGF) protein family (in particular BMI1), and is responsible for the modification of histone H2A. PRC1-dependent modification of histone H2A involves the monoubiquitylation of (predominantly) lysine 119 (K119Ub) (Vidal and Starowicz 2017). Belonging to the same family of repressors, the Polycomb Repressive Complex 2 (PRC2) is composed of four main subunits: RbAp48, SUZ12, EZH2 and EED. PRC2 can work alone by adding the repressive mark of trimethylated histone H3 at lysine 27 (H3K27me3) to its target genes or in cooperation with NIPP1 (Nuytten et al. 2007) to increase its repressive action. As previously mentioned, DNA methyltransferases (DNMTs) are the proteins in charge of DNA methylation. They have in general a transcriptional repressive role, especially on CpG-rich promoters because their role is to methylate cytosines, and they do this through the covalent transfer of a methyl group from S-adenosyl me-

thionine (SAM) to the carbon C-5 of cytosines to produce 5-methylcytosine, and methylation can influence the recruitment of transcriptional regulators. Conversely, transcription factors bound to DNA can also directly recruit the DNA methylation machinery (Laisné et al. 2018), as in those cases in which DNMTs are known to interact with HDACs (Geiman et al. 2004), which are the fourth regulators we considered for our analysis. HDACs are the effectors of histone deacetylation: their action lead to an increase in the ionic interactions between the histones, with positive charge, and DNA, that has a natural negative charge, compacting the chromatin structure, thus limiting the access to genes to the transcriptional complexes and consequently lowering the levels of gene transcription (Ropero and Esteller 2007).

As well as DNMTs and HDACs, all the mentioned epigenetic pathways can interact and, for instance, there is evidence of an interaction between HDAC and PRC1 (and possibly PRC2) (Bommi et al. 2010).

Notwithstanding their importance, the number of experimentally validated miRNAs involved in regulatory circuitries with the above-mentioned epi-Rs, those that have been referred to as epi-miRNAs in this dissertation, is still very small, their identification episodic, and no attempt was performed up to now for a systematic search. The main goal of this work is to fill this gap. We propose a simple protocol to identify candidate epi-miRNAs and try to unveil their mechanism of action by combining gene expression data coming from experiments of overexpression of a given miRNA and data of knock-down, knock-out, drug-mediated downregulation or ChIP experiments (from here on referred to as epi-R ablation) of a given epi-R, taking advantage of the large amount of gene expression data available online, often not fully exploited.

The whole method is based on a very common tool in computational biology, that is the intersection between gene sets. It is a widely used technique to spot common functions, pathways, and underlying biological interactions. One of the most used online tool to compare a list of genes to many other is MSigDB (Subramanian et al. 2005; Liberzon et al. 2015), that we used to download the epigenetic signatures described later in the text. The intersection method is also used to check the composition of a gene list in terms of functional annotation. For example DAVID (Huang et al. 2009; Huang et al. 2008) is an online tool able to investigate enrichment in various fields: biological terms (GO (Ashburner et al. 2000; Consortium 2019)), BioCarta KEGG pathway and many others. In the same way, PANTHER (Mi et al. 2019) uses gene sets overlap to check for GO functional enrichment of gene lists. Another tool exploiting this simple concept is CMap (Connectivity Map), a repository of expression profiles of multiple cell-types that underwent systematic perturbations, with various types of perturbagens, in order to obtain an inventory of connections between molecules that allows to functionally investigate the genome (*Connectivity Map (CMAP)* 2018).

This procedure can also be used to identify candidate epigenetic interactors of a given miRNA of interest. As an example we studied the case of miR-214, a promestastatic miRNA (Penna, Orso, Cimino, Tenaglia, et al. 2011;

Penna, Orso, Cimino, Vercellino, et al. 2013) involved in the coordination of melanoma tumor cell migration, invasion, adhesion to extracellular matrices, transendothelial migration, and survival to anoikis in vitro, as well as extravasation from blood vessels and metastatization in vivo (Penna, Orso, Cimino, Tenaglia, et al. 2011). miR-214 acts through the simultaneous coordination of a network including over 70 protein coding genes and the anti-metastatic small non-coding RNA, miR-148b (Penna, Orso, Cimino, Vercellino, et al. 2013). In the present work, it was experimentally demonstrated that the histone methyltransferase enhancer of zeste homolog 2 (EZH2), the main component of the polycomb-repressive complex 2 (PRC2), is an interactor of miR-214, involved, at least partially, in the activation of miR-214 downstream players: ITGA5, ALCAM and the small non coding RNA, miR-148b. Due to the importance of EZH2 as a crucial epi-R involved in the initiation and advancement of melanoma (Mahmoud et al. 2016), the described miR-214-EZH2 axis could be of potential interest to understand tumor progression.

Moreover, we investigated the interactions between miRNAs and epigenetic components in TCGA data from cancer tissue samples. As previously described, miRNAs are often linked to the dysregulations observed in cancer and dysregulated miRNAs have a demonstrated role in the known hallmarks of cancer, from keeping up signals of proliferation, to avoiding growth suppressors, they can also induce resistence to apoptosis, activate invasiveness and metastatization, and induce angiogenesis. miRNAs are often subjected to studies as biomarkers for human cancer diagnosis, prognosis and therapeutic targets (Peng and Croce 2016). Analysis of epi-miRNA behaviour in cancer tissue samples highlighted a strong epi-miRNA dysregulation in cancer and their role in important cancer-related pathways.

### 4.1.1 Rationale

The rationale behind our approach is that, since most of the considered epigenetic pathways are repressors of gene expression, if they are controlled (e.g. downregulated) by a miRNA then all the genes targeted by that specific epigenetic component should be upregulated when the miRNA is overexpressed.

To clarify this concept, lets consider this case: if an epi-R is directly repressed by a specific miRNA, the overexpression of the miRNA itself should upregulate the epi-R's direct targets, given that the epi-R here considered are mainly repressors. Thanks to this interaction between the miRNA and the epi-R it would therefore be possible to find a significant intersection between the genes upregulated in a miRNA overexpression experiment and the upregulated genes in a epi-R ablation experiment, that should therefore lead to the same output.

In general, the interaction between a miRNA and a global epi-R could lead to an epigenetic "amplification" of the miRNA action that may explain the episodic empirical observation of a large number of genes varying their expression in miRNA transfection experiments. Evidence that this is indeed

the case will be provided later in this text.

## 4.2 Materials and methods

### 4.2.1 The epigenetic components related datasets: MSigDB

Gene sets were downloaded from the MSigDB database (MSigDB, v5.2 updated October 2016) (Subramanian et al. 2005; Liberzon et al. 2015) which contains gene sets collected from various sources, from on-line pathway databases to PubMed publications. We selected a panel of gene sets broadly related to epigenetic pathways as discussed in the main text. We chose, when possible, multiple sets for a specific component, spanning over different tissues or experimental conditions. We ended up with a panel of 43 sets, summarized in Supplementary File S1. These sets can be divided into two main classes: results of ChIP experiments, whose entries allow to directly identify putative targets of the selected factor; results of gene expression experiments after knockdown, knockout or drug-mediated downregulation of a particular epi-R. In this case, we downloaded separately both the upregulated and the downregulated genes: these allowed us to fix also the sign and in some cases the direction of the regulatory interactions we found.

### 4.2.2 The miRNA transfection datasets: GEO

Fourteen series of miRNA transfection experiments on five different platforms by Seki et al. (Seki 2011a; Enokida 2012; Naoko Kikkawa et al. 2014; Seki 2011b; Nijiro Nohata, Toyoyuki Hanazawa, et al. 2011; Moriya et al. 2012; Fuse et al. 2012; Yoshino et al. 2011; Yamada, Sugawara, et al. 2018; N. Nohata, T. Hanazawa, N. Kikkawa, et al. 2011; Nijiro Nohata, Sone, et al. 2011; Hidaka et al. 2012; Yamasaki et al. 2012; Takashi Kinoshita et al. 2012; T. Kinoshita et al. 2013; Goto, Kojima, Nishikawa, et al. 2015; R. Matsushita et al. 2015; Goto, Kojima, Kurozumi, et al. 2016; Ryosuke Matsushita et al. 2016; Kumamoto et al. 2016; Yonemori et al. 2017; Yamada, Koshizuka, et al. 2018; Yamada, Arai, Sugawara, et al. 2018; Yamada, Arai, Kojima, et al. 2018a; Arai et al. 2018; Yamada, Arai, Kojima, et al. 2018b), for a total of 157 gene expression datasets involving 59 miRNAs transfected in different human cancer cell lines were downloaded from GEO. An experiment series from Misiewicz-Krzeminska et al. (Misiewicz-Krzeminska et al. 2013), containing data from miR-214 expression in H929 myeloma cell lines was also downloaded to support our experimental data. As reported by the authors, these gene expression data were obtained using Agilent whole genome microarrays. The data were LOWESS normalized and background subtracted from log10 of processed Red signal/processed Green signal by the author through Agilent software. Some of the gene names annotated to the probes were obsolete and we updated them, exploiting the bioconductoR (Bioconductor 3.2, R 3.2.3) package biomaRt (v. 2.34.2) (Durinck et al. 2009). Then the probes with the highest fold change across each platform

were selected, and a unique set of genes conform to all the experiments considered was computed in order to have the universe gene set to use for the hypergeometric test. This contained 17024 genes. The last step was to compute for each experiment the group of the up- and down- regulated genes. To do this, for each transfection experiment we applied an arbitrary threshold of t = $\pm$0.5 on the log10 (fold change).

### 4.2.3 The hypergeometric test

As already mentioned, the interaction between miRNAs and the epi-Rs could possibly lead to a group of genes in common between the genes differentially regulated in a miRNA overexpression experiment and in an epi-R ablation experiment. Finding this kind of intersection, whether significant, would be the evidence that there is indeed an interaction. The method that will be described to identify candidate epi-miRNA is based on this simple concept. For each experiment analyzed with our pipeline, we computed the intersection between the miRNA and the epi-R gene sets and the significance of this intersection was computed with the hypergeometric distribution test.

The hypergeometric distribution is built to compute the probability in a scenario similar to the one described here: in a jar containing 10 red marbles and 90 blue marbles, 10 of them are collected by the user. The hypergeometric distribution computes the probability of collecting *k* red marbles. In this case, collecting none or one red marble seems intuitively most likely.

If the result would instead be shifted towards a higher number of red marbles, this could be a sign of some non-random process acting underneath. These kind of non-random processes are those in which we were interested. If an intersection contained a significantly higher than expected number of genes, this would have been the sign of a possible interesting interaction between the miRNA and the epi-R.

More in detail, the hypergeometric distribution is characterized by 3 parameters:

- The population size, usually denoted *N*, that describes the total number of items. In the jar case it is 100.
- The number of "successes" in the population, usually denoted *K*, in the example described as the number of red marbles in the jar: 10.
- The sample size, usually denoted *n*. Here it is the number of draws from the jar, 10.

The hypergeometric distribution is defined as:

$$P(x = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

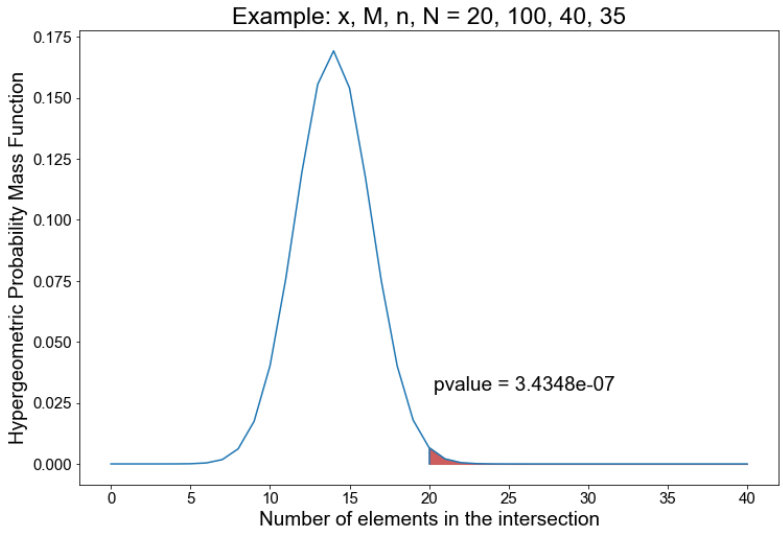



Figure 4.1: Example of application of the hypergeometric distribution.

where $N$ is the population size, $K$ is the number of success states in the population, $n$ is the number of draws (e.g. quantity drawn in each sample) while $k$ is the number of observed successes and the notation

$$\binom{a}{b}$$

depicts a binomial coefficient.

In our case, $N$ describes the universe of genes we were working with, meaning the whole group of genes used in our experiments. $K$, instead, is the number of genes present in the epi-R experiments; n is the number of genes present in the miRNA expression gene set and $k$ is the the number of genes found in the intersection. In practice, we developed a Python (v3.7.1) algorithm with the aim to compute the intersections and their significance, retrieving the `|scipy.stats.hypergeom|` method from the Scipy package (v. 1.2.0). In this method, the parameters are named after a different naming convention than what have been described earlier:

- $M$ is the population size (previously $N$)
- $n$ is the number of successes in the population (previously $K$)
- $N$ is the sample size (previously $n$)
- $x$ is the number of drawn "successes".

With this algorithm, we were able to compute the probability of finding a number of genes in the intersection equal or bigger of the one obtained. We computed the p-value as the area under curve of the right tail of the the Probability Mass Function distribution, that comprised those values that were equal or higher than the number of genes in the obtained intersections, as shown in Fig. 4.1.

### 4.2.4 miRTarBase

To identify putative targets broadly related to the epigenetic pathways we were interested in, we analyzed data from miRTarBase (`http://mirtarbase.mbc.nctu.edu.tw/`, release 6.0), a database which brings information related to experimentally validated miRNA-target interactions (MTIs) (Chou et al. 2017). We differentiated among weak (validated by CLIP-seq) or strong (validated using reporter assays or western blots) miRNA-target interactions by the database annotation.

### 4.2.5 TargetScan

TargetScan (`http://www.targetscan.org/vert72/`) (Agarwal et al. 2015) is a database of predicted biological targets of miRNAs. It was used to compare our results with those expected by a canonical miRNA-target direct-repressive interaction. We downloaded the files for conserved and not conserved miRNA families and sites of interaction (TargetScan release 7.2, March 2018) and we computed the number of predicted targets with no regard for the conservation or the score given to a specific site, in order to maximize the number of miRNA we could add to the analysis. To avoid repetitions, we performed the statistical analysis only on those miRNAs that resulted either normal miRNAs or candidate epi-miRNAs.

### 4.2.6 TCGA data retrieval

Data were downloaded from TCGA GDC Data Portal (`https://portal.gdc.cancer.gov/` , v16.0). For each project, matching files for miRNA expression and mRNA expression were downloaded for each patient case, for both tumor and normal tissue samples. For mRNA expression, FPKM (a normalized expression value that takes into account each gene length and the number of reads mapped to all protein-coding genes (*GDC Data User Guide* n.d.)) data were retrieved; for miRNA expression data, miRNA Expression Quantification files were downloaded (a table that associates miRNA IDs with read count and a normalized count in reads-per-million-miRNA-mapped (*GDC Data User Guide* n.d.)).

### 4.2.7 Differential expression of miRNAs in TCGA data

Matching tumor-normal tissue samples for each miRNA expression dataset were analyzed to compute the differential expression. The analysis was performed using DESeq2 (Love et al. 2014) package of Bioconductor (DESeq2 v. 1.14.1, R v. 3.6.1), after removal of the probes whose row counts sum was very low ($<10$), with standard parameters, comparing tumor samples vs. normal samples. We have set the threshold for differential expression to $\pm 0.5$ of the log2(Fold Change). Values of differential expression were considered significant where their p-values resulted lower than 0.01 (p-value $<0.01$).

### 4.2.8 micro RNA - epi-R correlation in TCGA data

Correlation was computed across-samples with the Pearson method, between each of the analyzed miRNAs and each epi-R component for every TCGA project analyzed. We computed Pearson correlation exploiting the Python (v3.7.1) `scipy.stats.pearsonr(x, y)` command retrieved from the Scipy package (v1.2.0). Only significant correlations were considered (p-value $<0.01$).

## 4.3 Results

### 4.3.1 Pipeline

For this analysis, we focused on the four main epigenetic and repressive pathways that are known to be target of or interact with epi-miRNAs (Amodio et al. 2015). These pathways, as listed before, are PRC1, PRC2, DNMT and HDAC. We searched for the gene sets deriving from epi-R ablation in the state of the art repository of gene signatures, the Molecular Signatures Database (Subramanian et al. 2005; Liberzon et al. 2015), which stores annotated gene sets to use with the Gene Set Enrichment Analysis software. From this web site, we selected and downloaded all the gene sets related to all the sub-components of the aforementioned complexes that were present in the database and that were built according to our criteria of selection.

We obtained a very heterogeneous group of gene sets, including sets of epi-R ablation experiments obtained through knock-out, knock-down or drug treatments or sets of genes that derived from ChIP experiments on epigenetic marks caused by the pathways that we analyzed.

Following these criteria 43 gene sets were retrieved, distributed as described in Figure 4.2. The most abundant group of gene sets was related to the Polycomb Repressive Complex 2 (PRC2), that covered the 46% of epi-R gene sets; Histone Deacetylases (HDAC) gene sets covered the 26% of gene sets; 19% of gene sets were related to Polycomb Repressive Complex 1 (PRC1) and 9% to DNA Methyltransferases (DNMTs) (see Figure 4.2). For the gene sets resulted from epi-R ablation, we downloaded separately both the upregulated and the downregulated gene sets. The number of genes contained in the sets ranged from 11 to more than 900, the majority of them (around 60%) contained up to 150 genes. Some of the sets had a strong overlap in their gene content and we used them as a cross-check of the procedure.

To search extensively for epi-miRNAs exploiting available gene expression datasets, we searched the Gene Expression Omnibus (GEO) database (Barrett et al. 2012; Edgar et al. 2002). In this database are stored and distributed for free genomics data, provided by the research community, derived from a variety of high-throughput analysis (*GEO Overview - GEO - NCBI* n.d.). Searching for a consistent and well curated group of gene sets, we finally chose 14 experiment series (Seki 2011a; Enokida 2012; Naoko Kikkawa

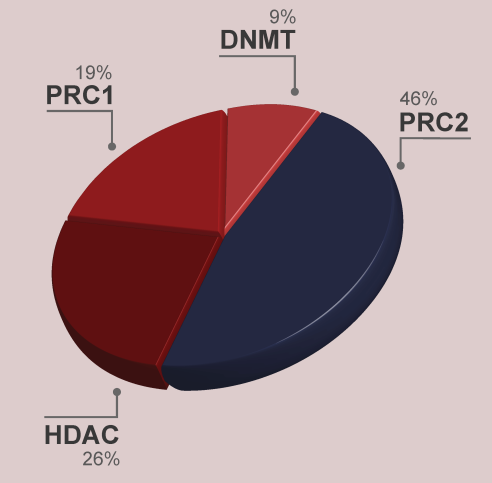



Figure 4.2: Percentage distribution of the MSigDB gene sets. Percentage distribution of the MSigDB gene sets used on the epigenetic pathway taken into consideration.

et al. 2014; Seki 2011b; Nijiro Nohata, Toyoyuki Hanazawa, et al. 2011; Moriya et al. 2012; Fuse et al. 2012; Yoshino et al. 2011; Yamada, Sugawara, et al. 2018; N. Nohata, T. Hanazawa, N. Kikkawa, et al. 2011; Nijiro Nohata, Sone, et al. 2011; Hidaka et al. 2012; Yamasaki et al. 2012; Takashi Kinoshita et al. 2012; T. Kinoshita et al. 2013; Goto, Kojima, Nishikawa, et al. 2015; R. Matsushita et al. 2015; Goto, Kojima, Kurozumi, et al. 2016; Ryosuke Matsushita et al. 2016; Kumamoto et al. 2016; Yonemori et al. 2017; Yamada, Koshizuka, et al. 2018; Yamada, Arai, Sugawara, et al. 2018; Yamada, Arai, Kojima, et al. 2018a; Arai et al. 2018; Yamada, Arai, Kojima, et al. 2018b) (Supplementary file S2). This particular panel of transfections was chosen among the hundreds of miRNA transfection experiments available in the GEO database, because they are characterized by a high level of homogeneity and consistency both from the point of view of the experimental protocol and from that of the data curation. Each series contained an high number of experiments, many of which were performed following protocols of siRNA transfection or drug treatment. Given that we were only interested in the effects of miRNAs on the epigenetic machinery, we selected only those experiments where cell lines were subjected to miRNA transfection. This led us to a total of 157 miRNA transfection experiments (Supplementary file S3) involving 59 different miRNAs on 26 different cell lines, providing us a good variety of both.

We applied our pipeline as depicted in Figure 4.3. After the first step of data acquisition and preparation, we have implemented a script to intersect each of the miRNA transfection gene sets of up-/down-regulated genes with each of the gene sets derived from the MSigDB. A significance score was given to each of these intersections computing the hypergeometric test as described before.

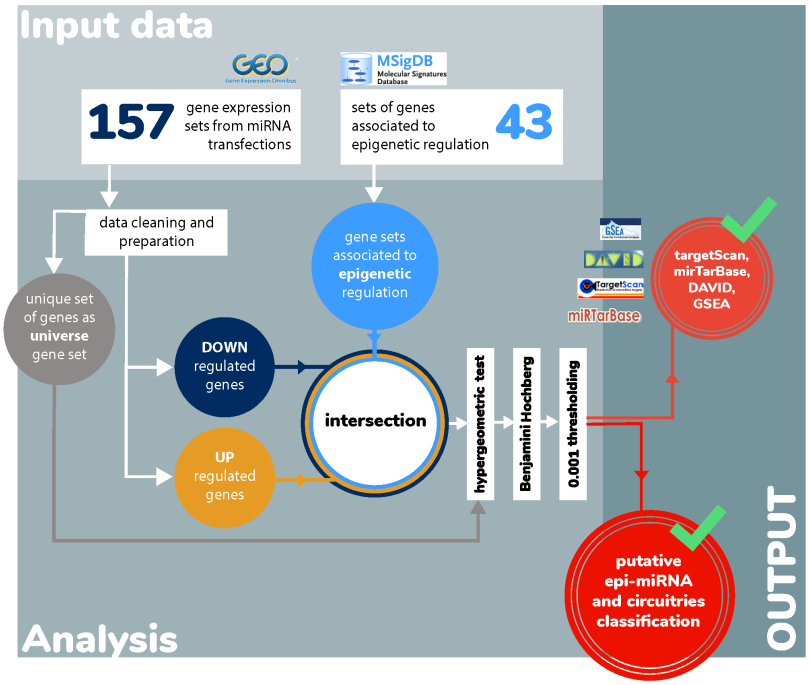



Figure 4.3: Graphical description of the applied pipeline. Each miRNA transfection experiment gene set was divided in up-/down-regulated gene sets, and for each of them was computed the intersection with the MSigDB gene sets. The significance of these intersections has been evaluated through a Hypergeometric test and corrected for multiple testing by the Benjamini–Hochberg procedure. All of the miRNAs involved in intersections with BH corrected p-value lower than 0.001 were selected as candidate epi-miRNAs. Results were further refined through GSEA, DAVID, miRTarBase or TargetScan analysis, and each miRNA-epi-R interaction was labeled on a specific theoretical circuit based on the sign of the miRNA gene set and the type of epigenetic gene set involved in the intersection. BH = Benjamini-Hochberg.

Next, we corrected the p-value obtained with the hypergeometric test with the Benjamini-Hochberg procedure. Looking at the distribution of the obtained BH corrected p-values (Figure 4.4), it was clear that there was a general propensity of miRNAs to interact with the epigenetic layer. Still, these interactions seemed to follow two different behaviours: their bulk resulted in high BH corrected p-values, and could be therefore considered as shallow interactions; on the other side there was, instead, a long tail of interactions that obtained very low BH corrected p-values, and these may represent higher-level interactions.

We decided to set a threshold on the BH corrected p-value that could discriminate between these two behaviors and, looking at the BH corrected p-values' distribution, the choice of the threshold fell on 0.001. All the miRNAs that appeared in the intersections with a lower BH corrected p-value were considered as candidate epi-miRNAs and the epi-Rs analyzed in the gene sets showing significant intersections were considered as their candidate epigenetic interactors.

To exclude possible bias in our analysis, we generated a simple random model, randomizing the original gene expression data by shuffling the gene

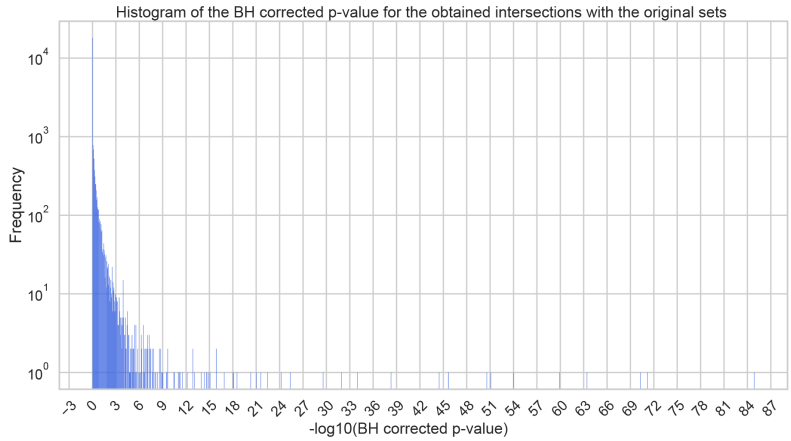



Figure 4.4: The frequency of the BH corrected p-values obtained with our analysis is plotted on the y axis in log scale, while the BH corrected p-values are plotted on the x axis as -log10(BH corrected p-value) transformed values. BH = Benjamini-Hochberg.

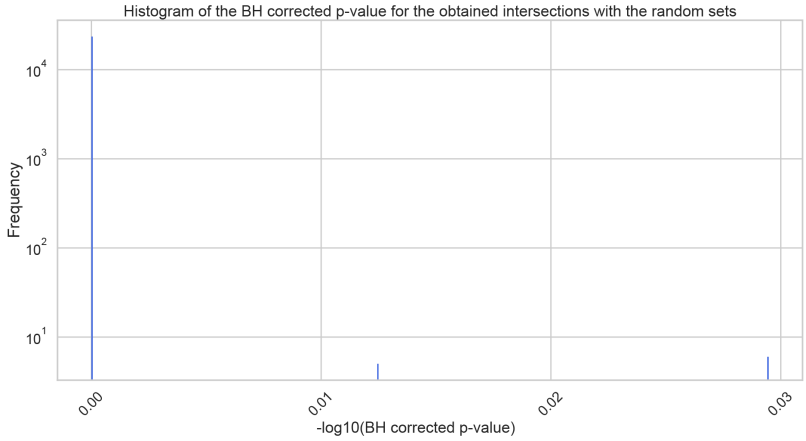



Figure 4.5: The analysis performed on the randomized data didn't lead to any significant BH corrected p-value, excluding the possibility of bias in our analysis. BH = Benjamini-Hochberg.

labels. New random gene sets were created using the same threshold t = $\pm$0.5 on the log10 (fold change) of the expression levels and the analysis was repeated on these random gene sets as described before. The results obtained from the random data didn't lead to any significant BH corrected p-value, as shown in Figure 4.5 indicating that our algorithm was bias-free.

From this pipeline, we found that more than half (57%) of the analyzed miRNAs are involved in epigenetic interactions. As a matter of fact, we found that 34 out of 59 miRNAs resulted as candidate epi-miRNAs: 19 of them were already known to interact with the epigenetic layer of regulation and are reported in Table 4.1, while 15 were identified as new epi-miRNAs; a piechart showing the distribution of the different classes of miRNA/epi-miRNA is shown in Figure 4.6.

For each of the analyzed miRNAs, we tried to map the interactions found between them and the epi-Rs showing significant intersections on simple

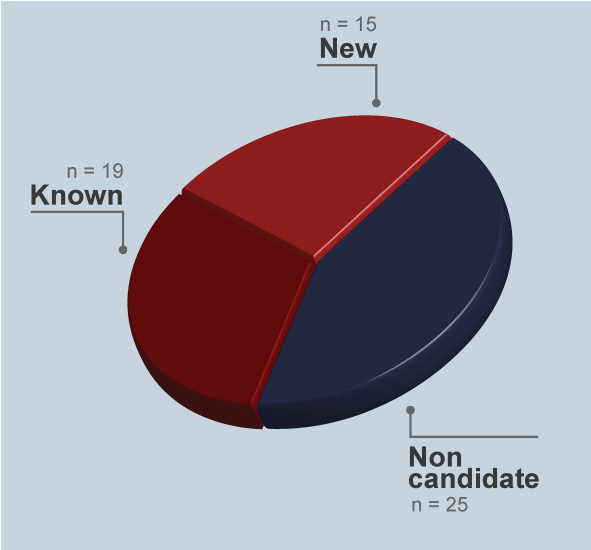



Figure 4.6: Number of known-new epi-miRNAs as resulted by our analysis.

| miRNA | Known | Reference |
|---|---|---|
| miR-1 | Yes | (Gruber and Zavolan 2013) |
| miR-133a | Yes | (Chavali et al. 2012) |
| miR-138 | Yes | (Rastgoo et al. 2016; Osella, Riba, et al. 2014) |
| miR-141 | Yes | (Tamagawa et al. 2014) |
| miR-144 | Yes | (Suzuki et al. 2013) |
| miR-145 | Yes | (Noh et al. 2013; Ramassone et al. 2018) |
| miR-145-5p | No | |
| miR-149 | No | |
| miR-150 | No | |
| miR-183 | Yes | (Gruber and Zavolan 2013) |
| miR-185-5p | No | |
| miR-199 | Yes | (Osella, Riba, et al. 2014) |
| miR-200a | Yes | (Bracken et al. 2008; Memari et al. 2018) |
| miR-205 | No | |
| miR-210 | No | |
| miR-218 | Yes | (Ramassone et al. 2018) |
| miR-221 | Yes | (Ramassone et al. 2018) |
| miR-222 | Yes | (Memari et al. 2018) |
| miR-223 | Yes | (Vasilatou et al. 2013) |
| miR-223-5p | No | |
| miR-224 | No | |
| miR-23b | No | |
| miR-24 | Yes | (Yao et al. 2019) |
| miR-27b | Yes | (Poon et al. 2016) |
| miR-29a | Yes | (Gruber and Zavolan 2013; Ramassone et al. 2018; Osella, Riba, et al. 2014) |
| miR-30a-5p | No | |
| miR-31 | No | |
| miR-375 | Yes | (Bi et al. 2018) |
| miR-451 | No | |
| miR-504 | No | |
| miR-517a | No | |
| miR-874 | Yes | (N. Nohata, T. Hanazawa, T. Kinoshita, et al. 2013) |
| miR-96 | No | |
| miR-99a | Yes | (Coppola et al. 2014) |

Table 4.1: Known and new epi-miRNAs among the 34 candidate epi-miRNAs.

theoretical circuits. To do this, we relied on the kind of information that we could retrieve from the gene expression dataset. The sign of the gene sets, whether they contained up or down regulated genes, has been fundamental to investigate the kind of circuitries in which the two elements where involved. What we obtained is depicted in figure 4.7. All cases in which no significant intersection was observed between the miRNA and the epi-R regulated genes (see Fig. 4.7A) were not further considered in our analysis because we assumed that the two players did not act cooperatively. Regulatory circuitries were instead considered when we found significant intersections in one of the following occurrences:

- between the downregulated genes following miRNA overexpression and the upregulated genes in epi-R ablation experiments (Fig.4.7B);
- between the downregulated genes following miRNA overexpression and the epi-R ablation (Fig.4.7C);
- between the upregulated genes observed in both conditions (miRNA overexpression and epi-R ablation, Fig. 4.7D);
- when the intersections described in Fig. 4.7C and 4.7D were identified for the same miRNA in the same epigenetic pathway (Fig. 4.7E).

It is clear that in this simplified view we cannot take into account undirect interactions and more complex regulatory networks that can lead to similar expression patterns. Therefore, the regulatory circuits in Fig. 4.7 should be seen as the most basic interactions between miRNAs and epi-Rs compatible with the experimental expression patterns. There are cases that are difficult to categorize on the base of these data, for example those in which intersections were found between genes upregulated after miRNA overexpression and genes downregulated after epi-R ablation. This kind of interactions cannot be explained by these simple circuitries and would instead require additional regulatory players.

The overall results of the best intersections obtained for each candidate epi-miRNA and the associated epigenetic pathways are summarized in Figure 4.8. The first column lists the 34 candidate epi-miRNAs we identified. In the second column the association of each miRNA to the putative targeted epigenetic pathway is shown. In third, fourth, fifth and sixth column, the potential loops between the candidate epi-miRNA and epi-R belonging to the categories described in Figure 4.7 are indicated. Finally, in the last column the results obtained searching the miRTarBase database of experimentally validated miRNA-target interactions are indicated (complete list of miRTarBase results in Supplementary file S4). Pathways displaying very strong significance on their intersections (BH corrected p-value$< 10^{-8}$) are reported in bold face (complete list of intersections in Supplementary file S5).

As said at the beginning, more than half (34) of the 59 miRNAs transfected in the analyzed panel of experiments turned out to be candidate epi-miRNAs. It is important to notice, however, that this observation can be

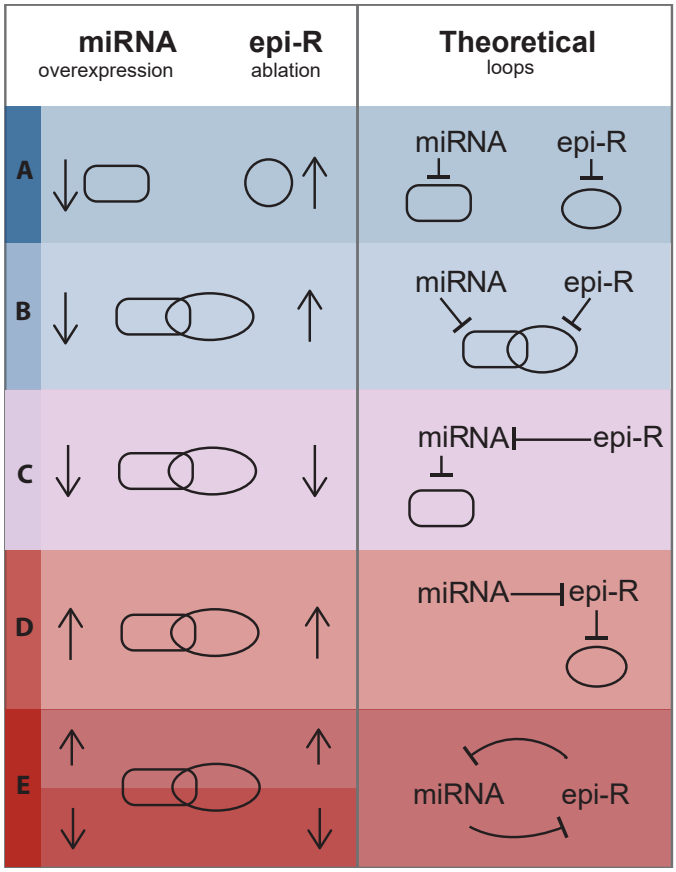



Figure 4.7: miRNA overexpression or epi-R ablation or ChIP experiments were used to identify the pattern of miRNA and epi-R interactions. The putative circuitries between miRNA and epi-R are here depicted. In **A**, miRNA and epi-R do not interact and regulate distinct sets of genes; in **B**, they repress an overlapping set of genes; in **C**, the epi-R represses the miRNA; in **D**, the miRNA represses the epi-R; in **E**, both (C) and (D) situations are present for the same miRNA and the same epigenetic pathway. The up arrows indicate the sets of genes upregulated in the depicted experiments, while the down arrows indicate the sets of downregulated genes. Cases in which the genes were upregulated after miRNA overexpression and downregulated after epi-R ablation were not took into consideration because it was not possible to identify an hypothetical circuit describing their interaction.

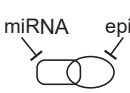
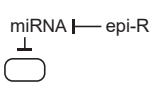
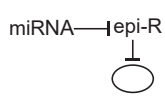
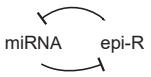
| miRNA | Associated Pathway | B | C | D | E | MirTarBase |
|---|---|---|---|---|---|---|
| | |   |   |   |   | |
| miR-1 | PRC1, **PRC2**, **HDAC** | **PRC2**, HDAC | **PRC2** | **PRC2** | **PRC2** | *HDAC* |
| miR-133a | PRC1, **PRC2**, HDAC | **PRC2** | **PRC2**, HDAC | PRC1, **PRC2**, HDAC | **PRC2**, HDAC | |
| miR-138 | PRC1, DNMT, **PRC2**, **HDAC** | PRC1, DNMT, **PRC2, HDAC** | | | | |
| miR-141 | PRC2 | | | **PRC2** | | |
| miR-144 | PRC2, HDAC | | | PRC2, HDAC | | *PRC2* |
| miR-145 | **PRC2**, **HDAC**, **PRC1**, DNMT | **PRC2** | PRC2 | **PRC2** **PRC1**, DNMT, **HDAC** | PRC2 | |
| miR-145-5p | **PRC2**, PRC1 | | | PRC1, **PRC2** | | |
| miR-149 | PRC1, PRC2, HDAC | | HDAC | PRC1, PRC2 | | |
| miR-150 | PRC2 | | | PRC2 | | |
| miR-183 | PRC2, HDAC | PRC2 | PRC2,HDAC | PRC2 | PRC2 | |
| miR-185-5p | DNMT | | DNMT | | | *DNMT* |
| miR-199 | PRC2 | | | PRC2 | | |
| miR-200a | PRC2 | | | PRC2 | | *PRC2* |
| miR-205 | **PRC2** | **PRC2** | **PRC2** | **PRC2** | **PRC2** | |
| miR-210 | PRC1, PRC2, HDAC | | | PRC1, PRC2, HDAC | | |
| miR-218 | **PRC2**, DNMT, PRC1 | **PRC2** | PRC2 | PRC1, **PRC2** | PRC2 | |
| miR-221 | HDAC, **PRC2**, PRC1 | PRC2, HDAC | PRC2 | **PRC2** | PRC2 | |
| miR-222 | **PRC2**, PRC1, HDAC | PRC2 | HDAC | PRC2 | | |
| miR-223 | PRC1, **PRC2**, HDAC | HDAC | | **PRC2** | | |
| miR-223-5p | **PRC2** | **PRC2** | **PRC2** | | | |
| miR-224 | PRC2 | PRC2 | | PRC2 | | |
| miR-23b | HDAC, PRC2, PRC1 | PRC2 | HDAC, PRC1 | PRC2, HDAC | HDAC | |
| miR-24 | **PRC2**, HDAC, PRC1 | PRC2 | PRC2, HDAC | PRC1, **PRC2** | PRC2 | PRC1, HDAC |
| miR-29a | PRC1, PRC2 | PRC1, PRC2 | | | | |
| miR-27b | PRC2 | | | PRC2 | | |
| miR-30a-5p | PRC1, PRC2 | | | PRC2, PRC1 | | PRC2 |
| miR-31 | DNMT, PRC2, HDAC | | | PRC2, HDAC | | |
| miR-375 | PRC2, PRC1, **HDAC** | PRC2 | **HDAC** | PRC1, PRC2, HDAC | HDAC | |
| miR-451 | PRC2, PRC1 | PRC1, PRC2 | | | | |
| miR-504 | PRC2 | | | PRC2 | | |
| miR-517a | PRC2, HDAC | PRC2, HDAC | | | | |
| miR-874 | PRC2 | PRC2 | | | | |
| miR-96 | **PRC2**, PRC1, HDAC | PRC2 | | PRC1, **PRC2**, HDAC | | |
| miR-99a | **PRC2**, HDAC | **PRC2** | **PRC2**, HDAC | PRC2 | PRC2 | |

Figure 4.8: Table of the identified epi-miRNA interactions. For each candidate epi-miRNA, we list in the second column the pathways in which they are involved (associated pathways): in some cases, a significant intersection was found between the genes upregulated after miRNA overexpression and downregulated after epi-R ablation. These cases were not further analyzed because it was not possible to group them under a single common theoretical circuit. The other cases are listed in columns third, fourth, fifth and sixth, where the involvement of the epi-miRNA in the different loops described in Figure 4.6 is indicated in the headings of the table. The bold face in these columns refers to the pathways associated with very strong signatures (BH corrected p-value $< 10^{-8}$ ). Finally, in the last column, the results of miRTarBase analysis are reported: the italic face in this column indicates the functional targets supported by strong experimental evidence, based on the indications of miRTarBase. BH = Benjamini-Hochberg.

influenced by the fact that most of the miRNAs included in the panel are known to be involved in cancer or in differentiation processes and thus more likely to be involved in epigenetic pathways. In a high number of cases (182 intersections corresponding to 55 different miRNA transfection experiments) we found very low BH corrected p-value for the intersections of the genes upregulated after miRNA transfection with the MSigDB sets, which could be compatible with a direct interaction of the miRNA with an epi-R.

Interestingly, we also found quite a few examples (93 intersections corresponding to 38 different transfection experiments) of significant intersections for the downregulated genes. In a few cases, when transfection experiments of the same miRNA in different cell lines were available, we found that the same miRNA may act as an epi-miRNA in one particular cell line and show no epigenetic effect in the other cell lines.

The results showed a few clear correlations between the different epigenetic pathways analyzed. In general, for several miRNAs, we found signatures for more than one epigenetic pathway (23/34 epi-miRNAs, see Figure 4.8). In some cases, it is difficult to disentangle if these are due to a direct interaction of the miRNA with the epi-R or are triggered by the dysregulation of other epigenetic pathways.

Furthermore, we observed that transfection experiments for our candidate epi-miRNAs are characterized by a large number of regulated genes (Figure 4.9 A-B) with a slight tendency toward upregulated versus downregulated genes (Figure 4.9A), a signature compatible with a widespread action caused by the interactions with the epi-Rs.

As shown by the box-plot analysis, the total number of differentially expressed genes resulted significantly higher for the experiments where the miRNA turned out to be a candidate epi-miRNA compared with experiments where the same miRNA was not identified as a candidate epi-miRNA (Figure 4.9B).

The categorical scatter plot (Figure 4.10A) shows the amount of the differentially expressed genes in the different transfection experiments for each candidate epi-miRNA. Histogram in Figure 4.10B shows that the difference between the distribution of the total number of differentially expressed genes for each transfection experiment and each analyzed miRNA is significantly higher in miRNAs acting as epi-miRNAs as assessed by Mann–Whitney U two-tailed test.

Moreover, a TargetScan analysis of predicted targets of the candidate epi-miRNAs did not show a significant difference in number compared with the predicted target genes of the other miRNAs (Figure 4.11B).

We describe below a few instances of our results, which we chose both for the relevance of the pathways in which the candidate epi-miRNAs are involved and because they represent paradigmatic examples of the different combination of entries that we found in our analysis.

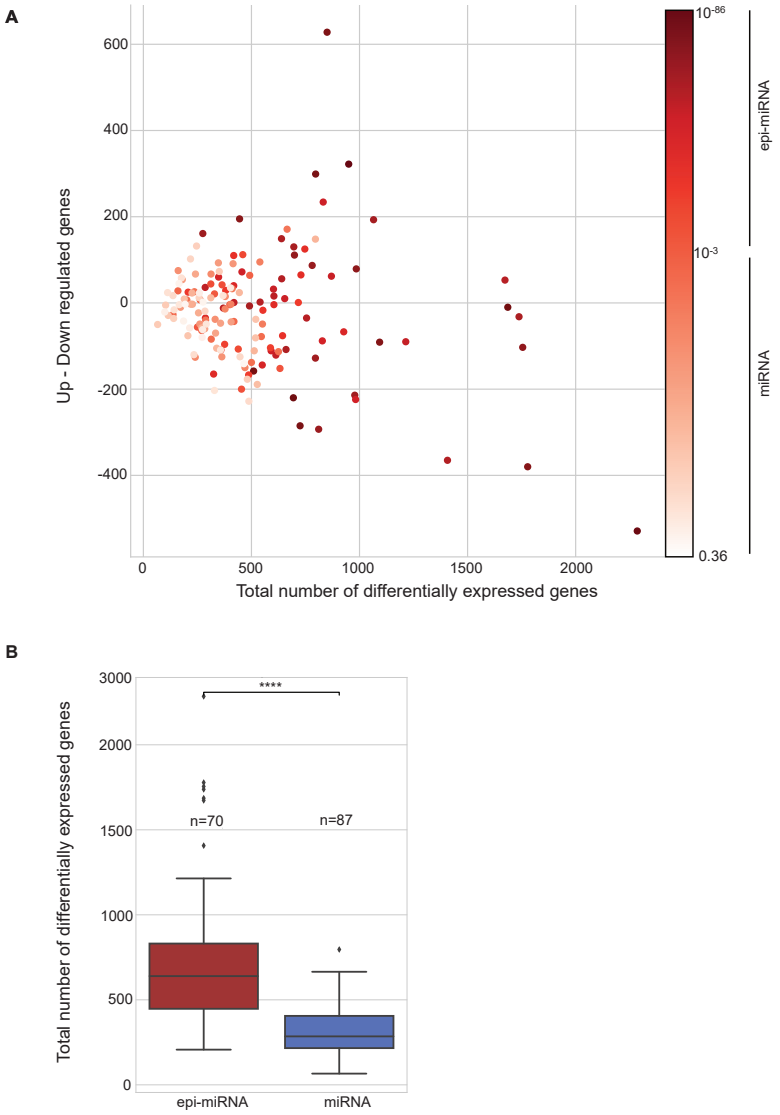



Figure 4.9: epi-miRNAs regulate a higher number of genes than miRNAs not interacting with the epigenetic layer. **A** Scatter plot showing the distribution of the difference among up- and downregulated genes (upregulated genes – downregulated genes) versus the total number of differentially expressed genes for all analyzed transfection experiments on the basis of the significance of their intersection with the MSigDB sets. Each symbol represents one of the transfection experiments listed in Supplementary File S3. The epi-miRNAs and miRNAs identified by the MSigDB intersection analysis are shown in red color gradation as indicated in the BH corrected p-value color bar. **B** Box plots representing the total number of differentially expressed genes for the indicated (n) number of experiments, discriminating between those where the miRNAs turned out to behave as epi-miRNAs (red, left) and those where they acted as regular miRNAs (blue, right). In (A) BH corrected p-value was obtained using a Benjamini–Hochberg procedure. In (B), significance was assessed using a Mann–Whitney U two tailed test, ****p-value $<$ 0.0001; - a non statistically significant p-value. BH = Benjamini-Hochberg.

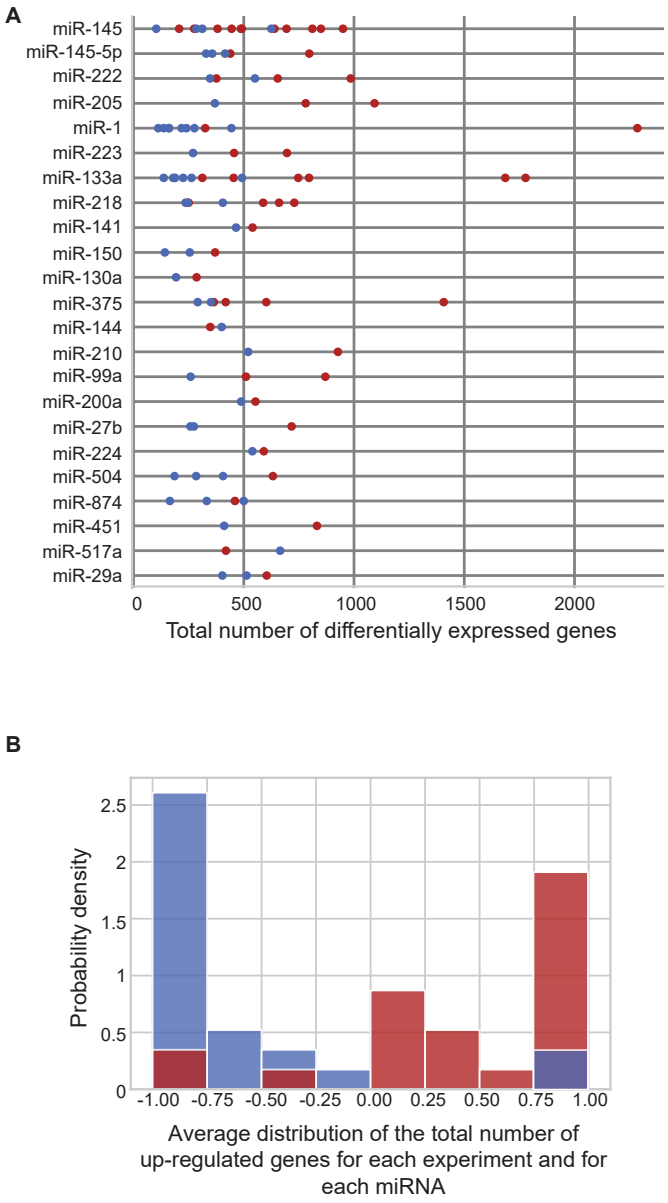



Figure 4.10: Single candidate epi-miRNA experiments show a larger number of differentially expressed genes than miRNAs not interacting with the epigenetic layer.**A** Categorical scatter plot representing the total number of differentially expressed genes in all the considered experiments for each identified epi-miRNA. Red color indicates experiments in which the miRNA was considered an epi-miRNA (BH corrected p-value threshold 0.001), while blue color indicates experiments where the miRNA was not considered an epi-miRNA (BH corrected p-value $>$0.001).**B** Histogram showing the distribution of the average of the total number of differentially expressed genes for each transfection experiments and each analyzed miRNA. Red: experiments where the miRNA was considered an epi-miRNA. Blue: all the experiments in which the miRNA was not considered an epi-miRNA. BH = Benjamini-Hochberg.

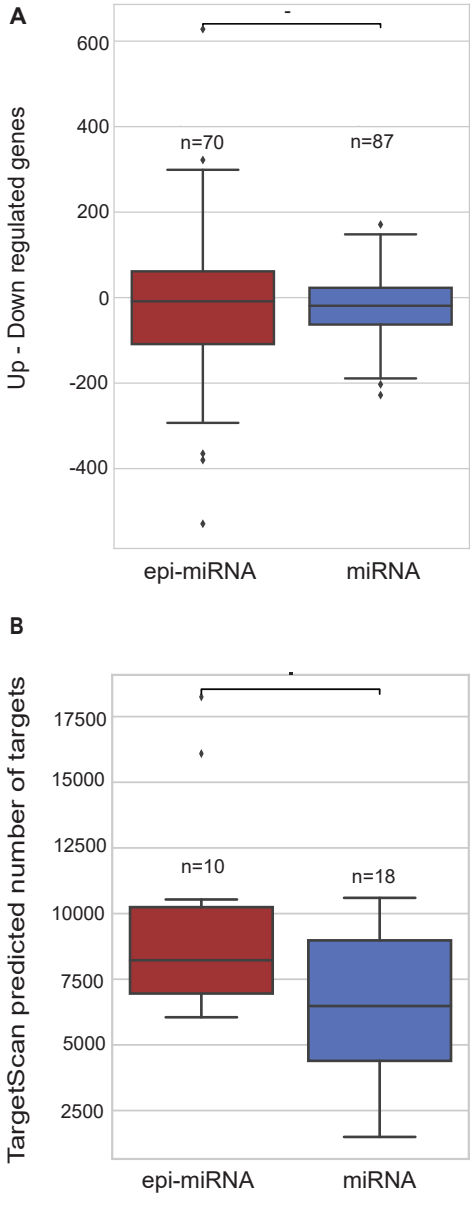



Figure 4.11: Analysis of the difference between upregulated and downregulated genes for candidate epi-miRNA and miRNA. **A** Box-plots representing the difference among the number of up and downregulated genes of the indicated (n) number of experiments for both candidate epi-miRNAs and miRNAs. **B** Box-plots representing the number of TargetScan predicted genes for both candidate epi-miRNAs and miRNAs. miRNAs behaving only as epi-miRNA or only as a plain miRNA in all related experiments were considered for the analysis. In (A) and (B) significance was assessed using a Mann-Whitney-U two tailed test, ****=p-value $< 0.0001$; - a non-statistically significant P-value.

### 4.3.2 Examples of candidate epi-miRNAs identified with our pipeline

#### miR-31

In the panel of the analyzed transfection experiments, miR-31, a typical miRNA deregulated in cancer, was present in a single transfection experiment in the human highly metastatic prostate cancer cell line, PC3. Looking at the intersections with our gene sets, four entries were observed (BH corrected p-value $< 0.001$, see Supplementary File S6). All of them were for the set of genes upregulated after transfection (NUYTTEN_EZH2_TARGETS_UP, SENESE_HDAC1_TARGETS_UP, SENESE_HDAC3_TARGETS_UP and GSE27434WT_VS_DNMT1_KO_TREG_DN). Three sets contained genes upregulated after knockout of EZH2, HDAC1 and HDAC3, suggesting a direct inhibition of these three genes by miR-31 or more generally on the corresponding epigenetic machinery as shown in the third column of Figure 4.8.

#### miR-205

miR-205 was not previously known as an epi-miRNA. For miR-205, three experiments on three prostate cancer cell lines (PC3, DU145 and C4-2) were present in our panel. A larger number (seven) of significant intersections with our gene sets both for the upregulated and downregulated genes was observed with very low BH corrected p-value (Supplementary File S6). Interestingly, for one of the EZH2-related files, the BH corrected p-values are really low (down to $10^{-46}$) and genes show consistent signatures in the upregulated and downregulated sets. More precisely, the upregulated genes are the same found upregulated after knockout of EZH2 (NUYTTEN_EZH2_TARGETS_UP, BH corrected p-value$= 10^{-21}$) and downregulated those which are downregulated after knockout of EZH2 (NUYTTEN_EZH2_TARGETS_DN, BH corrected p-value $= 10^{-46}$).

#### miR-138

Differently from the previously described miRNA, miR-138 is a well-known epi-miRNA (Gruber and Zavolan 2013), in this case only one transfection experiment in the A498 kidney carcinoma cell line was present in the considered datasets. Eight intersections (BH corrected p-value $< 0.001$) were observed and all of them in the set of genes downregulated after transfection. Noteworthy, in all these cases, the gene sets are incoherent with respect to the expected epi-miRNA interaction. For instance, we found among the downregulated genes a large number (BH corrected p-value $= 10^{-18}$) of genes upregulated after knockout of EZH2. This means that these genes are simultaneously targets of miR-138 and EZH2. We report as an example the intersection corresponding to the file NUYTTEN_EZH2_TARGETS_UP mentioned above. This intersection is composed by 95 genes which are listed

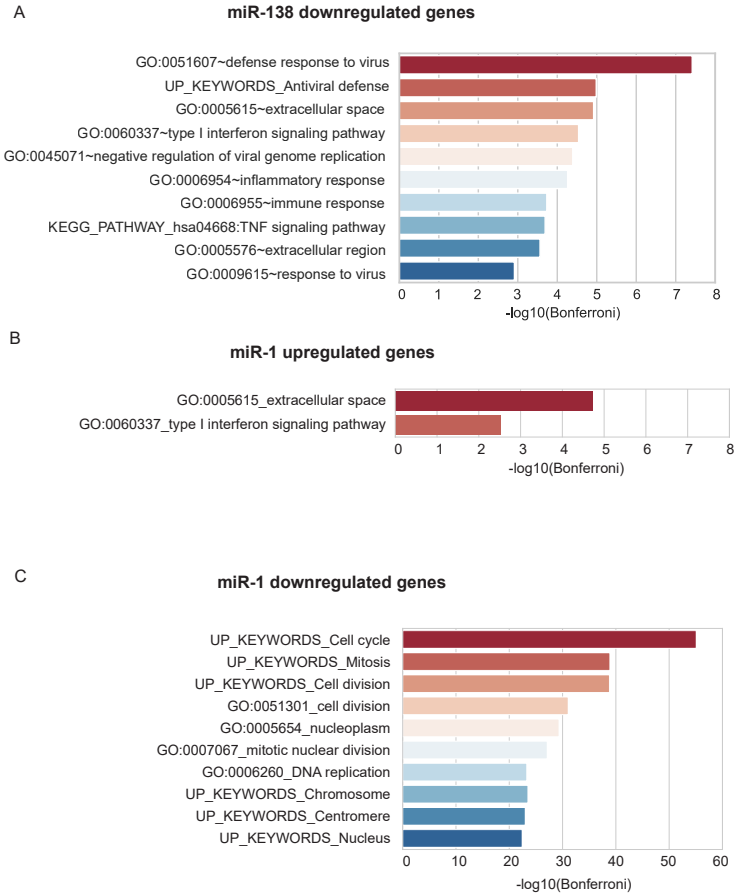



Figure 4.12: GO analysis for miR-138 and miR-1. **A–C** Bar plots of the best entries for DAVID GO analysis performed for the genes of the best intersections with MSigDB sets obtained for (A) miR-138 and (B–C) miR-1. For each entry, the log10 of the Bonferroni corrected p-value is shown (cutoff on the Bonferroni corrected p = 0.001).

in the Supplementary File S6 (BH corrected p-value = $10^{-18}$). Performing a Gene Ontology analysis (DAVID tool online `https://david.ncifcrf.gov/home.jsp`) (Huang et al. 2009; Huang et al. 2008), entries related to immune response with a low Bonferroni corrected p-value of $10^{-8}$ were observed (Figure 4.12A), suggesting that a large portion of these genes is activated only following a perturbation of the immune system and should be otherwise kept silenced.

## miR-145

The well-known epi-miRNA, miR-145, is the miRNA with the largest number of available transfection experiments of our panel, in some cases on the same cell lines. miR-145 is simultaneously involved in several different epigenetic pathways (see Figure 4.8). In particular, we found strong evidence of a direct involvement in the regulation of the two Polycomb pathways, PRC1 and PRC2, and of the DNMT machinery. Most relevantly, the same pathway is regulated in several different cell lines, for instance, PRC2 is downregulated

not only in LNCap but also in PC3 and DU145 prostate cancer cell lines. Interestingly, a direct inspection shows that the effect of the downregulation of PRC2 is very similar in the three cell lines. In other words, there is a large overlap among the genes that are upregulated in the three experiments (notwithstanding the difference in cell lines) and these upregulated genes are exactly those which are targeted by PRC2 (data not shown).

#### miR-1

A very interesting pattern was found in the case of the well-known epi-miRNA, miR-1 (Gruber and Zavolan 2013), which appeared in ten entries of our panel of transfection experiments, but only for two of them, corresponding to the PC3 (prostate cancer) and EBC-1 (lung cancer) cell lines, the intersection had a BH corrected p-value $< 0.001$. In particular, in lung cancer EBC-1 cells, there was a strong indication supporting a direct downregulation of PRC2. What is relevant is that this signature appears in a consistent way both in the set of the upregulated and downregulated genes (see data in Supplementary File S6). We found a large overlap (BH corrected p-value $= 10^{-72}$) between the set of genes downregulated after transfection of the miRNA and the genes downregulated after knockout of EZH2 and, simultaneously, a large overlap (BH corrected p-value $= 10^{-15}$) of the genes upregulated after transfection with those upregulated after knockout of EZH2. The set of common upregulated genes shows a strong overlap with the same GO categories mentioned above on immune response (Figure 4.12B), while a GO analysis of the set of downregulated genes shows that most of these genes are associated to cell cycle (Figure 4.12C). GO categories for miR-31, miR-145 and miR-205 are shown in Figure 4.13.

### 4.3.3 Identification of the putative epigenetic interactors of a given miRNA

#### The Dataset

The procedure we proposed can be used, in addition to identifying epi-miRNA candidates, given a miRNA of interest, to identify its candidate epigenetic interactors. Due to the relevance of miR-214 in tumor progression, the Taverna laboratory at the Molecular Biotechnology Center (MBC) in Turin is continuously investigating unknown mechanisms through which miR-214 could coordinate tumor metastatization, and this gave us the opportunity to test our approach on the original dataset that they provided.

In the attempt to identify new genes directly or indirectly modulated by miR-214, MA-2 (Penna, Orso, Cimino, Tenaglia, et al. 2011) aggressive melanoma cells were transiently transfected with a precursor for miR-214 (pre-miR-214) or negative controls (pre-control). RNA was extracted 72 h following transient transfection and miR-214 expression levels verified by real-time PCR (data not shown). RNA was used to perform whole human genome gene expression analysis (GSE124965). A total of 493 differentially expressed

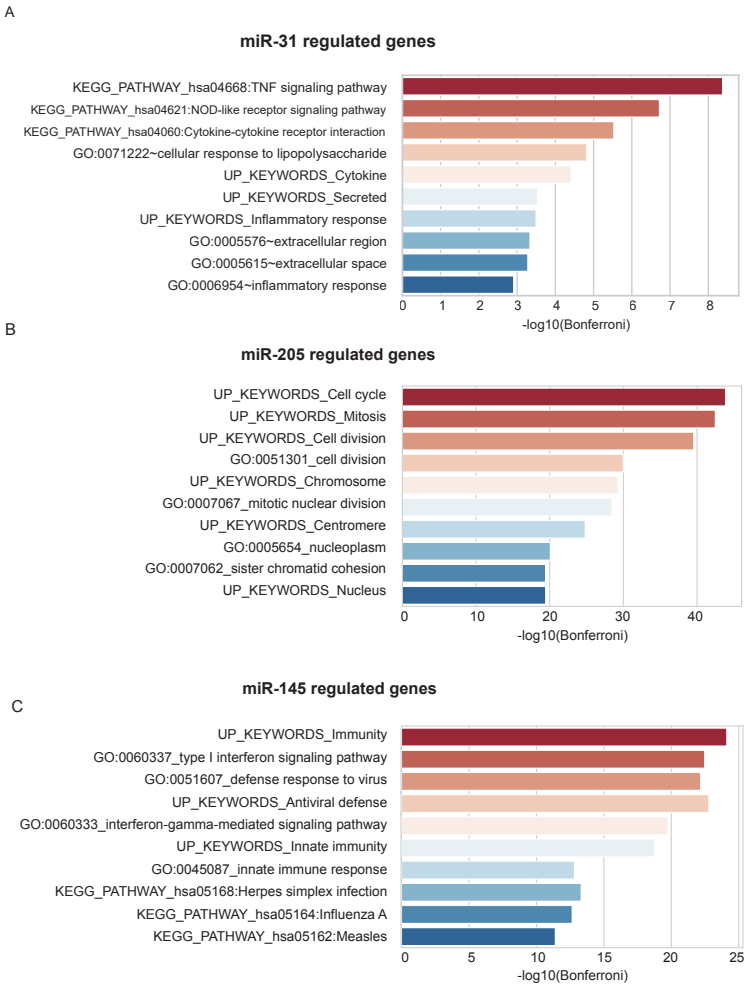



Figure 4.13: GO analysis for miR-31, miR-145 and miR-205. **A-C** Bar-plots of the ten best entries for DAVID GO analysis performed for the genes of the best intersections with GSEA sets obtained for (A) miR-31 and (B) miR-205 (C) miR-145. For each entry, the log10 of the Bonferroni corrected p-value is shown.

genes (342 upregulated; 88 downregulated) were found, considering a 14.2% FDR. Surprisingly, 342/493 of the differentially expressed genes resulted to be upregulated, suggesting a possible indirect regulation of these genes by miR-214 via a potential crosstalk with an epi-R. Similar observation came from the analysis of another miR-214 dataset (GSE35948) in human myeloma cell lines, NCI-H929 and MM1S (Misiewicz-Krzeminska et al. 2013). The first hypothesis we made was that miR-214 could influence the expression of this huge amount of genes via an epigenetic mechanism. For this reason, the list of the differentially expressed genes was subjected to the analysis described above to identify candidate epigenetic interactors of miR-214. The lists of the differentially expressed genes used for the analysis is available in Supplementary File S7. The heatmaps of differential expression for both datasets are shown in Figure 4.14 A and B, where it is qualitatively appreciable the increased number of upregulated genes compared with downregulated ones following miR-214 overexpression.

Eight MSigDB sets were identified, across up- and downregulated gene sets using both our (GSE124965) and GSE35948 (Misiewicz-Krzeminska et al. 2013) datasets (Supplementary File S8), showing evidence of a direct regulation of the miRNA on the considered epigenetic pathways. Five of them correspond to the PRC2 pathway, two correspond to HDACs and one to the PRC1 pathway. In particular, the PRC2 sets are related to the silencing of the enhancer of zeste 2 polycomb-repressive complex 2 subunit (EZH2) and to the silencing of the EZH2 interactor NIPP1 (Supplementary File S8), suggesting EZH2 as the candidate gene directly targeted by miR-214 and responsible for the vast upregulation observed in the experimental dataset. GSEA analysis of the genes belonging to the best PRC2 intersection, described above, shows a significant enrichment with sets related to epithelial-to-mesenchymal transition and TGFb pathway in agreement with the prometastatic role of miR-214 (Figure 4.15).

### miR-214 downregulates EZH2

In order to prove that miR-214 targets EZH2 gene in a melanoma cell context, miR-214 was overexpressed in MA-2 melanoma cells, and EZH2 protein expression levels were analyzed. As shown in Figure 4.16A, miR-214 overexpression leads to a 60% reduction of EZH2 protein levels.

Further analysis were performed to check if EZH2 silencing could phenocopy miR-214 overexpression effects on the activation of miR-214-dependent downstream pathway. ITGA5, ALCAM and miR-148b expression levels were analyzed. As shown in Figure 4.16B and C, EZH2 silencing (60%) was able to induce an upregulation of ALCAM (32%) and ITGA5 (60%) expression, thus phenocopying the effects observed following miR-214 overexpression (Bommi et al. 2010). Interestingly, ALCAM and ITGA5 were not consistently modulated at the mRNA level. In order to asses if EZH2 could be partially involved in miR-214-mediated miR-148b down-modulation, miR-148b expression levels were tested following EZH2 silencing as shown in Figure 4.16D. EZH2 silencing was able to decrease miR-148b expression, suggesting a partial con-

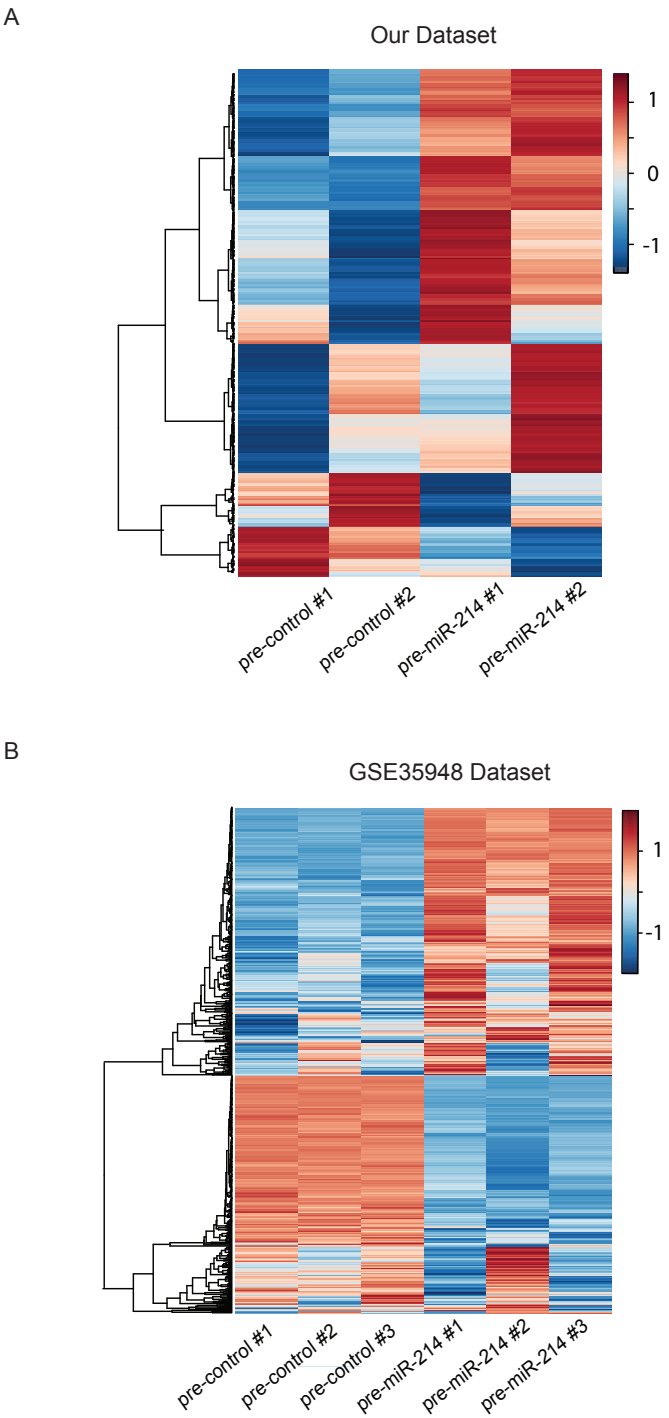



Figure 4.14: Heatmaps of the expression levels of miR-214-regulated genes. **A** Heatmap representing the results of the microarray analysis performed on MA-2 melanoma cells 72 h following miR-214 transient overexpression (GSE124965). **B** Heatmap representing the results of the microarray analysis of GSE35948 dataset.

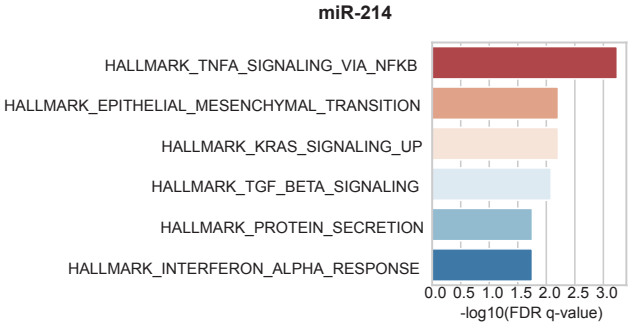



Figure 4.15: Functional enrichment for the best miR-214 intersection with MSigDB sets. Best entries for enrichment analysis performed on GSEA hallmark collection for the best intersection of miR-214 regulated genes with MSigDB sets. For each entry, the log10 of the FDR q-values is shown. FDR: False discovery rate

tribution of EZH2 in miR-214-mediated control of miR-148b. All these data suggest that miR-214 could potentially act, at least partially, through an epigenetic mechanism together with the classical mechanism of action in controlling metastatization.

### 4.3.4 TCGA Analysis

The results that we are going to report are based upon data generated by the TCGA Research Network (`https://www.cancer.gov/tcga`). The Cancer Genome Atlas (TCGA) is an important tool for cancer research: it contains more than twenty thousands molecularly characterized samples, both from cancer and normal tissue, comprising matching cancer-normal samples, derived from 33 different cancer types (*The Cancer Genome Atlas Program* n.d.).

More in detail, we began selecting a subset of the projects representing tumors in which miRNAs were already known to have an important role, obtaining 18 projects. These 18 projects comprised a great variability in terms of tumor types and number of samples, from those more common containing a huge number of samples, to those more rare for which a smaller number of samples was available.

For this analysis we needed both mRNA and miRNA expression data for each patient case. Furthermore, each patient case should have had both tumor and normal tissue samples. Adding these more stringent criteria to our search, some of the more rare tumors or those tumors that derived from tissues on which it is difficult to make a biopsy (as brain, e.g.), or which were obtained from liquid tissues, were excluded. Samples from two of them, TCGA-READ and TCGA-COAD, were considered together in order to obtain colorectal cancer data, and will be therefore named TCGA-READ-COAD from here on. The final number of analyzed TCGA projects is 12. A list of the analyzed TCGA projects is reported in table 4.2.

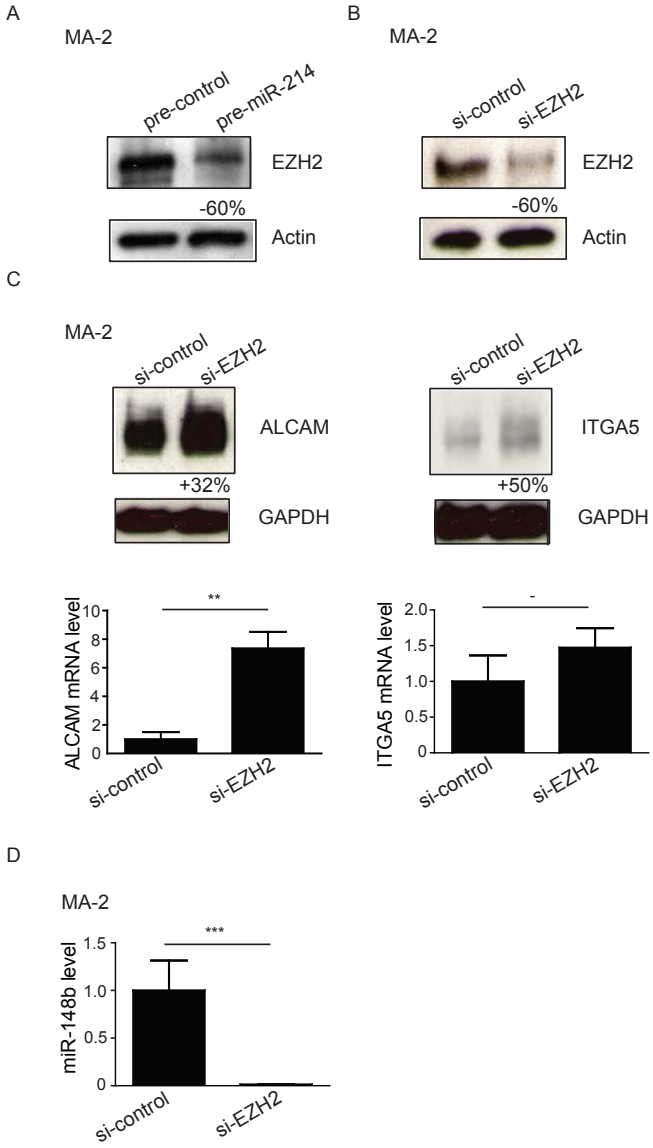



Figure 4.16: Histone methyltransferase enhancer of zeste homolog 2 is a target of miR-214. **A** and **B** EZH2 protein levels were assessed by WB in MA-2 cells 72 h after transfection with miR-214 precursors or their negative controls (pre-214 vs control) or after EZH2 transient downmodulation (si-EZH2 vs si-control). **C** ALCAM and ITGA5 protein and mRNA levels were analyzed by WB and qRT-PCR in MA-2 cells 72 h after EZH2 transient downmodulation (si-EZH2 vs si-control). **D** miR-148b expression levels were assessed by qRT-PCR in MA-2 cells 72 h after EZH2 transient downmodulation (si-EZH2 vs si-control). Protein modulations were compute relative to controls, normalized on Actin (B and C) or GAPDH (D) loading controls and expressed as percentages; results in (C) and (D) are shown as fold changes (mean  SEM) relative to controls, normalized on GAPDH mRNA or U6 snRNA levels. Two experiments with independent protein and RNA preparations were performed and representative ones are shown. EZH2: Histone methyltransferase enhancer of zeste homolog 2; WB: Western blot.

| Project ID | Project Name | Number of Samples |
|---|---|---|
| TCGA-BLCA | Bladder Urothelial Carcinoma | 405 |
| TCGA-PRAD | Prostate Adenocarcinoma | 491 |
| TCGA-CHOL | Cholangiocarcinoma | 36 |
| TCGA-HNSC | Head and Neck Squamous Cell Carcinoma | 495 |
| TCGA-STAD | Stomach Adenocarcinoma | 372 |
| TCGA-ESCA | Esophageal Carcinoma | 161 |
| TCGA-KIRC | Kidney Renal Clear Cell Carcinoma | 865 |
| TCGA-CESC | Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma | 304 |
| TCGA-READ-COAD | Rectum Adenocarcinoma, Colon Adenocarcinoma | 598 |
| TCGA-LIHC | Liver Hepatocellular Carcinoma | 367 |
| TCGA-BRCA | Breast Invasive Carcinoma | 1072 |
| TCGA-PAAD | Pancreatic Adenocarcinoma | 177 |

Table 4.2: Description of the TCGA projects analyzed, as reported on the GDC Data Portal.

### miRNA differential expression

Given that the main goal of this last part of the project was to analyze if the miRNA-epigenetic components interactions that we found were retrievable also from tissue samples, the analysis was focused on the 34 miRNAs that resulted candidate epi-miRNAs from the prior analysis. We tried to mimic the experimental conditions of the previous pipeline. The condition most similar to miRNA transfection that we could imagine in tissue samples was computing where these miRNAs were differentially regulated. As a consequence we performed a miRNA differential expression analysis on the matching tumor-normal tissue samples to search for those tumors in which our candidate epi-miRNAs resulted either up or downregulated. The results from the differential expression analysis are shown in Figure 4.17.

From the data shown in 4.17, some general considerations can be drawn:

- some miRNAs are downregulated in most of the samples: this behaviour is clear for miR-133a, miR-29a, miR-99a, miR-145 and miR-23b, and it is possible that they may act as oncosoppressors (Frixa et al. 2015);

- similarly, some other miRNAs are generally upregulated: miR-517, miR-141, miR-183, miR-210, miR-96 and miR-31 are the examples of miRNAs that resulted almost always upregulated from the differential expression analysis; this is usually the case of miRNAs that are oncogenic (Frixa et al. 2015);

- there are miRNAs whose behaviour is more heterogeneous and therefore difficult to categorize: they could be highly upregulated in some cases and downregulated in others ( like the case of miR-205, or miR-144);

- for a few miRNAs we obtained a small number of cases in which they were differentially expressed (like miR-874 or miR-150), and for them it is difficult to evince a general behaviour.

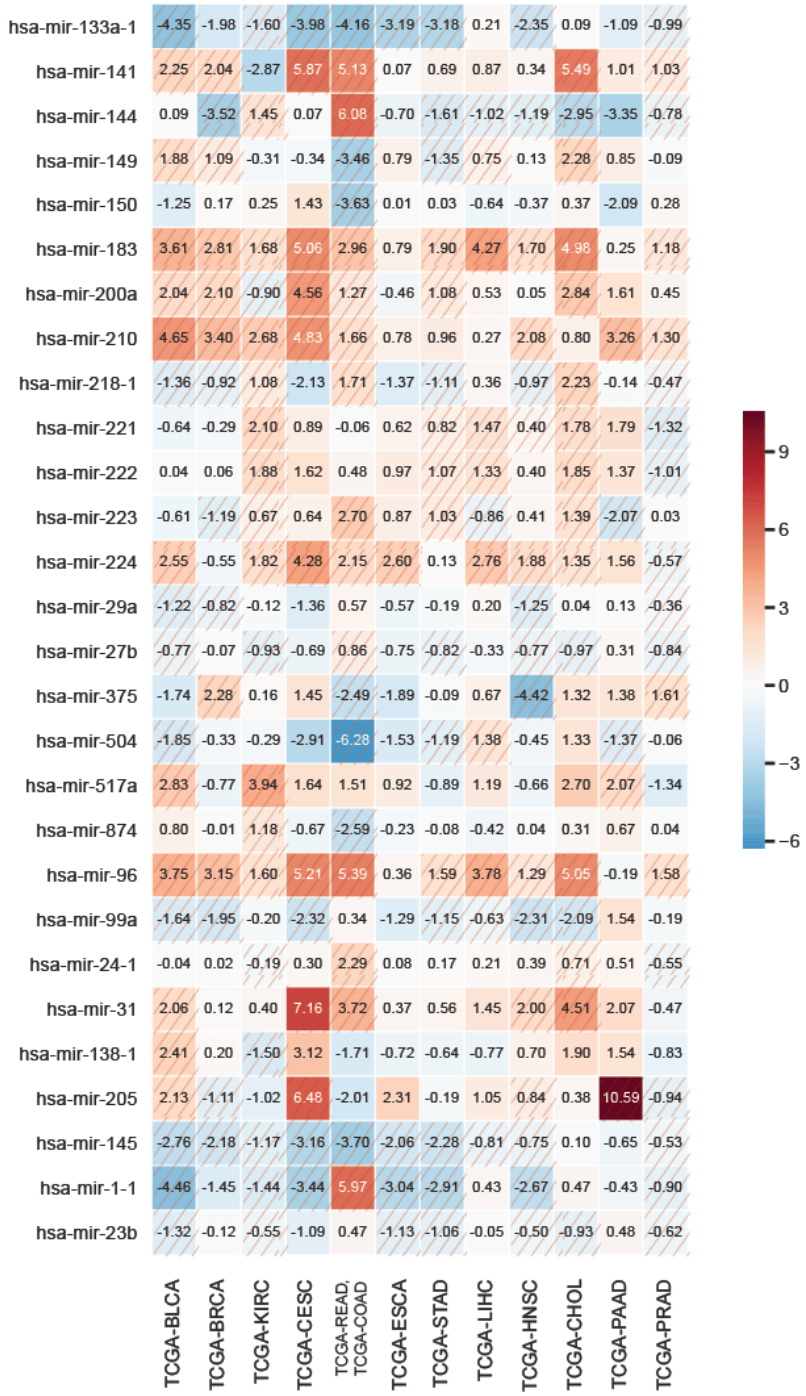



Figure 4.17: miRNA differential expression analysis for each TCGA project analyzed. For each miRNA, the values of differential expression are shown as annotation on the corresponding cell (tumor tissue samples vs normal tissue samples), and the color coding helps to visualize whether it is positive or negative. The colorbar indicates the log2FC of the differential expression. The values in red-dashed cells are those reporting a significant differential expression (p-value$<$0.01).

The overall results show that candidate epi-miRNAs seem strongly involved in cancer, with almost half of them (16/34) being differentially expressed in 6 or more of the TCGA projects analyzed. Obviously, the recorded dysregulation may be partially due to the overall gene expression dysregulation of cancer, and we put a low threshold on the significance of the differential expression to be more confident that these fluctuations were not random.

### miRNA - epigenetic components correlations

To identify which miRNA resulted up or down-regulated in each tumor type, a threshold of $\pm 0.5$ was set on the log2(Fold Change). Next, we focused on miRNAs interactions with the epigenetic components. More precisely, we analyzed their correlation with the specific epigenetic components taken into account in our previous pipeline in each of the TCGA projects considered.

We computed the correlation between the 34 candidate epi-miRNAs and the epigenetic components, to investigate their relationships. We found a very heterogeneous landscape from which it was very difficult to evince precise patterns of interactions. This can be due to two main reasons: we already realyzed from the previous analysis that miRNAs can act as epi-miRNA in a cell-type specific manner, and this may be true even for different contexts. miRNAs may be part of the epigenetic regulatory network in some specific tumors and not in others. Moreover, correlation does not imply causality, and a direct interaction between the miRNA and the epigenetic component is not the only way to explain the correlation between the expression levels of the two components, but many other undirect interactions could take place in the middle.

Anyway, the results of the correlations could still be a good measure of the presence of relationships between the miRNAs and the epigenetic components, and can be considered as signatures of the importance of these miRNAs in epigenetic regulation.

The miRNAs showed many positive correlations. As a matter of fact, in some cases positive correlations are almost the only resulting correlations: one peculiar example is that of miR-183, reported in Figure 4.18B. From the prior analysis, it resulted involved in feedback loops with HDAC and PRC2. From the differential expression analysis, it resulted upregulated in 10/12 TCGA projects, and showed only positive correlations (and no negative correlations) in 9/10 of them, with all the epigenetic pathways analyzed. A similar behaviour was observed for miR-96 too, and this is not strange, considering that miR-183 and miR-96 belong to the same cluster of miRNAs. miRNAs belonging to the miR-183 family are recurrently highly expressed in cancer, suggesting that these miRNAs may lead to the gain of some function which may be beneficial for carcinogenesis (Dambal et al. 2015). Generally speaking, the miRNAs that resulted upregulated in a high number of tumors, also resulted strongly positively correlated with the epigenetic components, suggesting that those miRNAs could be involved in pathways that need to be upregulated in cancer. Unfortunately, no hypothesis could be made on

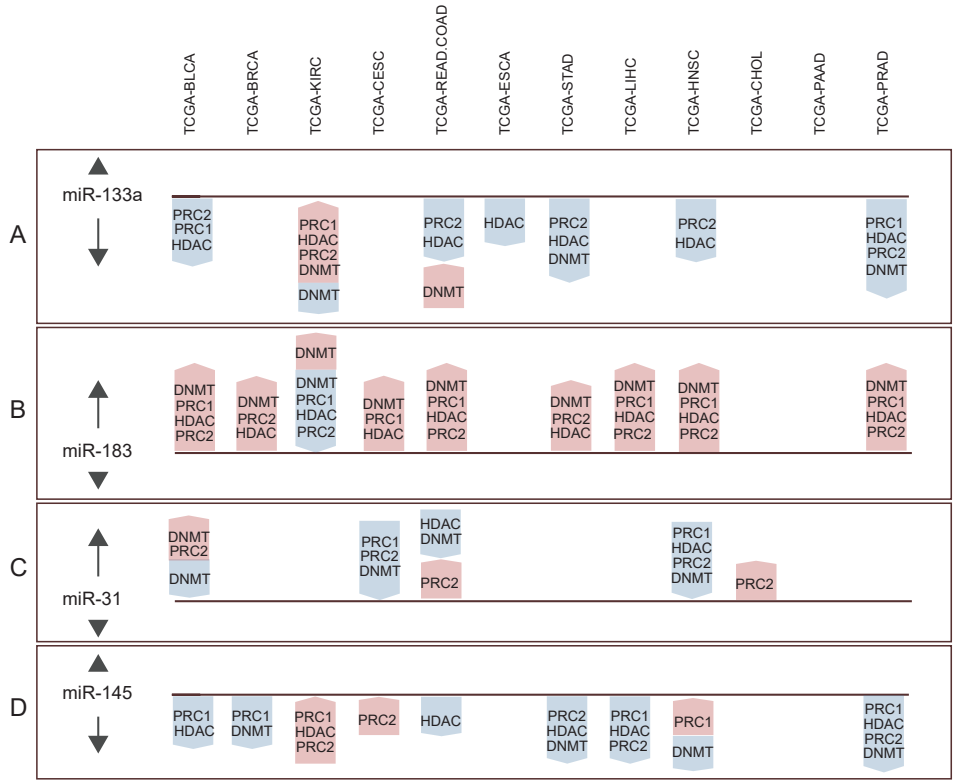



Figure 4.18: Correlations between miRNAs and the epigenetic components in the different TCGA projects. In each box, different miRNAs and their correlations are depicted. For each miRNA the correlations within the tumors where they resulted upregulated are depicted above the midline, while those within the tumors where they resulted downregulated are depicted below the line. The pathways that resulted negatively correlated with the miRNA are listed on blue down arrows, those that resulted positively correlated are depicted on red up arrows.

the kind of circuits in which the miRNAs and the epigenetic components are involved based on these kind of information.

In other cases, a high number of negative correlations have been found. A negative correlation would suggest that the miRNA and the epigenetic component are involved in some kind of negative feedback loop (similar to the theoretical loops depicted in 4.7C-D).

For example miR-133a (Figure 4.18A), member of the miR-133 family, which appears to be dysregulated in many human tumors including gastric cancer, non-small cell lung cancer and pituitary tumors (Li et al. 2018), resulted downregulated in 8/12 TCGA projects and in most of them (7/8), showed negative correlations with all the four epigenetic pathways analyzed. In TCGA-PRAD, containing samples derived from prostate cancer tissues, it showed negative correlations with all the four epigenetic pathways, suggesting a strong epigenetic interplay in that cancer type.

miR-145 expression is reduced in various cancer cell lines (Zeinali et al. 2019) and resulted always downregulated (where significantly differentially expressed, in 10/12 projects, Figure 4.17) from our analysis too. Again, in prostate cancer, showed significant negative correlation with all the four

pathways (4.18D), and they all resulted as its targets in the previous analysis too (see Figure 4.8D).

Another interesting case is that of miR-31. It is notoriously involved in cancer, either as tumor suppressor or as carcinogen (T. Yu et al. 2018). In our analysis, resulted significantly differentially expressed in 5 TCGA projects, and upregulated in all of them (Figure 4.17). Interestingly, it showed negative correlations in 4/5 tumors, and the same pathways that resulted to be associated with its action from our analysis, DNMT, PRC2 and HDAC, resulted negatively correlated in these samples (4.18C). Moreover, also PRC1 resulted negatively correlated to its expression, in TCGA-CESC and TCGA-HNSC. It is possible that miR-31 interacts in a feedback loop with this pathway only in the specific tissues from which these samples derived. Many other correlations have been found, between the other miRNAs and the epigenetic components, and are reported in Supplementary figure S9.

# Chapter 5

# Conclusion and future perspectives

## 5.1 Discussion

Epigenetics encloses a wide range of pathways and modifications that affect the DNA and its transcription, without affecting the sequence. The importance of this layer of regulation is even greater once that all the interactors with the epigenetic pathways are taken into account. Among these interactors, miRNAs result fundamental for the establishment of specific dynamics. miRNAs that control epigenetic components, and in some cases are controlled by them generating feedback loops, are called epi-miRNAs. This specific category of ncRNAs has been seen to interact with the epigenetic components at many levels, but their identification was still far from being exaustive.

We tried to pose the foundations for epi-miRNA identification, developing a pipeline able to detect which miRNAs have epi-miRNA function or which are the interactors of a miRNA of interest. We selected a panel of gene sets which can be considered *bona-fide* targets of the above discussed main epigenetic pathways and intersected them with the set of genes respectively upregulated or downregulated in a series of gene expression datasets coming from publicly available transfection experiments of various miRNAs. Intuitively, a large intersection could be considered as a signature of an epigenetic amplification. The p-value of the intersection was assessed using the hypergeometric distribution corrected for multiple testing.

The main result of our analysis is that the number of putative epi-miRNAs is much larger than expected: it is indeed difficult to estimate how broad the epi-miRNome can actually be. Also, from our analysis resulted that epi-miRNAs are typically involved in differentiation processes and are likely to play an important role in cancer development. Indeed, accumulating evidence shows that epigenetic alterations are very common in cancer and exert pivotal roles in progression and metastasis. When transfection experiments in more than one cell line were available we found that the epigenetic role of miRNAs is strongly tissue dependent. The same miRNA can act as an epigenetic

switch in a particular tissue or cancer cell line and may have no effect at all in another cell line. An interesting feature of most of the known epi-miRNAs is that they are typically involved in complex double inhibitory feedback loops with their epigenetic targets. It has been shown in Osella et al. (Osella, Riba, et al. 2014) that this circuit is perfectly suited for tissue differentiation and more generally to choose among different cell fates and keep memory of this choice. The nature of the two partners of the feedback loop (miRNA on one side and epi-R on the other side) ensures an optimal resistance to stochastic fluctuations and avoids erroneous switching of the loop (Osella, Riba, et al. 2014). Several of the epi-miRNAs that we identified seem to follow this same pattern, but due to the nature of the data used in our analysis we cannot close the circuits for all the identified candidate epi-miRNA except for the ones presented in Figure 4.8, column sixth. Obviously, the circuits described are highly hypothetical, as it is not possible to confirm each interaction from the available data, but still they remain a good approximation of the kind of interactions we were looking for. We discuss here a few topical examples of these interactions, chosen among the miRNAs that we presented in the result section.

miR-31 is an important regulator of tumorigenesis. It can function either as oncomiR or as a tumor suppressive miRNA and it is typically deregulated in cancer cells. This miRNA has been recently shown to be a transcriptional target of EZH2 (Kurihara et al. 2016; Q. Zhang et al. 2014). Together with our finding this observation closes a perfect toggle switch (see Fig.4.7) which, as mentioned above, is a typical signature of epi-miRNAs. Remarkably enough, it has been recently shown that such a feedback loop is present with the other epigenetic pathways. Indeed it was shown that HDAC inhibitors enhance the expression level of miR-31 (Cho et al. 2015). Moreover, also the association between the HDAC and the PRC2 pathways that we observed in our data is confirmed in the literature, for instance it has been shown recently that in some classes of tumors the inhibition of miR-31 requires the joint action of EZH2 and HDAC (Koumangoye et al. 2015).

As for the previous miRNA, also miR-205 may act either as a tumor suppressor through inhibiting proliferation and invasion, or as an oncogene through facilating tumor initiation and proliferation, depending on the specific tumor context and target genes. It belongs to the miR-200 family and it is thus involved in the regulation of EMT. miR-205 was not previously reported as an epi-miRNA. Also in this case, there is evidence of a double inhibitory feedback loop with EZH2 (Q. Zhang et al. 2014). Differently from the previously described miRNA, miR-138 is a well-known epi-miRNA, it is involved in a double negative feedback loop both with SIRT1 (C.-M. Liu et al. 2013), which is a member of the HDAC family, and with EZH2 (H. Zhang et al. 2013; Zhu et al. 2016). These circuits are prototypical examples of epigenetic feedback loops (FBLs) and were discussed for instance in (Osella, Riba, et al. 2014). What is interesting is that in all these cases the gene sets are incoherent with respect to the expected epi-miRNA interaction. This means that these genes are simultaneously targets of miR-138 and EZH2.

The type of circuit that we found is depicted in Fig.4.7B. This circuit is very interesting as it is easy to see it ensures that the target genes are always silenced, independently from the orientation of the regulatory switch. What is more important, due to the incoherent nature of the circuit, their expression levels are not only low but are stable against fluctuations. These target genes are likely to be important regulator themselves, which in physiological conditions must be kept under control and are dysregulated during cancer progression. miR-145 is a known tumor suppressor which impacts on tumor cell growth and invasiveness, plays important roles in the differentiation of stem cells and vascular smooth muscle cells and regulates the expression of core stemness-associated factors, such as OCT4, SOX2, and KLF4 (Xu et al. 2009). It is interesting to notice that also for this miRNA there are strong indications suggesting the presence of a double inhibitory feedback loop with DNMT (Xue et al. 2015). This agrees with independent observations: for instance it was shown to be suppressed by hypermethylation in tumors (Suh et al. 2011) and it is down-regulated in lung adenocarcinoma tissues in cooperation with incremented DNA methylation (Xia et al. 2015). The DNMT pathway is indeed one of the pathways that we found as direct targets of miR-145 in our analysis (see Figure 4.8). Finally, a very interesting pattern is found in the case of miR-1 (Gruber and Zavolan 2013), for which GO categories show an enrichment in immune response and cell cycle inhibition categories and our findings suggest that miR-1 exerts these functions by targeting the PRC2 pathway.

Our proposed pipeline can be also used to identify putative epigenetic interactors of any given miRNA. The prometastatic miR-214 has been used as an example of the potentiality of our method. Based both on a dataset from our transfection experiments and from a publicly available dataset, we identified EZH2 as an epigenetic interactor of miR-214. We demonstrated that miR-214 regulates EZH2 in melanoma cells and EZH2 is in turn at least partially responsible of the downstream pathway activated by miR-214 involving ALCAM, ITGA5 and miR-148b. These results were in agreement with previous data regarding miR-214 regulation of EZH2 during muscle differentiation (Juan et al. 2009). Moreover, the GSEA analysis performed for the best intersection of miR-214 regulated genes with the selected MSigDB sets used in our analysis shows an enrichment in EMT and TGF-b related genes, perfectly in agreement with prometastatic role of miR-214 during tumor progression. The identification of miR-214 as an epi-miRNA in a tumor cell context can highlight new layers of regulation of this miRNA in the control of tumor progression.

Applying our pipeline to transfection experiments led to appreciable results, proving that transfection experiments can give good hints of the interactions underlying miRNAs and the epigenetic layer of regulation. It was one of our goals to be able to develop a method that could help spot epi-miRNAs in a simple way, being able both to use only one kind of data and at the same time to exploit the huge amount of data available online. These results, though, being based on data from transfection experiments, don't

necessarily report the natural effect that miRNAs have in the cells where they are endogenously present. miRNA overexpression experiments can generate a certain level of skepticism for their ability to produce false positive results through the artificial increase of miRNAs' concentration levels over the physiological standards, generally obtained with transient transfection (Thomson et al. 2011). Moreover, miRNA transfection experiments are usually executed in a cell environment that is not the natural context of the analyzed miRNA, and in these cases the native targets of the specific cell-type under object may be missed (Thomson et al. 2011).

Furthermore, miRNA are known to be involved in cancer dysregulation, and their differential regulation in cancer is often related or caused by malfunctions in important cancer-related pathways. To evaluate the relationship between miRNAs and epigenetic components in cancer tissue samples, we performed a differential expression and correlation analysis on TCGA data. We indeed spotted each of the 34 candidate epi-miRNAs as up- or downregulated in at least one of the analyzed TCGA projects. Some of these were always upregulated or always downregulated, and may therefore be miRNA that behave as oncomiR, in the first case, or as oncosoppressors, as for the second case, but this can also indicate a cell type specificity related to the complex mixture of cell types of the tumoral environment. Moreover, some miRNAs were widely dysregulated, indicating that they may be involved in pathways fundamental for the cancer cells and that are independent from the tissue of origin, while some other were differentially expressed in very few of the analyzed cancer types, suggesting a more cell-type or context specific function. Cell-type specificity may be evinced also from those cases in which the behaviour is more heterogeneous, involving different degrees and directions of differential expression in different tumor types.

Analyzing the correlations between each of the 34 miRNAs that resulted candidate epi-miRNAs and the epigenetic pathways in the TCGA projects, we found that many of the interactions spotted by our pipeline remained true. In some cases it is difficult to disentangle the kind of underlyning interaction, as for miR-183 and miR-96. In some other cases, the direction of the regulation and of the correlations gave us the possibility to hypothesize that there should be at least some negative feedback loop between the two components, as in the case of miR-133a and miR-31, confirming an interaction between the epigenetic components and the putative epi-miRNAs even when they are not artificially overexpressed as in the transfection experiments.

In conclusion, some of our candidate epi-miRNAs agree with already published findings, representing a positive test of our procedure, but some of them are new and represent one of our main results, and it would seem that their role can also be confirmed by in tissue analysis. Altogether our findings point to a much stronger role of epi-miRNAs in the regulatory network of higher eukaryotes and more generally to a strong interplay between the post-transcriptional and the epigenetic layers of regulation in shaping the differentiation process of complex tissues, that lead them to be important elements in cancer regulation too.

## 5.2 Future perspectives

The alteration of the epigenetic mechanisms can cause several serious pathologies, including cancers, cardiovascular, metabolic, neurodegenerative disorders as well as inherited syndromes. Our findings support the evidence collected during the years about the importance of miRNAs influence over epigenetic regulation. Given this importance and the complexity of their interplay, the epigenetic mechanisms can be only understood considering the regulatory circuitries composed of miRNAs and epi-Rs. Our work contributes to better elucidate and to identify new players involved in these complex regulatory circuits. Although the presented pipeline was applied on gene sets from tumor origin, it could be applied to different pathologies beside cancers.

Notably, our approach could be used in the future to go more in details in these regulations to verify how much these circuitries are similar among cells of the same tissue of origin or cells derived from the same kind of tumors as well as to unravel the pathological implications of new epi-miRNAs. Potentially, the level of resolution could also be increased in order to discriminate among the different members of the epi-R family involved in the circuits. The analysis on the TCGA data may profit a higher resolution too, stratifying the data on a more fine-grained basis and therefore performing the analysis on specific tumor subtypes, where possible.

The pipeline is designed to find interactions between miRNAs and epigenetic components based on the overlap between gene sets, evaluating the impact of miRNA on the epigenetic layer of regulation. The same pipeline could be used to analyze the effect of miRNAs on other cellular systems too, if it would be possible to have access to the right data. For example, it is known that miRNAs strongly interact with transcription factors (Osella, Bosia, et al. 2011), and would be therefore itneresting to assess if it is possible to investigate this interactions with our pipeline.

We have seen that the epi-miRNA-epi-R interactions seem to be strictly related to the cell type and conditions; also, in some cases, miRNAs result to regulate epigenetic components, but not the other way around. The strength of the project was based on being able to detect epi-miRNA in the simplest way and with the simplest-to-obtain data: the gene expression data. Still, one step forward could be made by analyzing data coming from ENCODE or other databases of ChipSeq experiments to search for the presence of epigenetic marks on miRNA genes.

With our work we have identified several new miRNAs involved in epigenetic regulation, highlighting the strong importance of the interplay between miRNAs and epi-Rs.

# Appendix A

# Materials and methods from the MBC wet experiments

## A.1 Reagents and antibodies

Pre-miR miRNA Precursor Molecules Negative Control 1, miRNA precursors hsa-miR-214 (PM12124; Ambion, TX, USA). TaqMan miRNA assays for miRNA detection: hsa-miR-148b (000471) and U6 snRNA (001973) (Applied Biosystems, CA, USA). siRNAs: siEZH2 (Hs-EZH2- FlexiTube siRNA SI02665166) and All Stars Negative Control siRNA were purchased from Qiagen (CA, USA). Primary antibodies: anti-actin pAb I-19 and anti-GAPDH pAb V-18 were from Santa Cruz Biotechnology (TX, USA); anti-ITGA5 pAb RM10 was kindly provided by G Tarone (University of Torino, Torino, Italy); anti-EZH2 mAb 612666 (BD, Transduction Laboratories, CA, USA); anti-CD166/ALCAM mAb MOG/07 (Novocastra Laboratories, Newcastle Upon Tyne, UK). Secondary antibodies: goat antimouse and goat antirabbit HRP-conjugated IgG were from Santa Cruz Biotechnology. All antibodies were used at the producer's suggested concentrations. Real-Time assays: QuantiTect Primer Assay (200) Hs-GAPDH-1-SG QuantiTect Primer Assay QT00079247, QuantiTect Primer Assay (200) Hs-ALCAM-1-SG QuantiTect Primer Assay QT00026824; QuantiTect Primer Assay (200) Hs-ITGA5-1-SG QuantiTect Primer Assay QT00080871 (Qiagen, CA, USA).

## A.2 Cell culture

MA-2 cells were provided by R.O. Hynes (Massachusetts Institute of Technology, MA, USA) and maintained as described previously (Penna, Orso, Cimino, Tenaglia, et al. 2011; Penna, Orso, Cimino, Vercellino, et al. 2013).

## A.3 Transient transfections of pre-miRs and siRNAs

To obtain transient pre-miR or siRNA expression, cells were transfected using RNAiFect (Qiagen, CA, USA) reagent, 75 nM pre-miR and 170 nM siRNA, as described in (Penna, Orso, Cimino, Tenaglia, et al. 2011).

## A.4 RNA isolation and quantitative Reverse Transcription (qRT)-PCR for miRNA or mRNA Detection

Total RNA was isolated from cells using TRIzol Reagent (Invitrogen Life Technologies, CA, USA) according to manufacturers' protocol. qRT-PCRs were performed as described in (Penna, Orso, Cimino, Tenaglia, et al. 2011).

## A.5 Gene expression profiling

Protein-coding expression profiling was carried out using the Whole Human Genome Oligo Microarray (41,000, 60-mer oligonucleotide probes) from Agilent Technologies. Total RNA (800 ng) was labeled by the Agilent One-Color Microarray-Based Gene Expression protocol, according to the manufacturer's instructions. Slides were scanned on an Agilent microarray scanner. Analysis was performed as described in (Cimino et al. 2013). Gene expression data are available in the U.S. Center for Biotechnology Information Gene Expression Omnibus (GEO) database (GSE124965).

## A.6 Protein preparation and western blotting

Total protein preparation and western blot analysis were performed as described in (Penna, Orso, Cimino, Tenaglia, et al. 2011).

## A.7 Statistical analyses for biological experiments

Data are presented as mean standard deviation (SD) or as mean standard error of the mean (SEM), as indicated, and two-tailed Student's t-test was used for comparison, with, $p < 0.05$; $p < 0.01$; $p < 0.001$ considered to be statistically significant. - indicates a non-statistically-significant p-value.

# Appendix B

# Links to pipeline packages and supplementary files

The pipeline followed for the epi-miRNA identification described in 4.3.1 and the dissertation's supplementary files are present at https://github.com/elisaReale/epi-miRNA. Pipeline for the TCGA analysis in paragraph 4.3.4 is still in a preliminary phase, and will be uploaded as soon as it is completed.

# Bibliography


Agarwal, Vikram et al. (Aug. 2015). "Predicting effective microRNA target sites in mammalian mRNAs". In: *eLife* 4. Ed. by Elisa Izaurralde, e05005. ISSN: 2050-084X.

Allis, C. David and Thomas Jenuwein (June 2016). "The molecular hallmarks of epigenetic control". In: *Nature Reviews Genetics* 17. Perspective, 487 EP -.

Alon, Uri (June 2007). "Network motifs: theory and experimental approaches". In: *Nature Reviews Genetics* 8. Review Article, 450 EP -.

Amodio, Nicola et al. (May 2015). "miR-29s: a family of epi-miRNAs with therapeutic implications in hematologic malignancies". In: *Oncotarget* 6.15. 3805[PII], pp. 12837–12861. ISSN: 1949-2553.

Arai, Takayuki et al. (May 2018). "Regulation of NCAPG by miR-99a-3p (passenger strand) inhibits cancer cell aggressiveness and is involved in CRPC". In: *Cancer medicine* 7.5. PMC5943442[pmcid], pp. 1988–2002. ISSN: 2045-7634.

Aravin, Alexei A. et al. (Sept. 2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice". In: *Molecular cell* 31.6. S1097-2765(08)00619-9[PII], pp. 785–799. ISSN: 1097-4164.

Ashburner, M. et al. (May 2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium". In: *Nat. Genet.* 25.1, pp. 25–29.

Baek, Daehyun et al. (July 2008). "The impact of microRNAs on protein output". In: *Nature* 455. Article, 64 EP -.

Bannister, A. J. and T. Kouzarides (Mar. 2011). "Regulation of chromatin by histone modifications". In: *Cell Res.* 21.3, pp. 381–395.

Bannister, Andrew J. and Tony Kouzarides (Mar. 2011). "Regulation of chromatin by histone modifications". In: *Cell research* 21.3. cr201122[PII], pp. 381–395. ISSN: 1748-7838.

Bao, Ning, Khar-Wai Lye, and M. Kathryn Barton (Nov. 2004). "MicroRNA Binding Sites in Arabidopsis Class III HD-ZIP mRNAs Are Required for Methylation of the Template Chromosome". In: *Developmental Cell* 7.5, pp. 653–662. ISSN: 1534-5807.

Barrett, Tanya et al. (Nov. 2012). "NCBI GEO: archive for functional genomics data sets—update". In: *Nucleic Acids Research* 41.D1, pp. D991–D995. ISSN: 0305-1048.

Bartel, David P. (2009). "MicroRNAs: Target Recognition and Regulatory Functions". In: *Cell* 136.2, pp. 215–233. ISSN: 0092-8674.

Behm-Ansmant, Isabelle et al. (2006). "mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes". In: *Genes  Development* 20.14, pp. 1885–1898.

Benetti, Roberta et al. (Mar. 2008). "A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases". In: *Nature Structural &Amp; Molecular Biology* 15. Article, 268 EP -.

Bhaskaran, M. and M. Mohan (Sept. 2013). "MicroRNAs: History, Biogenesis, and Their Evolving Role in Animal Development and Disease". In: *Veterinary Pathology* 51.4, pp. 759–774.

Bi, Laixi et al. (2018). "A novel miR-375-HOXB3-CDCA3/DNMT3B regulatory circuitry contributes to leukemogenesis in acute myeloid leukemia". In: *BMC Cancer* 18.1, p. 182. ISSN: 1471-2407.

Bianchi, Marzia, Katarzyna Augoff, et al. (Jan. 2012). "miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer". In: *Molecular cancer* 11. 1476-4598-11-5[PII], pp. 5–5. ISSN: 1476-4598.

Bianchi, Marzia, Alessandra Renzini, et al. (Apr. 2017). "Coordinated Actions of MicroRNAs with other Epigenetic Factors Regulate Skeletal Muscle Development and Adaptation". In: *International journal of molecular sciences* 18.4. PMC5412424[pmcid], p. 840. ISSN: 1422-0067.

Blanc, Roméo S. and Stéphane Richard (2017). "Arginine Methylation: The Coming of Age". In: *Molecular Cell* 65.1, pp. 8–24. ISSN: 1097-2765.

Bommi, Prashant V. et al. (2010). "The polycomb group protein BMI1 is a transcriptional target of HDAC inhibitors". In: *Cell Cycle* 9.13. PMID: 20543557, pp. 2663–2673.

Bracken, Cameron P. et al. (2008). "A Double-Negative Feedback Loop between ZEB1-SIP1 and the microRNA-200 Family Regulates Epithelial-Mesenchymal Transition". In: *Cancer Research* 68.19, pp. 7846–7854. ISSN: 0008-5472.

Braconi, Chiara, Nianyuan Huang, and Tushar Patel (2010). "MicroRNA-dependent regulation of DNA methyltransferase-1 and tumor suppressor gene expression by interleukin-6 in human malignant cholangiocytes". In: *Hepatology* 51.3, pp. 881–890.

Brockdorff, N. et al. (May 1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome". In: *Nature* 351.6324, pp. 329–331.

Brockers, Kevin and Robert Schneider (2019). "Histone H1, the forgotten histone". In: *Epigenomics* 11.4. PMID: 30793938, pp. 363–366.

Buchwald, Gretel et al. (2006). "Structure and E3-ligase activity of the Ring–Ring complex of Polycomb proteins Bmi1 and Ring1b". In: *The EMBO Journal* 25.11, pp. 2465–2474.

Burk, Ulrike et al. (2008). "A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells". In: *EMBO reports* 9.6, pp. 582–589.

Calin, George A. et al. (Apr. 2008). "MiR-15a and miR-16-1 cluster functions in human leukemia". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.13. 0800121105[PII], pp. 5166–5171. ISSN: 1091-6490.

Calin, George Adrian et al. (Nov. 2002). "Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.24. 242606799[PII], pp. 15524–15529. ISSN: 0027-8424.

Cao, Jian and Qin Yan (Mar. 2012). "Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer". In: *Frontiers in oncology* 2. PMC3355875[pmcid], pp. 26–26. ISSN: 2234-943X.

Castel, Stephane E. and Robert A. Martienssen (Jan. 2013). "RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond". In: *Nature Reviews Genetics* 14. Review Article, 100 EP -.

Catalanotto, Caterina, Carlo Cogoni, and Giuseppe Zardo (Oct. 2016). "MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions". In: *International journal of molecular sciences* 17.10. PMC5085744[pmcid], p. 1712. ISSN: 1422-0067.

Chakravarthy, Srinivas et al. (2005). "Structural Characterization of the Histone Variant macroH2A". In: *Molecular and Cellular Biology* 25.17, pp. 7616–7624. ISSN: 0270-7306.

Chavali, Vishalakshi, Suresh C. Tyagi, and Paras K. Mishra (Aug. 2012). "MicroRNA-133a regulates DNA methylation in diabetic cardiomyocytes". In: *Biochemical and biophysical research communications* 425.3. S0006-291X(12)01412-X[PII], pp. 668–672. ISSN: 1090-2104.

Cho, Joon-Ho, Manjari Dimri, and Goberdhan P. Dimri (Apr. 2015). "MicroRNA-31 is a transcriptional target of histone deacetylase inhibitors and a regulator of cellular senescence". In: *The Journal of biological chemistry* 290.16. M114.624361[PII], pp. 10555–10567. ISSN: 1083-351X.

Chou, Chih-Hung et al. (Nov. 2017). "miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions". In: *Nucleic Acids Research* 46.1, pp. 296–302. ISSN: 0305-1048.

Chow, Jennifer C. et al. (2010). "LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation". In: *Cell* 141.6, pp. 956–969. ISSN: 0092-8674.

Cimino, Daniela et al. (2013). "miR148b is a major coordinator of breast cancer progression in a relapse-associated microRNA signature by targeting ITGA5, ROCK1, PIK3CA, NRAS, and CSF1". In: *The FASEB Journal* 27.3. PMID: 23233531, pp. 1223–1235.

Cimmino, Amelia et al. (Sept. 2005). "miR-15 and miR-16 induce apoptosis by targeting BCL2". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.39. 0506654102[PII], pp. 13944–13949. ISSN: 0027-8424.

*Connectivity Map (CMAP)* (July 2018). URL: `https://www.broadinstitute.org/connectivity-map-cmap`.

Consortium, The Gene Ontology (Jan. 2019). "The Gene Ontology Resource: 20 years and still GOing strong". In: *Nucleic Acids Res.* 47.D1, pp. D330–D338.

Cooper, D. N., M. H. Taggart, and A. P. Bird (Feb. 1983). "Unmethylated domains in vertebrate DNA". In: *Nucleic Acids Res.* 11.3, pp. 647–658.

Coppola, Antonietta et al. (2014). "Cardiomyogenesis is controlled by the miR-99a/let-7c cluster and epigenetic modifications". In: *Stem Cell Research* 12.2, pp. 323–337. ISSN: 1873-5061.

Dakhlallah, Duaa et al. (Feb. 2013). "Epigenetic regulation of miR-17~92 contributes to the pathogenesis of pulmonary fibrosis". In: *American journal of respiratory and critical care medicine* 187.4. rccm.201205-0888OC[PII], pp. 397–405. ISSN: 1535-4970.

Dambal, Shweta et al. (July 2015). "The microRNA-183 cluster: the family that plays together stays together". In: *Nucleic Acids Research* 43.15, pp. 7173–7188. ISSN: 0305-1048.

Datta, Jharna et al. (July 2008). "Methylation mediated silencing of MicroRNA-1 gene and its role in hepatocellular carcinogenesis". In: *Cancer research* 68.13. 68/13/5049[PII], pp. 5049–5058. ISSN: 1538-7445.

Denli, Ahmet M. et al. (2004). "Processing of primary microRNAs by the Microprocessor complex". In: *Nature* 432.7014, pp. 231–235. ISSN: 1476-4687.

Dews, Michael et al. (2006). "Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster". In: *Nature Genetics* 38.9, pp. 1060–1065. ISSN: 1546-1718.

Di Lorenzo, Alessandra and Mark T. Bedford (2011). "Histone arginine methylation". In: *FEBS Letters* 585.13, pp. 2024–2031.

Dogini, Danyella et al. (Mar. 2014). "The new world of RNAs". In: *Genetics and molecular biology* 37, pp. 285–293.

Durinck, Steffen et al. (2009). "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt". In: *Nature Protocols* 4.8, pp. 1184–1191. ISSN: 1750-2799.

Duursma, Anja M. et al. (May 2008). "miR-148 targets human DNMT3b protein coding region". In: *RNA (New York, N.Y.)* 14.5. rna.972008[PII], pp. 872–877. ISSN: 1469-9001.

Edgar, Ron, Michael Domrachev, and Alex E. Lash (Jan. 2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository". In: *Nucleic acids research* 30.1. PMC99122[pmcid], pp. 207–210. ISSN: 1362-4962.

Egger, Gerda et al. (2004). "Epigenetics in human disease and prospects for epigenetic therapy". In: *Nature* 429.6990, pp. 457–463. ISSN: 1476-4687.

Enokida, Hideki (2012). "Novel molecular targets regulated by tumor suppressors microRNA-1 and microRNA-133a in bladder cancer". In: *International Journal of Oncology*.

Fabbri, Muller et al. (Oct. 2007). "MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B". In: *Proceedings of the National Academy of Sciences of the United*

*States of America* 104.40. 0707628104[PII], pp. 15805–15810. ISSN: 0027-8424.

Fabian, Marc R. and Nahum Sonenberg (June 2012). "The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC". In: *Nature Structural &Amp; Molecular Biology* 19. Review Article, 586 EP -.

Fasanaro, Pasquale et al. (June 2008). "MicroRNA-210 modulates endothelial cell response to hypoxia and inhibits the receptor tyrosine kinase ligand Ephrin-A3". In: *The Journal of biological chemistry* 283.23. M800731200[PII], pp. 15878–15883. ISSN: 0021-9258.

*Four Common Histone Modifications* (n.d.). `https://www.cusabio.com/c-20829.html`, last accessed on 2019-08-10. URL: `https://www.cusabio.com/c-20829.html`.

Frixa, Tania, Sara Donzelli, and Giovanni Blandino (Dec. 2015). "Oncogenic MicroRNAs: Key Players in Malignant Transformation". In: *Cancers* 7.4. cancers7040904[PII], pp. 2466–2485. ISSN: 2072-6694.

Fuse, Miki et al. (2012). "Tumor suppressive microRNAs (miR-222 and miR-31) regulate molecular pathways based on microRNA expression signature in prostate cancer". In: *Journal of Human Genetics* 57.11, pp. 691–699. ISSN: 1435-232X.

Gao, J. et al. (2013). "MiR-26a inhibits proliferation and migration of breast cancer through repression of MCL-1". In: *PLoS ONE* 8.6, e65138.

Gao, L. et al. (July 2002). "Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family". In: *J. Biol. Chem.* 277.28, pp. 25748–25755.

Gardner, Timothy S., Charles R. Cantor, and James J. Collins (2000). "Construction of a genetic toggle switch in Escherichia coli". In: *Nature* 403.6767, pp. 339–342. ISSN: 1476-4687.

Garzon, Ramiro et al. (2009). "MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1". In: *Blood* 113.25, pp. 6411–6418. ISSN: 0006-4971.

*GDC Data User Guide* (n.d.). URL: `https://docs.gdc.cancer.gov/Data/PDF/`.

Gebert, Luca F. R. and Ian J. MacRae (2019). "Regulation of microRNA function in?animals". In: *Nature Reviews Molecular Cell Biology* 20.1, pp. 21–37. ISSN: 1471-0080.

Geiman, Theresa M. et al. (2004). "DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system". In: *Biochemical and Biophysical Research Communications* 318.2, pp. 544–555. ISSN: 0006-291X.

*GEO Overview - GEO - NCBI* (n.d.). URL: `https://www.ncbi.nlm.nih.gov/geo/info/overview.html`.

Ghildiyal, Megha et al. (Jan. 2010). "Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway".

In: *RNA (New York, N.Y.)* 16.1. rna.1972910[PII], pp. 43–56. ISSN: 1469-9001.

Glass, Christopher K. and Michael G. Rosenfeld (2000). "The coregulator exchange in transcriptional functions of nuclear receptors". In: *Genes Development* 14.2, pp. 121–141.

Goll, M. G. et al. (Jan. 2006). "Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2". In: *Science* 311.5759, pp. 395–398.

Goto, Yusuke, Satoko Kojima, Akira Kurozumi, et al. (May 2016). "Regulation of E3 ubiquitin ligase-1 (WWP1) by microRNA-452 inhibits cancer cell migration and invasion in prostate cancer". In: *British journal of cancer* 114.10. bjc201695[PII], pp. 1135–1144. ISSN: 1532-1827.

Goto, Yusuke, Satoko Kojima, Rika Nishikawa, et al. (Sept. 2015). "MicroRNA expression signature of castration-resistant prostate cancer: the microRNA-221/222 cluster functions as a tumour suppressor and disease progression marker". In: *British journal of cancer* 113.7. bjc2015300[PII], pp. 1055–1065. ISSN: 1532-1827.

Gruber, Andreas J and Mihaela Zavolan (2013). "Modulation of epigenetic regulators and cell fate decisions by miRNAs". In: *Epigenomics* 5.6. PMID: 24283881, pp. 671–683.

Gu, Bingnan and Min Gyu Lee (2013). "Histone H3 lysine 4 methyltransferases and demethylases in self-renewal anddifferentiation of stem cells". In: *Cell & Bioscience* 3.1, p. 39. ISSN: 2045-3701.

Ha, Minju and V. Narry Kim (July 2014). "Regulation of microRNA biogenesis". In: *Nature Reviews Molecular Cell Biology* 15. Review Article, 509 EP -.

Han, Jinju et al. (Dec. 2004). "The Drosha-DGCR8 complex in primary microRNA processing". In: *Genes & development* 18.24. gad.1262504[PII], pp. 3016–3027. ISSN: 0890-9369.

He, Lin et al. (2005). "A microRNA polycistron as a potential human oncogene". In: *Nature* 435.7043, pp. 828–833. ISSN: 1476-4687.

Henikoff, S. and M. M. Smith (Jan. 2015). "Histone variants and epigenetics". In: *Cold Spring Harb Perspect Biol* 7.1, a019364.

Hidaka, Hiedo et al. (Jan. 2012). "Tumor suppressive microRNA-1285 regulates novel molecular targets: aberrant expression and functional significance in renal cell carcinoma". In: *Oncotarget* 3.1. 417[PII], pp. 44–57. ISSN: 1949-2553.

Holliday, R. (Jan. 1990). "DNA methylation and epigenetic inheritance". In: *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 326.1235, pp. 329–338.

Holliday, R and JE Pugh (1975). "DNA modification mechanisms and gene activity during development". In: *Science* 187.4173, pp. 226–232. ISSN: 0036-8075.

Holoch, Daniel and Danesh Moazed (Feb. 2015). "RNA-mediated epigenetic regulation of gene expression". In: *Nature reviews. Genetics* 16.2. nrg3863[PII], pp. 71–84. ISSN: 1471-0064.

Hotchkiss, R. D. (Aug. 1948). "The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography". In: *J. Biol. Chem.* 175.1, pp. 315–332.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki (Nov. 2008). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists". In: *Nucleic Acids Research* 37.1, pp. 1–13. ISSN: 0305-1048.

— (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". In: *Nature Protocols* 4.1, pp. 44–57. ISSN: 1750-2799.

Huntzinger, Eric and Elisa Izaurralde (Jan. 2011). "Gene silencing by microRNAs: contributions of translational repression and mRNA decay". In: *Nature Reviews Genetics* 12. Review Article, 99 EP -.

Hyun, Kwangbeom et al. (Apr. 2017). "Writing, erasing and reading histone lysine methylations". In: *Experimental &Amp; Molecular Medicine* 49. Review, e324 EP -.

Illingworth, Robert S. and Adrian P. Bird (2009). "CpG islands – 'A rough guide'". In: *FEBS Letters* 583.11, pp. 1713–1720.

Iorio, Marilena V., Claudia Piovan, and Carlo M. Croce (2010). "Interplay between microRNAs and the epigenetic machinery: An intricate network". In: *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1799.10. Small molecule modulators of chromatin and co activators - A link between Epigenetics and Gene expression, pp. 694–701. ISSN: 1874-9399.

Ishibashi, T. et al. (Apr. 2010). "H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis". In: *Nucleic Acids Res.* 38.6, pp. 1780–1789.

Jones, Peter L. et al. (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription". In: *Nature Genetics* 19.2, pp. 187–191. ISSN: 1546-1718.

Joosten, Sophie C. et al. (2018). "Epigenetics in renal cell cancer: mechanisms and clinical applications". In: *Nature Reviews Urology* 15.7, pp. 430–451. ISSN: 1759-4820.

Juan, Aster H. et al. (Oct. 2009). "Mir-214-Dependent Regulation of the Polycomb Protein Ezh2 in Skeletal Muscle and Embryonic Stem Cells". In: *Molecular Cell* 36.1, pp. 61–74. ISSN: 1097-2765.

Karlic, Rosa et al. (Feb. 2010). "Histone modification levels are predictive for gene expression". In: *Proceedings of the National Academy of Sciences* 107.7, p. 2926.

Kikkawa, Naoko et al. (2014). "microRNA-504 inhibits cancer cell proliferation via targeting CDK6 in hypopharyngeal squamous cell carcinoma". In: *International Journal of Oncology* 44.6, pp. 2085–2092.

Kinoshita, T. et al. (Nov. 2013). "Tumour-suppressive microRNA-29s inhibit cancer cell migration and invasion by targeting laminin-integrin signalling in head and neck squamous cell carcinoma". In: *British journal of cancer* 109.10. bjc2013607[PII], pp. 2636–2645. ISSN: 1532-1827.

Kinoshita, Takashi et al. (Nov. 2012). "Tumor suppressive microRNA-218 inhibits cancer cell migration and invasion through targeting laminin-332 in head and neck squamous cell carcinoma". In: *Oncotarget* 3.11. 709[PII], pp. 1386–1400. ISSN: 1949-2553.

Kong, William et al. (Nov. 2008). "MicroRNA-155 is regulated by the transforming growth factor beta/Smad pathway and contributes to epithelial cell plasticity by targeting RhoA". In: *Molecular and cellular biology* 28.22. MCB.00941-08[PII], pp. 6773–6784. ISSN: 1098-5549.

Koumangoye, Rainelli B. et al. (2015). "SOX4 interacts with EZH2 and HDAC3 to suppress microRNA-31 in invasive esophageal cancer cells". In: *Molecular Cancer* 14.1, p. 24. ISSN: 1476-4598.

Kumamoto, Tomohiro et al. (Nov. 2016). "Regulation of TPD52 by antitumor microRNA-218 suppresses cancer cell migration and invasion in lung squamous cell carcinoma". In: *International journal of oncology* 49.5. PMC5063422[pmcid], pp. 1870–1880. ISSN: 1791-2423.

Kunej, Tanja et al. (2011). "Epigenetic regulation of microRNAs in cancer: An integrated review of literature". In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 717.1. MicroRNA in Environmental Mutagenesis, pp. 77–84. ISSN: 0027-5107.

Kuramochi-Miyagawa, Satomi et al. (Apr. 2008). "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes". In: *Genes & development* 22.7. 22/7/908[PII], pp. 908–917. ISSN: 0890-9369.

Kurdistani, Siavash K, Saeed Tavazoie, and Michael Grunstein (2004). "Mapping Global Histone Acetylation Patterns to Gene Expression". In: *Cell* 117.6, pp. 721–733. ISSN: 0092-8674.

Kurihara, Hiroyoshi et al. (2016). "The relationship between EZH2 expression and microRNA-31 in colorectal cancer and the role in evolution of the serrated pathway". In: *Oncotarget* 7.11, pp. 12704–12717. ISSN: 1949-2553.

Laisné, Marthe et al. (Dec. 2018). "Mechanisms of DNA Methyltransferase Recruitment in Mammals". In: *Genes* 9.12. PMC6316769[pmcid], p. 617. ISSN: 2073-4425.

Lee, Rosalind C., Rhonda L. Feinbaum, and Victor Ambros (Dec. 1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14". In: *Cell* 75.5, pp. 843–854.

Lee, Sun-Min et al. (2017). "Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes". In: *Proceedings of the National Academy of Sciences* 114.10, E1885–E1894. ISSN: 0027-8424.

Levine, Erel et al. (Aug. 2007). "Quantitative Characteristics of Gene Regulation by Small RNA". In: *PLOS Biology* 5.9, pp. 1–13.

Li, Ning, Heng Zhou, and Qizhu Tang (2018). "miR-133: A Suppressor of Cardiac Remodeling?" In: *Frontiers in Pharmacology* 9, p. 903. ISSN: 1663-9812.

Liberzon, Arthur et al. (2015). "The Molecular Signatures Database Hallmark Gene Set Collection". In: *Cell Systems* 1.6, pp. 417–425. ISSN: 2405-4712.

Liu, Chang-Mei et al. (July 2013). "MicroRNA-138 and SIRT1 form a mutual negative feedback loop to regulate mammalian axon regeneration". In: *Genes & development* 27.13. gad.209619.112[PII], pp. 1473–1483. ISSN: 1549-5477.

Liu, Nian and Tao Pan (Jan. 2015). "RNA epigenetics". In: *Translational research : the journal of laboratory and clinical medicine* 165.1. S1931-5244(14)00127-3[PII], pp. 28–35. ISSN: 1878-1810.

Liu, Qing and Ming Wei Wang (July 2016). "Histone lysine methyltransferases as anti-cancer targets for drug discovery". In: *Acta Pharmacologica Sinica* 37. Review, 1273 EP -.

Love, Michael I., Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12, p. 550. ISSN: 1474-760X.

Lu, Jun et al. (2005). "MicroRNA expression profiles classify human cancers". In: *Nature* 435.7043, pp. 834–838. ISSN: 1476-4687.

Lu, Mingyang et al. (2013). "MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination". In: *Proceedings of the National Academy of Sciences* 110.45, pp. 18144–18149. ISSN: 0027-8424.

Luger, Karolin et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution". In: *Nature* 389.6648, pp. 251–260. ISSN: 1476-4687.

Mahmoud, Fade et al. (June 2016). "Role of EZH2 histone methyltrasferase in melanoma progression and metastasis". In: *Cancer biology & therapy* 17.6. PMC4990393[pmcid], pp. 579–591. ISSN: 1555-8576.

Malik, Harmit S. and Steven Henikoff (2003). "Phylogenomics of the nucleosome". In: *Nature Structural & Molecular Biology* 10.11, pp. 882–891. ISSN: 1545-9985.

Mariño-Ramírez, Leonardo et al. (Oct. 2005). "Histone structure and nucleosome stability". In: *Expert review of proteomics* 2.5. PMC1831843[pmcid], pp. 719–729. ISSN: 1744-8387.

Martin, Cyrus and Yi Zhang (2005). "The diverse functions of histone lysine methylation". In: *Nature Reviews Molecular Cell Biology* 6.11, pp. 838–849. ISSN: 1471-0080.

Matsushita, R. et al. (July 2015). "Tumour-suppressive microRNA-144-5p directly targets CCNE1/2 as potential prognostic markers in bladder cancer". In: *British journal of cancer* 113.2. bjc2015195[PII], pp. 282–289. ISSN: 1532-1827.

Matsushita, Ryosuke et al. (May 2016). "Regulation of UHRF1 by dual-strand tumor-suppressor microRNA-145 (miR-145-5p and miR-145-3p): Inhibition of bladder cancer cell aggressiveness". In: *Oncotarget* 7.19. 8668[PII], pp. 28460–28487. ISSN: 1949-2553.

Mattick, John S. et al. (2009). "RNA regulation of epigenetic processes". In: *BioEssays* 31.1, pp. 51–59.

McGinty, Robert K. and Song Tan (2014). "Histone, Nucleosome, and Chromatin Structure". In: *Fundamentals of Chromatin*. Ed. by Jerry L. Workman and Susan M. Abmayr. New York, NY: Springer New York, pp. 1–28. ISBN: 978-1-4614-8624-4.

Meijer, Hedda A., Ewan M. Smith, and Martin Bushell (2014). "Regulation of miRNA strand selection: follow the leader?" In: *Biochemical Society Transactions* 42.4, pp. 1135–1140. ISSN: 0300-5127.

Memari, Fatemeh et al. (2018). "Epigenetics and Epi-miRNAs: Potential markers/therapeutics in leukemia". In: *Biomedicine Pharmacotherapy* 106, pp. 1668–1677. ISSN: 0753-3322.

Metzger, Eric et al. (Jan. 2008). "Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation". In: *Nature cell biology* 10.1. ncb1668[PII], pp. 53–60. ISSN: 1476-4679.

Mi, H. et al. (Jan. 2019). "PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools". In: *Nucleic Acids Res.* 47.D1, pp. D419–D426.

Mierlo, Guido van et al. (2019). "The Complexity of PRC2 Subcomplexes". In: *Trends in Cell Biology* 29.8, pp. 660–671. ISSN: 0962-8924.

Misiewicz-Krzeminska, Irena et al. (Apr. 2013). "Restoration of microRNA-214 expression reduces growth of myeloma cells through positive regulation of P53 and inhibition of DNA replication". In: *Haematologica* 98.4. haematol.2012.070011[PII], pp. 640–648. ISSN: 1592-8721.

Moriya, Yasumitsu et al. (2012). "Tumor suppressive microRNA-133a regulates novel molecular networks in lung squamous cell carcinoma". In: *Journal of Human Genetics* 57.1, pp. 38–45. ISSN: 1435-232X.

Morrison, Ashby J. and Xuetong Shen (2005). "DNA Repair in the Context of Chromatin". In: *Cell Cycle* 4.4. PMID: 15753660, pp. 513–512.

Noh, Ji Heon et al. (2013). "MiR-145 functions as a tumor suppressor by directly targeting histone deacetylase 2 in liver cancer". In: *Cancer Letters* 335.2, pp. 455–462. ISSN: 0304-3835.

Nohata, N., T. Hanazawa, N. Kikkawa, et al. (Sept. 2011). "Tumour suppressive microRNA-874 regulates novel cancer networks in maxillary sinus squamous cell carcinoma". In: *British journal of cancer* 105.6. bjc2011311[PII], pp. 833–841. ISSN: 1532-1827.

Nohata, N., T. Hanazawa, T. Kinoshita, et al. (Apr. 2013). "Tumour-suppressive microRNA-874 contributes to cell proliferation through targeting of histone deacetylase 1 in head and neck squamous cell carcinoma". In: *British journal of cancer* 108.8. bjc2013122[PII], pp. 1648–1658. ISSN: 1532-1827.

Nohata, Nijiro, Toyoyuki Hanazawa, et al. (2011). "Tumor suppressive microRNA-375 regulates oncogene AEG-1/MTDH in head and neck squamous cell carcinoma (HNSCC)". In: *Journal of Human Genetics* 56.8, pp. 595–601. ISSN: 1435-232X.

Nohata, Nijiro, Yaeko Sone, et al. (2011). "miR-1 as a tumor suppressive microRNA targeting TAGLN2 in head and neck squamous cell carcinoma". In: *Oncotarget* 2.1-2. 213[PII], pp. 29–42. ISSN: 1949-2553.

Noonan, E. J. et al. (Mar. 2009). "miR-449a targets HDAC-1 and induces growth arrest in prostate cancer". In: *Oncogene* 28. Original Article, 1714 EP -.

Nuytten, M. et al. (Sept. 2007). "The transcriptional repressor NIPP1 is an essential player in EZH2-mediated gene silencing". In: *Oncogene* 27. Original Article, 1449 EP -.

O'Brien, Jacob et al. (Aug. 2018). "Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation". In: *Frontiers in endocrinology* 9. PMC6085463[pmcid], pp. 402–402. ISSN: 1664-2392.

O'Donnell, Kathryn A. et al. (2005). "c-Myc-regulated microRNAs modulate E2F1 expression". In: *Nature* 435.7043, pp. 839–843. ISSN: 1476-4687.

Osella, Matteo, Carla Bosia, et al. (Mar. 2011). "The Role of Incoherent MicroRNA-Mediated Feedforward Loops in Noise Buffering". In: *PLOS Computational Biology* 7.3, pp. 1–16.

Osella, Matteo, Andrea Riba, et al. (2014). "Interplay of microRNA and epigenetic regulation in the human regulatory network". In: *Frontiers in Genetics* 5, p. 345. ISSN: 1664-8021.

Palazzo, Alexander F. and Eliza S. Lee (2015). "Non-coding RNA: what is functional and what is junk?" In: *Frontiers in Genetics* 6, p. 2. ISSN: 1664-8021.

Park, June Hyun and Chanseok Shin (Aug. 2014). "MicroRNA-directed cleavage of targets: mechanism and experimental approaches". In: *BMB reports* 47.8. 2837[PII], pp. 417–423. ISSN: 1976-670X.

Patel, Prajal H., Scott A. Barbee, and J. Todd Blankenship (Mar. 2016). "GW-Bodies and P-Bodies Constitute Two Separate Pools of Sequestered Non-Translating RNAs". In: *PloS one* 11.3. PONE-D-15-53487[PII], e0150291–e0150291. ISSN: 1932-6203.

Peng, Yong and Carlo M. Croce (2016). "The role of MicroRNAs in human cancer". In: *Signal Transduction and Targeted Therapy* 1.1, p. 15004. ISSN: 2059-3635.

Penna, Elisa, Francesca Orso, Daniela Cimino, Enrico Tenaglia, et al. (2011). "microRNA-214 contributes to melanoma tumour progression through suppression of TFAP2C". In: *The EMBO Journal* 30.10, pp. 1990–2007.

Penna, Elisa, Francesca Orso, Daniela Cimino, Irene Vercellino, et al. (2013). "miR-214 Coordinates Melanoma Progression by Upregulating ALCAM through TFAP2 and miR-148b Downmodulation". In: *Cancer Research* 73.13, pp. 4098–4111. ISSN: 0008-5472.

Peserico, Alessia and Cristiano Simone (2011). "Physical and functional HAT/HDAC interplay regulates protein acetylation balance". In: *Journal of biomedicine & biotechnology* 2011. PMC2997516[pmcid], pp. 371832–371832. ISSN: 1110-7251.

Pichiorri, Flavia et al. (Oct. 2010). "Downregulation of p53-inducible microRNAs 192, 194, and 215 impairs the p53/MDM2 autoregulatory loop in multiple myeloma development". In: *Cancer cell* 18.4. S1535-6108(10)00342-9[PII], pp. 367–381. ISSN: 1878-3686.

Pichler, M. and G. A. Calin (July 2015). "MicroRNAs in cancer: from developmental genes in worms to their clinical application in patients". In: *British Journal Of Cancer* 113. Minireview, 569 EP -.

Poon, Vivian Y. et al. (2016). "miR-27b shapes the presynaptic transcriptome and influences neurotransmission by silencing the polycomb group protein Bmi1". In: *BMC Genomics* 17.1, p. 777. ISSN: 1471-2164.

Portela, Anna and Manel Esteller (Oct. 2010). "Epigenetic modifications and human disease". In: *Nature Biotechnology* 28, 1057 EP -.

Ramassone, Alice et al. (Feb. 2018). "Epigenetics and MicroRNAs in Cancer". In: *International journal of molecular sciences* 19.2. PMC5855681[pmcid], p. 459. ISSN: 1422-0067.

Rastgoo, Nasrin et al. (Dec. 2016). "Restoration of Mir-138 Expression By Epigenetic Modulation Overcomes Drug Resistance in Multiple Myeloma Cells". In: *Blood* 128.22, pp. 2060–2060. ISSN: 0006-4971.

Reinhart, B. J. et al. (Feb. 2000). "The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans". In: *Nature* 403.6772, pp. 901–906.

*RNA interference* (n.d.). https://www.abnova.com/support/pathway.asp?switchfunctionid={D045D93C-F533-4E46-AE4A-CE11C41028A2}, last accessed on 2019-08-22. URL: https://www.abnova.com/support/pathway.asp?switchfunctionid=%7BD045D93C-F533-4E46-AE4A-CE11C41028A2%7D.

*Roadmap Epigenomics Project* (n.d.). http://www.roadmapepigenomics.org/overview, last accessed on 2019-07-03. URL: http://www.roadmapepigenomics.org/overview.

Robertson, Keith D. (2005). "DNA methylation and human disease". In: *Nature Reviews Genetics* 6.8, pp. 597–610. ISSN: 1471-0064.

Rocha, Simao Teixeira da et al. (Jan. 2014). "Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome". In: *Molecular Cell* 53.2, pp. 301–316. ISSN: 1097-2765.

Ropero, Santiago and Manel Esteller (2007). "The role of histone deacetylases (HDACs) in human cancer". In: *Molecular Oncology* 1.1, pp. 19–25. ISSN: 1574-7891.

Ryu, B., S. Hwang, and R. M. Alani (May 2013). "MicroRNAs as an emerging target for melanoma therapy". In: *J. Invest. Dermatol.* 133.5, pp. 1137–1139.

Sato, Fumiaki et al. (2011). "MicroRNAs and epigenetics". In: *The FEBS Journal* 278.10, pp. 1598–1609.

Schmitter, Daniela et al. (Sept. 2006). "Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells". In: *Nucleic Acids Research* 34.17, pp. 4801–4815. ISSN: 0305-1048.

Schoeftner, Stefan et al. (July 2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing". In: *The EMBO journal* 25.13. 7601187[PII], pp. 3110–3122. ISSN: 0261-4189.

Scott, Gary K. et al. (2006). "Rapid Alteration of MicroRNA Levels by Histone Deacetylase Inhibition". In: *Cancer Research* 66.3, pp. 1277–1281. ISSN: 0008-5472.

Seki, Naohiko (2011a). "Identification of novel molecular targets regulated by tumor suppressive miR-1/miR-133a in maxillary sinus squamous cell carcinoma". In: *International Journal of Oncology*.

— (Dec. 2011b). "Tumor suppressive microRNA-375 regulates lactate dehydrogenase B in maxillary sinus squamous cell carcinoma". In: *International Journal of Oncology*.

Shi, Yujiang et al. (2004). "Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1". In: *Cell* 119.7, pp. 941–953. ISSN: 0092-8674.

Smith, Christopher D. et al. (2007). "The Release 5.1 Annotation of Drosophila melanogaster Heterochromatin". In: *Science* 316.5831, pp. 1586–1591. ISSN: 0036-8075.

Song, Q. C. et al. (Mar. 2014). "Downregulation of microRNA-26a is associated with metastatic potential and the poor prognosis of osteosarcoma patients". In: *Oncol. Rep.* 31.3, pp. 1263–1270.

Storz, Gisela (2002). "An Expanding Universe of Noncoding RNAs." In: *Science* 296.5571, pp. 1260–1263. ISSN: 0036-8075.

Strahl, B. D. and C. D. Allis (Jan. 2000). "The language of covalent histone modifications". In: *Nature* 403.6765, pp. 41–45.

Su, Hong et al. (July 2011). "Mammalian hyperplastic discs homolog EDD regulates miRNA-mediated gene silencing". In: *Molecular cell* 43.1. S1097-2765(11)00454-0[PII], pp. 97–109. ISSN: 1097-4164.

Subramanian, Aravind et al. (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43, pp. 15545–15550. ISSN: 0027-8424.

Suh, Seong O. et al. (May 2011). "MicroRNA-145 is regulated by DNA methylation and p53 gene mutation in prostate cancer". In: *Carcinogenesis* 32.5. bgr036[PII], pp. 772–778. ISSN: 1460-2180.

Suzuki, Hiromu et al. (Dec. 2013). "Epigenetic alteration and microRNA dysregulation in cancer". In: *Frontiers in genetics* 4. PMC3847369[pmcid], pp. 258–258. ISSN: 1664-8021.

Talbert, Paul B. and Steven Henikoff (2010). "Histone variants – ancient wrap artists of the epigenome". In: *Nature Reviews Molecular Cell Biology* 11.4, pp. 264–275. ISSN: 1471-0080.

Tamagawa, Shunji et al. (2014). "Role of miR-200c/miR-141 in the regulation of epithelial-mesenchymal transition and migration in head and neck squamous cell carcinoma". In: *International Journal of Molecular Medicine* 33.4, pp. 879–886.

Tamaru, Hisashi (July 2010). "Confining euchromatin/heterochromatin territory: jumonji crosses the line". In: *Genes & development* 24.14. 24/14/1465[PII], pp. 1465–1478. ISSN: 1549-5477.

*The Cancer Genome Atlas Program* (n.d.). URL: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

Thomson, Daniel W., Cameron P. Bracken, and Gregory J. Goodall (June 2011). "Experimental strategies for microRNA target identification". In: *Nucleic Acids Research* 39.16, pp. 6845–6853. ISSN: 0305-1048.

Tian, Tianhai and Kevin Burrage (2006). "Stochastic models for regulatory networks of the genetic toggle switch". In: *Proceedings of the National Academy of Sciences* 103.22, pp. 8372–8377. ISSN: 0027-8424.

Trojer, Patrick and Danny Reinberg (2007). "Facultative Heterochromatin: Is There a Distinctive Molecular Signature?" In: *Molecular Cell* 28.1, pp. 1–13. ISSN: 1097-2765.

Tuddenham, Lee et al. (2006). "The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells". In: *FEBS Letters* 580.17, pp. 4214–4217.

Varambally, Sooryanarayana et al. (2008). "Genomic Loss of microRNA-101 Leads to Overexpression of Histone Methyltransferase EZH2 in Cancer". In: *Science* 322.5908, pp. 1695–1699. ISSN: 0036-8075.

Vasilatou, Diamantina et al. (2013). "Epigenetic alterations and microRNAs". In: *Epigenetics* 8.6. PMID: 23760524, pp. 561–570.

Vidal, Miguel and Katarzina Starowicz (2017). "Polycomb complexes PRC1 and their function in hematopoiesis". In: *Experimental Hematology* 48, pp. 12–31. ISSN: 0301-472X.

Vidigal, Joana A. and Andrea Ventura (2015). "The biological functions of miRNAs: lessons from in vivo studies". In: *Trends in Cell Biology* 25.3, pp. 137–147. ISSN: 0962-8924.

Waddington, C. H. (Dec. 2011). "The Epigenotype". In: *International Journal of Epidemiology* 41.1, pp. 10–13. ISSN: 0300-5771.

Wang, Hengbin et al. (2004). "Role of histone H2A ubiquitination in Polycomb silencing". In: *Nature* 431.7010, pp. 873–878. ISSN: 1476-4687.

Wang, Sumei, Wanyin Wu, and Francois X. Claret (Mar. 2017). "Mutual regulation of microRNAs and DNA methylation in human cancers". In: *Epigenetics* 12.3. PMC5406215[pmcid], pp. 187–197. ISSN: 1559-2308.

Wang, Xiangting et al. (July 2008). "Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription". English. In: *Nature* 454.7200. Copyright - Copyright Nature Publishing Group Jul 3, 2008; Caratteristica del documento - Graphs; Diagrams; ; Ultimo aggiornamento - 2019-09-06; CODEN - NATUAS, pp. 126–30.

Wang, Zifeng et al. (2013). "Transcriptional and epigenetic regulation of human microRNAs". In: *Cancer Letters* 331.1, pp. 1–10. ISSN: 0304-3835.

Weinhold, Bob (Mar. 2006). "Epigenetics: the science of change". In: *Environmental health perspectives* 114.3. PMC1392256[pmcid], A160–A167. ISSN: 0091-6765.

*What is chromatin, heterochromatin and euchromatin?* (May 2018). https://www.mechanobio.info/genome-regulation/what-is-chromatin-

heterochromatin-and-euchromatin/, last accessed on 2019-08-19. URL: https://www.mechanobio.info/genome-regulation/what-is-chromatin-heterochromatin-and-euchromatin/.

Xia, Wenjie et al. (Nov. 2015). "DNA methylation mediated silencing of microRNA-145 is a potential prognostic marker in patients with lung adenocarcinoma". In: *Scientific reports* 5. srep16901[PII], pp. 16901–16901. ISSN: 2045-2322.

Xie, Mingyi et al. (Dec. 2013). "Mammalian 5'-capped microRNA precursors that generate a single microRNA". In: *Cell* 155.7. S0092-8674(13)01478-5[PII], pp. 1568–1580. ISSN: 1097-4172.

Xu, Na et al. (May 2009). "MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells". In: *Cell* 137.4, pp. 647–658. ISSN: 0092-8674.

Xue, Gang et al. (2015). "A feedback regulation between miR-145 and DNA methyltransferase 3b in prostate cancer cell and their responses to irradiation". In: *Cancer Letters* 361.1, pp. 121–127. ISSN: 0304-3835.

Yamada, Yasutaka, Takayuki Arai, Satoko Kojima, et al. (June 2018a). "Antitumor roles of both strands of the miR-455 duplex: their targets SKA1 and SKA3 are involved in the pathogenesis of renal cell carcinoma". In: *Oncotarget* 9.42. 25410[PII], pp. 26638–26658. ISSN: 1949-2553.

— (Sept. 2018b). "Regulation of antitumor miR-144-5p targets oncogenes: Direct regulation of syndecan-3 and its clinical significance". In: *Cancer science* 109.9. PMC6125479[pmcid], pp. 2919–2936. ISSN: 1349-7006.

Yamada, Yasutaka, Takayuki Arai, Sho Sugawara, et al. (Apr. 2018). "Impact of novel oncogenic pathways regulated by antitumor miR-451a in renal cell carcinoma". In: *Cancer science* 109.4. PMC5891191[pmcid], pp. 1239–1253. ISSN: 1349-7006.

Yamada, Yasutaka, Keiichi Koshizuka, et al. (Jan. 2018). "Passenger strand of miR-145-3p acts as a tumor-suppressor by targeting MYO1B in head and neck squamous cell carcinoma". In: *International journal of oncology* 52.1. PMC5743364[pmcid], pp. 166–178. ISSN: 1791-2423.

Yamada, Yasutaka, Sho Sugawara, et al. (2018). "Molecular pathogenesis of renal cell carcinoma: Impact of the anti-tumor miR-29 family on gene regulation". In: *International Journal of Urology* 25.11, pp. 953–965.

Yamasaki, Takeshi et al. (Sept. 2012). "Tumor suppressive microRNA-138 contributes to cell migration and invasion through its targeting of vimentin in renal cell carcinoma". In: *International journal of oncology* 41.3. PMC3582944[pmcid], pp. 805–817. ISSN: 1791-2423.

Yang, Jr-Shiuan et al. (Aug. 2010). "Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.34. 1006432107[PII], pp. 15163–15168. ISSN: 1091-6490.

Yao, Qian, Yuqi Chen, and Xiang Zhou (2019). "The roles of microRNAs in epigenetic regulation". In: *Current Opinion in Chemical Biology* 51. Chemical Genetics and Epigenetics • Molecular Imaging, pp. 11–17. ISSN: 1367-5931.

Yonemori, Keiichi et al. (Jan. 2017). "ZFP36L2 promotes cancer cell aggressiveness and is regulated by antitumor microRNA-375 in pancreatic ductal adenocarcinoma". In: *Cancer science* 108.1. PMC5276842[pmcid], pp. 124–135. ISSN: 1349-7006.

Yoshino, H. et al. (Mar. 2011). "The tumour-suppressive function of miR-1 and miR-133a targeting TAGLN2 in bladder cancer". In: *British journal of cancer* 104.5. bjc201123[PII], pp. 808–818. ISSN: 1532-1827.

Yu, Li et al. (Apr. 2013). "miR-26a inhibits invasion and metastasis of nasopharyngeal cancer by targeting EZH2". In: *Oncology letters* 5.4. ol-05-04-1223[PII], pp. 1223–1228. ISSN: 1792-1074.

Yu, Tao et al. (2018). "Functions and mechanisms of microRNA-31 in human cancers". In: *Biomedicine Pharmacotherapy* 108, pp. 1162–1169. ISSN: 0753-3322.

Zeinali, Tahereh et al. (2019). "Regulatory mechanisms of miR-145 expression and the importance of its function in cancer metastasis". In: *Biomedicine Pharmacotherapy* 109, pp. 195–207. ISSN: 0753-3322.

Zhang, Huijun et al. (Jan. 2013). "MiR-138 Inhibits Tumor Growth Through Repression of EZH2 in Non-Small Cell Lung Cancer". In: *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 31, pp. 56–65.

Zhang, J. et al. (2019). "Regulation of histone arginine methylation/demethylation by methylase and demethylase." In: *Molecular Medicine Reports* 19, pp. 3963–3971. ISSN: 1791-2997.

Zhang, Q. et al. (Oct. 2014). "Polycomb protein EZH2 suppresses apoptosis by silencing the proapoptotic miR-31". In: *Cell death & disease* 5.10. cddis2014454[PII], e1486–e1486. ISSN: 2041-4889.

Zhang, Yijun et al. (Dec. 2014). "Cellular microRNAs up-regulate transcription via interaction with promoter TATA-box motifs". In: *RNA (New York, N.Y.)* 20.12. rna.045633.114[PII], pp. 1878–1889. ISSN: 1469-9001.

Zhang, Zhiqian et al. (Aug. 2011). "Epigenetic Silencing of miR-203 Upregulates SNAI2 and Contributes to the Invasiveness of Malignant Breast Cancer Cells". In: *Genes & cancer* 2.8. 10.1177_1947601911429743[PII], pp. 782–791. ISSN: 1947-6027.

Zhou, H. et al. (June 2014). "MicroRNA-26a acts as a tumor suppressor inhibiting gallbladder cancer cell proliferation by directly targeting HMGA2". In: *Int. J. Oncol.* 44.6, pp. 2050–2058.

Zhu, Ziqiang et al. (Mar. 2016). "MiR-138 Acts as a Tumor Suppressor by Targeting EZH2 and Enhances Cisplatin-Induced Apoptosis in Osteosarcoma Cells". In: *PloS one* 11.3. PONE-D-15-36768[PII], e0150026–e0150026. ISSN: 1932-6203.

Zhuang, C. et al. (Mar. 2016). "A double-negative feedback loop between EZH2 and miR-26a regulates tumor cell growth in hepatocellular carcinoma". In: *Int. J. Oncol.* 48.3, pp. 1195–1204.