



**HAL**  
open science

# Régulation réciproque entre les microARN et leurs cibles

Sophie Mockly

► **To cite this version:**

Sophie Mockly. Régulation réciproque entre les microARN et leurs cibles. Médecine humaine et pathologie. Université Montpellier, 2021. Français. NNT : 2021MONTT051 . tel-03558262

**HAL Id: tel-03558262**

**<https://theses.hal.science/tel-03558262>**

Submitted on 4 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Génétique et Biologie Moléculaire

École doctorale - CBS2  
(Sciences Chimiques et Biologiques pour la Santé)

Unité de recherche - Institut de Génétique Humaine  
(UMR 9002 CNRS- Université de Montpellier)

## Régulation réciproque entre les microARN et leurs cibles

Présentée par **Sophie MOCKLY**  
Le 08 décembre 2021

Sous la direction du Dr. Hervé SEITZ

Devant le jury composé de

Dr. Alena SHKUMATAVA, Institut Curie

Dr. Helge GROSSHANS, Friedrich Miescher Institute for Biomedical Research

Dr. Christine GASPIN, MIAT INRAE

Dr. Jérôme CAVAILLE, CBI CNRS-Université Toulouse III - Paul Sabatier

Dr. Séverine CHAMBEYRON, IGH CNRS-Université de Montpellier

Dr. Hervé SEITZ, IGH CNRS-Université de Montpellier

Rapporteuse

Rapporteur

Examinatrice

Examineur

Présidente

Directeur de thèse



UNIVERSITÉ  
DE MONTPELLIER



UNIVERSITY OF MONTPELLIER

DOCTORAL THESIS

---

# Reciprocal regulation between microRNAs and their targets

---

*Author:*  
Sophie MOCKLY

*Supervisor:*  
Dr. Hervé SEITZ

A thesis submitted in fulfillment of the requirements for the degree of  
*Ph.D. in Genetics and Molecular Biology*  
in the  
*Institute of Human Genetics, UMR 9002 – CNRS/University of Montpellier*

*Reviewers:*  
Dr. Alena SHKUMATAVA  
Pr. Helge GROSSHANS

*Examiners:*  
Dr. Séverine CHAMBEYRON  
Dr. Christine GASPIN  
Dr. Jérôme CAVAILLÉ

December 8, 2021



*“J’ai une bonne et une mauvaise nouvelle...”*

Hervé Seitz

## Acknowledgements

First and foremost, I would like to express my sincere gratitude to the reviewers of my Ph.D. Committee, Dr. Alena Shkumatava and Pr. Helge Großhans, for agreeing to take the time to review my thesis and for their insightful comments and suggestions, as well as to the examiners, Dr. Séverine Chambeyron, Dr. Christine Gaspin, and Dr. Jérôme Cavaillé, for accepting the invitation to participate in my thesis defense. I am also deeply grateful to the members of my annual Comité de thèse, Dr. Céline Gongora, Dr. Sébastien Pfeffer and Dr. Sylvain Egloff for their feedbacks and recommendations throughout my thesis.

Je suis extrêmement reconnaissante à mon superviseur, Hervé Seitz, de m'avoir accueillie dans son équipe et d'avoir su être un véritable mentor. Son soutien continu, ses précieux conseils, sa générosité intellectuelle, ses immenses connaissances et sa bonne humeur ont été autant de points qui m'ont permis de réaliser une thèse dans des conditions remarquables. Je tiens également à remercier l'ensemble des membres actuels et anciens de l'équipe, Isabelle Busseau, Elisabeth Houbron et Germain Busto, pour leur soutien scientifique mais surtout humain. Leur écoute et leur jovialité ont fait de ces quatre années un moment merveilleux. J'ai eu une chance infinie de travailler avec vous tous et je ne louerai jamais assez la qualité scientifique et humaine de l'équipe.

Mes remerciements vont aussi à l'ensemble du personnel de l'Institut de Génétique Humaine que j'ai pu côtoyer, avec un merci particulier aux équipes Chambeyron et Mochizuki pour leur écoute scientifique et leurs conseils éclairés lors de nos Joint Lab Meetings qui ont joué un rôle essentiel dans la construction de mon projet de thèse.

Un merci supplémentaire aux rencontres amicales que j'ai pu faire à l'IGH, et la liste est longue. En commençant par les anciens, je remercie Jessica, Ana Luiza, Kasia et Kamar pour leur bienveillance de Post-Docs mais aussi leur expérience et leur honnêteté, Lenka, Eric et Alexandre pour tous ces rires et ces discussions enflammées et enfin Germain qui, à défaut de n'être pas "fifou", aura apporté un surplus d'allégresse dans notre équipe. Poursuivons avec les irréductibles, je remercie les équipes Dejardin et De Luco qui m'ont toujours acceptée pendant les pauses-déjeuners, et qui sont devenues comme mes équipes subsidiaires. Merci à Amandine pour son soutien, surtout sur cette fin de thèse, à Yaiza pour sa bienveillance de grande soeur, et enfin à Andrew qui, tout en concurrençant Germain dans la catégorie "Chercheur absolument insupportable", ne m'a enfermé que trois fois dans mon bureau et fait peur que 468 fois pendant mes manip.

Je tiens à remercier chaleureusement tous les professeurs qui m'ont amenée à aimer les sciences, en général, et la biologie tout particulièrement. Mes professeurs du collège et du lycée Jean-Moulin à Saint-Amand-Montrond, mais aussi de classe préparatoire BCPST au lycée Blaise-Pascal à Clermont-Ferrand, d'Agrocampus Ouest et de l'Université de Rennes. Un merci tout particulier à Mme Ranty, ma professeure de SVT en Première et Terminale, qui a su me conseiller dans une période d'errance universitaire et qui m'a, sans le savoir, poussée dans

les bras de la recherche, ainsi qu'au Pr. Sandrine Lagarrigue pour m'avoir chaperonnée lors de mes premiers pas en génétique et dans la recherche académique.

Je souhaite aussi remercier tous mes amis qui m'ont accompagnée ces dernières années, voire même depuis le début. Merci à mes amis d'enfance, Chloé, Quentin, Tom, Bastien et Mathieu d'être toujours présents et de me soutenir à distance. Merci à mes camarades de Prépa, Anaïs, Vanille et tous les autres Bio 139 qui restent dans mon coeur, pour ces années riches en émotion que nous avons vécues ensemble et qui m'ont appris la résilience et l'effort. Merci à mes deux compères de l'Agro, Dr. Julien et Dr. Fred, si lourds et si drôles, qui après nos belles années à Rennes m'ont accompagnée dans l'aventure de la recherche académique et m'ont soutenue en cette fin de thèse, malgré la distance. Merci aussi à mes fillots/fillotes de l'Agro, en particulier Marianne et Soledad pour les rires et les goûters improvisés. Enfin, je souhaite remercier les membres du club de Course d'Orientation de Montpellier, l'AMSO34, pour tous ces moments sportifs et conviviaux que nous avons passés ensemble.

Enfin, je tiens à exprimer ma gratitude à ma famille et plus particulièrement à mes parents. Sans leur compréhension et leurs encouragements de ma première à ma dernière année dans le Supérieur, je n'aurais peut-être pas embrassé de si belles et passionnantes études. Un grand merci à mon papa pour m'avoir inculqué, de par son modèle, les valeurs du travail et du mérite, et à ma maman d'avoir toujours eu foi, plus que moi-même, en mes capacités et pour avoir toujours répondu à l'appel, et surtout à mes appels, dans les moments difficiles.

Pour terminer, je tiens à remercier Cyril qui m'a accompagnée et soutenue pendant ces quatre années de dur labeur, mais surtout d'amour, de rire et de joie.

# Contents

<b>Preface</b>	<b>1</b>
<b>1 History and Molecular Overview of microRNAs</b>	<b>3</b>
1.1 History of microRNA Discovery . . . . .	3
From heterochronic gene identification to stRNA definition . . . . .	3
stRNAs through the prism of RNAi . . . . .	4
A novel class of small RNAs defined by molecular features . . . . .	5
microRNAs across species . . . . .	6
1.2 microRNA Biogenesis in Metazoan . . . . .	7
1.2.1 Genomic Distribution and Annotation of microRNA Genes . . . . .	7
1.2.2 Transcription and Nuclear Processing of Canonical microRNAs . . . . .	8
Pri-miRNA transcription . . . . .	8
Maturation of pri-miRNA into pre-miRNA by the Microprocessor . . . . .	9
1.2.3 Cytoplasmic Maturation of Canonical microRNAs . . . . .	10
Nuclear export of pre-miRNA . . . . .	10
Processing of pre-miRNA into miRNA duplex by DICER . . . . .	10
1.2.4 microRNA-Inducing Silencing Complex Formation . . . . .	11
1.2.5 Non-Canonical microRNA Biogenesis Pathways . . . . .	12
Microprocessor-independent mirtrons . . . . .	12
Microprocessor and XPO5-independent 5'-capped miRNAs . . . . .	13
DICER-independent miR-451 . . . . .	14
TUT-dependant miRNAs . . . . .	14
1.3 Molecular Mechanisms of microRNA-Mediated Gene Silencing . . . . .	16
1.3.1 Principles of microRNA Targeting . . . . .	16
The notion of miRNA seed . . . . .	16
Canonical and non-canonical miRNA target sites . . . . .	17
MRE localization in transcriptional units . . . . .	18
1.3.2 microRNA-Mediated Gene Silencing . . . . .	18
The interplay between components of the miRISC silencing machinery . . . . .	19
RNAi-like target slicing . . . . .	21
<b>2 Dynamics and Physiological Relevance of microRNAs</b>	<b>23</b>
2.1 Physiological Importance of microRNA Abundance and Function in Metazoans . . . . .	23
2.1.1 Global Effect of microRNAs on Development and Homeostasis . . . . .	23
2.1.2 Physiological Contribution of Individual microRNAs . . . . .	24



2.1.3	Abnormal Regulation and Function of microRNAs in Cancer . . . . .	26
2.2	Regulation Mechanisms of microRNA Abundance . . . . .	27
2.2.1	Dependance Between microRNA Levels and Function . . . . .	27
2.2.2	Dynamics of microRNAs . . . . .	27
2.2.3	Control of Global microRNA Biogenesis . . . . .	28
	Regulatory loop of maturation factors . . . . .	28
	Post-transcriptional modifications of maturation factors . . . . .	28
2.2.4	Regulation of Individual microRNA Biogenesis . . . . .	29
	Transcriptional regulation . . . . .	29
	Precursor miRNA sequence modifications . . . . .	29
	Influence of RNA-binding proteins . . . . .	31
2.2.5	Control of microRNA Stability . . . . .	32
	3'-tailing and trimming impact on miRNA stability . . . . .	32
	Target-directed miRNA degradation . . . . .	33
2.3	Physiological Relevance of microRNA Targets . . . . .	37
2.3.1	Classes of microRNA-Guided Regulatory Effects . . . . .	37
	Developmental genetic switch . . . . .	37
	Fine-tuning and noise buffering . . . . .	37
	Neutral interactions . . . . .	38
2.3.2	The "Many" and "Few Targets" Hypotheses . . . . .	39
	"Many Targets" . . . . .	39
	"Few Targets" . . . . .	40
<b>3</b>	<b>Functional Relevance of microRNAs on Cellular Phenotypes</b>	<b>43</b>
3.1	Functional microRNA Targets Identification Methods . . . . .	43
3.1.1	Review: <i>Inconsistencies and limitations of current miRNA target identification methods</i> . . . . .	43
3.2	Identification of microRNA:Target Interactions Involved in Cell Proliferation . . . . .	64
	Rationale of the strategy . . . . .	64
	Methodology description . . . . .	64
	Quantification of the global proliferative effect of miR-34a . . . . .	65
	Generation of miR-21-deficient HAP1 cells . . . . .	68
3.3	Revisiting the Anti-Proliferative Function of miR-34a . . . . .	70
3.3.1	The Master Tumor-Suppressive Role of miR-34a in the Literature . . . . .	70
	Description of the miR-34 family . . . . .	70
	miR-34a, a candidate tumor-suppressor miRNA . . . . .	71
	miR-34a replenishment therapeutics and their limits . . . . .	72
	Aim of the project . . . . .	74
3.3.2	Paper: <i>A rationalized definition of tumor suppressor microRNAs excludes miR-34a</i> . . . . .	74
3.4	Discussion . . . . .	99

<b>4 Identification of Target-Directed microRNA Degradation Inducers</b>	<b>103</b>
4.1 Development of a TDMD Inducer Prediction Computational Approach . . . . .	104
Preliminary step: Establishment of TDMD-inducing pairing rules . . . . .	104
Step 1: Computational TDMD-inducing pattern search in vertebrates . . . . .	105
Step 2: TDMD-inducing site conservation score in vertebrates . . . . .	109
Step 3: TDMD-inducing candidate selection in a tissue of interest . . . . .	110
4.2 TDMD Inducer Candidates in Murine Neuronal Cells . . . . .	111
4.3 Experimental Validation of AGO2-bound miR-708 Cleavage Activity . . . . .	113
4.4 Experimental Validation of TDMD Inducer Candidates . . . . .	115
Strategy 1: RNAi with pre-designed shRNA constructs . . . . .	115
Strategy 2: RNAi with optimized shRNA constructs . . . . .	117
Strategy 3: dCas-KRAB-mediated gene silencing . . . . .	119
4.5 Analysis of Degradation Intermediates from small RNA-seq Data . . . . .	121
XomeTox samples preparation . . . . .	121
Development of a miRNA isoform quantification approach . . . . .	122
Inconsistencies in rat miRNA annotation . . . . .	123
4.6 Discussion and Perspectives . . . . .	124
<b>A Vector Maps</b>	<b>127</b>
<b>B Patent WO2021185766</b>	<b>129</b>
<b>Bibliography</b>	<b>143</b>
<b>Abstract</b>	<b>167</b>



# List of Abbreviations

<b>3'-UTR</b>	3'-Untranslated Region
<b>5'-UTR</b>	5'-Untranslated Region
<b>ceRNA</b>	competing endogenous RNA
<b>circRNA</b>	circular RNA
<b>DNA</b>	Deoxyribonucleic Acid
<b>dsRNA</b>	double-stranded RNA
<b>FACS</b>	Fluorescence-Activated Cell Sorting
<b>lincRNA</b>	long intervening non-coding RNA
<b>MEF</b>	Mouse Embryonic Fibroblasts
<b>miRNA</b>	microRNA
<b>MRE</b>	MicroRNA Response Element
<b>mRNA</b>	messenger RNA
<b>ncRNA</b>	non-coding RNA
<b>nt</b>	nucleotide
<b>ORF</b>	Open Reading Frame
<b>pri-microRNA</b>	primary precursor for microRNA
<b>pre-microRNA</b>	precursor for microRNA
<b>RBP</b>	RNA-Binding Protein
<b>RNA</b>	Ribonucleic Acid
<b>RNAi</b>	RNA interference
<b>siRNA</b>	small-interfering RNA
<b>stRNA</b>	small-temporal RNA
<b>TDMD</b>	Target-Directed MicroRNA Degradation
<b>TF</b>	Transcription Factor
<b>TSS</b>	Transcription Start Site
<b>TU</b>	Transduction Unit



# List of Figures

1.1	<i>lin-4</i> and <i>let-7</i> transcripts and complementary sites within <i>lin-14</i> and <i>lin-41</i> 3'-UTRs . . . . .	4
1.2	The miR-25 family nomenclature . . . . .	7
1.3	Examples of genomic location of miRNA loci and sequence of pri-miRNA . . . . .	8
1.4	Primary and secondary structures of canonical pri-miRNAs . . . . .	10
1.5	Structures of human AGO2 associated with a guide RNA . . . . .	12
1.6	Primary and secondary structures of mirtron pri-miRNAs . . . . .	13
1.7	Biogenesis of canonical and non-canonical miRNAs . . . . .	15
1.8	Canonical sites of mammalian miRNAs. . . . .	17
1.9	Atypical and non-canonical sites of mammalian miRNAs. . . . .	18
1.10	Model of miRNA-mediated gene silencing . . . . .	21
2.1	The repertoire of <i>dme-bantam</i> isomiRs . . . . .	30
2.2	The ZSIM8/Cullin-RING E3 ubiquitin ligase model of TDMD and its interplay with TDTT . . . . .	36
3.1	Overview of an unbiased CRISPR screen for the identification of miRNA targets involved in cell proliferation control . . . . .	65
3.2	Proliferative rate of wild-type and miR-34a-deficient clones in HCT-116 and HAP1 cells . . . . .	67
3.3	Proliferative rate of wild-type and miR-34a-deficient clones in HCT-116 from Navarro et al., 2015 . . . . .	67
3.4	titre . . . . .	68
3.5	Schematic of the single miR-21-mutant clone successfully isolated . . . . .	69
3.6	Structure and conservation of the miR-34 family in vertebrates. . . . .	70
4.1	Patterns of TDMD-inducing pairing geometry . . . . .	104
4.2	Input files pre-processing . . . . .	106
4.3	Rationale behind miRNA seed match search in sense and antisense strands . . . . .	107
4.4	Schematic of 3' mismatches and central bulge complementarity identification workflow . . . . .	108
4.5	TDMD inducer candidates in mouse cortex . . . . .	112
4.6	Multiple alignment of TDMD inducer candidates in vertebrates . . . . .	112
4.7	AGO2-cleaved Nnat transcripts are not detected by RLM-5'-RACE in murine primary cortical neurons . . . . .	114

4.8	Lentiviral delivery of pGIPZ shRNAs against TDMD inducer candidates resulted in poorly efficient mRNA knock-down . . . . .	116
4.9	Small RNA-Seq coverage profile of pGIPZ shRNAs . . . . .	117
4.10	Structural accessibility of target gene sequence . . . . .	118
4.11	Optimized synthetic pri-miRNA backbone from Fang et al., 2015. . . . .	118
4.12	Quantification of TDMD inducer candidates knockdown with optimized shRNA constructs . . . . .	119
4.13	Quantification of TDMD inducer candidates knockdown with dCas9-KRAB/ZIM3120	
4.14	Example of the mapping strategy to record canonical and non-canonical isomiRs.	122
4.15	Distribution of native miRNAs sequence percentage in Xometox libraries. . . . .	123
A.1	Vector map of pGIPZ lentiviral vector from Dharmacon . . . . .	127
A.2	Vector maps of plasmids for dCas9-KRAB/ZIM3-mediated gene silencing. . . . .	128

# Preface

This thesis was carried out in the team "Systemic impact of small regulatory RNAs" under the supervision of Dr. Hervé Seitz. During these four years, I studied a subfamily of small regulatory RNAs called microRNAs, discovered 25 years ago and considered as essential regulators of a wide range of cellular processes by modulating the expression of numerous genes.

My thesis project was developed around two axes of study, each corresponding to a regulatory direction between miRNAs and their target genes. In order to present the product of my work, this thesis is subdivided into four parts.

- The first part presents a general state of the art on miRNAs by first introducing the history of their discovery and then the molecular characteristics of these small RNAs: from their biogenesis to their mode of gene silencing.
- The second part focuses on the physiological function of miRNAs. For this purpose, the modes of control of the intracellular abundance of miRNAs, from which their functionality derives, are introduced but also the current theories concerning the role of the different types of miRNA targets.
- The third part relates to miRNA-mediated gene silencing which translates into a macroscopic phenotype, and the identification of such functional gene targets of miRNAs. A review on the different techniques of identifying miRNA targets introduces this part and is followed by the results of my work on the miR-34a miRNA, a supposedly anti-proliferative miRNA. The product of my work has resulted in the writing of a paper that is currently under revision.
- Finally, the fourth part of this thesis focuses on the reciprocal regulation of miRNAs by their targets. My project consisted in developing a bioinformatics tool to predict the targets inducing such miRNA degradation, and my efforts to experimentally validate our *in silico* results are presented.

I also had the opportunity to participate in collaborations during my thesis, and some of my work is presented in the text and the appendices.





## Chapter 1

# History and Molecular Overview of microRNAs

### 1.1 History of microRNA Discovery

#### From heterochronic gene identification to stRNA definition

The first described miRNAs result from dissecting the genetic control of developmental timing in *Caenorhabditis elegans*. This model organism exhibits some ideal traits for the study of development: a transparent body, a fixed number of somatic cells characteristic to eutely, an invariant cell lineage, and four larval stages defined by molts, which altogether facilitate the detection and reliable scoring of mutant phenotypes with altered post-embryonic development. Thereby, mutagenesis screening in *C. elegans* resulting in temporal defects in cell fates enabled the identification of heterochronic genes (Ambros et al., 1984; Sulston et al., 1981), *i.e.* genes controlling the proper timing and sequence of developmental stages.

Such heterochronic mutations in *C. elegans* initially defined four genes: *lin-4*, *lin-14*, *lin-28*, and *lin-29*, in reference to “cell lineage abnormal” mutants (Ambros et al., 1984). Later, more than twenty genes were described to constitute the heterochronic gene pathway in the larval development of *C. elegans* with stage-specific expression patterns and genetic interactions among them (Moss et al., 2014). Following the discovery of the first heterochronic mutants, the laboratories of Ambros and Ruvkun deciphered the molecular mechanism behind the genetic interrelationship between *lin-4* and *lin-14*: decrease of *lin-14* activity is required for proper progression from the first larval stage to later stages and is *lin-4*-dependent (Ambros, 1989; Ambros et al., 1987). They first demonstrated that this down-regulation was mediated by negative regulatory elements located in the 3′-UTR of *lin-14* mRNA, likely interacting with *lin-4* product (Wightman et al., 1991). However, they found that *lin-4* does not encode a protein, instead its product is a pair of small transcripts: a short one of approximately 22 nt and a long one of 61 nt, containing the shorter sequence and predicted to fold in a hairpin structure (Lee et al., 1993). The longer transcript was suggested to act as a precursor of the shorter one, and this hypothesis was later confirmed, as detailed in section 1.2. Interestingly, the *lin-4* product sequence was partially complementary to regulatory sites located in the 3′-UTR of *lin-14* mRNA and conserved in the *Caenorhabditis* genus (Wightman et al., 1993), suggesting that

the base pairing between *lin-4* small RNAs and *lin14* mRNA inhibits LIN-14 production.

A few years later, the Ruvkun laboratory described a novel heterochronic gene, *let-7* for “lethal-7” mutant, sharing similarities with *lin-4*. The product of the *let-7* gene was a 21 nt transcript complementary to elements in the *lin-41* 3'-UTR which mediate the post-transcriptional down-regulation of LIN-41 during the later larval stages (Reinhart et al., 2000; Slack et al., 2000). However, while *lin-4* is conserved in nematodes only, the *let-7* 21-nt RNA sequence and developmental control are conserved in bilaterians, and *let-7* homologs exhibit a similar stem-loop secondary structure. Furthermore, *let-7* complementary sites in *lin-41* 3'-UTR are also conserved in *D. melanogaster* and *Danio rerio* (Pasquinelli et al., 2000). These latter results led to the conclusion that *let-7* and *lin-4* 21-nt long RNAs form a new class of small RNAs able to regulate temporal development in *C. elegans* and likely in all bilaterian animals, and they were dubbed “small temporal RNAs” abbreviated stRNAs.

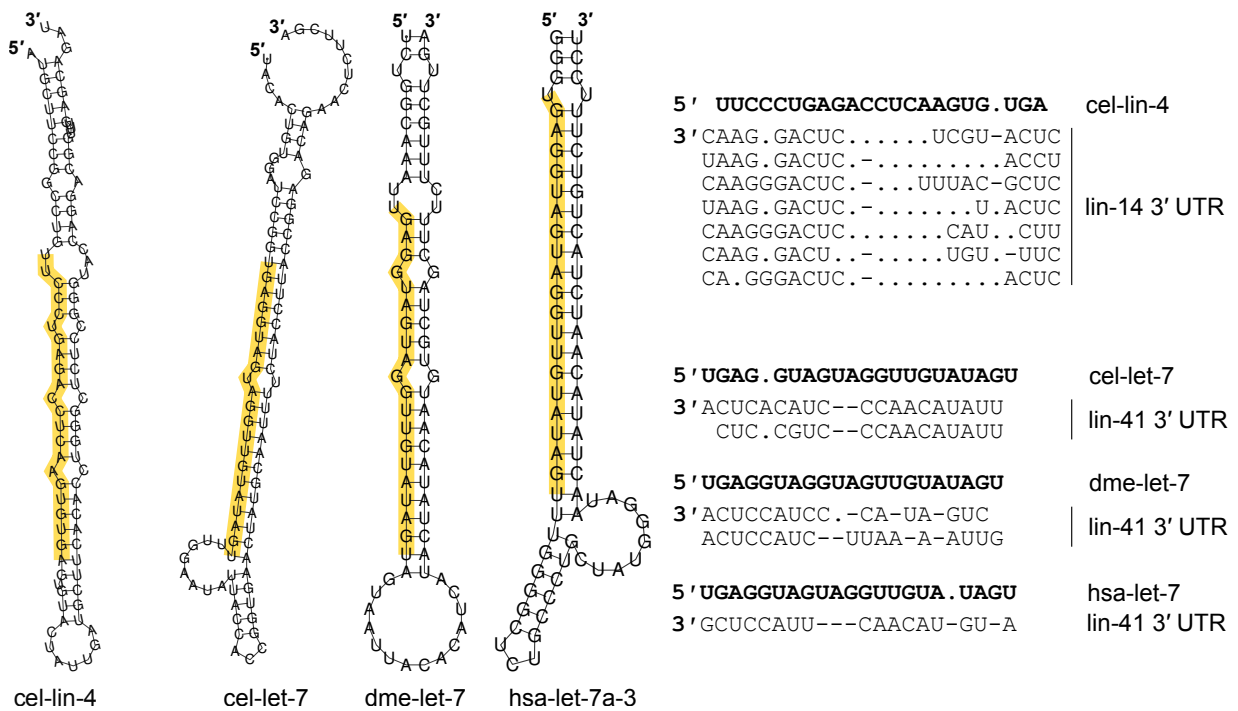


FIGURE 1.1: *lin-4* and *let-7* transcripts and complementary sites within *lin-14* and *lin-41* 3'-UTRs Left: Predicted stem-loop structures of *C. elegans lin-4* and *C. elegans*, *D. melanogaster* and *Homo sapiens let-7* longer transcripts. Graphical representation of the minimum free energy prediction are generated by RNAfold with default parameters. The ~21-nt region is highlighted in yellow. Right: Complementarity between *lin-4* and *let-7* ~21-nt regions and *lin-14* and *lin-41* 3'-UTRs. Dots indicate absence of a nucleotide and dashes one or more non-complementary nucleotides. Adapted from Lee et al., 1993; Pasquinelli et al., 2000

### stRNAs through the prism of RNAi

Meanwhile, another conserved RNA-mediated silencing phenomenon was attracting much attention: experimental introduction of double-stranded RNA (dsRNA) with one strand complementary to an endogenous gene can specifically induce its silencing (Fire et al., 1998). In animals, this process was called RNA interference (RNAi), and it is conserved from plants to

fungi but under other names, respectively *Post-Transcriptional Gene Silencing* (Hamilton et al., 1999) and *Quelling* (Romano et al., 1992). In very few years, numerous laboratories demonstrated in various model organisms, including nematodes, flies, fungi, and plants, that long dsRNAs are processed by an RNase III-family nuclease, named DICER in drosophila (Bernstein et al., 2001), into RNA species of 21-23 nt, called “*small interfering RNAs*” abbreviated siRNAs (Hamilton et al., 1999; Parrish et al., 2000; Yang et al., 2000; Zamore et al., 2000). Such siRNAs mediate the guidance of RNA-directed nucleases to induce complementary mRNA degradation (Hammond et al., 2000).

It did not take long for the similarities between siRNAs and stRNAs to be discussed. The most obvious are their conserved length of 20-23 nt and their processing from a double-stranded precursor: a long dsRNA for siRNAs and a stem-loop RNA for stRNAs. The first step to intersect these two pathways was the demonstration that both require DICER to generate this active small RNA species that induce gene expression silencing (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight et al., 2001).

### **A novel class of small RNAs defined by molecular features**

Following, three laboratories took advantage of knowledge gathered about siRNAs processing and RNAi *in vitro* models to identify concurrently new endogenous small RNAs. Notably, the Tuschl laboratory developed a cDNA cloning procedure to isolate siRNAs from lysate of *Drosophila* embryos that recapitulates RNAi *in vitro* (Elbashir et al., 2001; Tuschl et al., 1999). Each sequence was mapped to the reference genome and already characterized functional RNAs (rRNA, snRNAs, tRNAs), and sequences with no database entry were excluded. This protocol was used to uncover small endogenous RNAs in a human cell line (HeLa cells) and a *Drosophila* cell line (S2 cells) (Lagos-Quintana et al., 2001). The Bartel laboratory adapted this protocol to make it specific for RNAs with 5'-terminal phosphate and 3'-terminal hydroxyl, characteristic of RNase III-cleaved products (Bernstein et al., 2001). Cloned sequences from *C. elegans* total RNA mapping on predicted stem-loop transcripts were selected (Lau et al., 2001). The Ambros laboratory on its side performed a bioinformatics screening of stem-loop structures validated by Northern-blotting and a cDNA cloning of size-fractionated total RNA from *C. elegans* followed by an analysis of predicted stem-loop structure of most conserved sequences (Lee et al., 2001).

Altogether, they reported a total of approximately a hundred additional genes for such endogenous small non-coding RNAs: 16 new genes in flies, 21 in human cells, and 55 in nematodes. The phylogenetic conservation, abundance and diverse expression patterns of so-called stRNAs imply that they function in various regulatory pathways, in addition to their known role in the temporal control of developmental events. Eventually, because of their small size and the diversity of their potential biological function, this larger class of endogenous non-coding RNAs that are about 20-23 nt in length and are processed from hairpin secondary structures was renamed “*microRNAs*” abbreviated miRNAs.

### microRNAs across species

Since their discovery in *C. elegans* more than two decades ago, miRNAs or miRNA-like RNAs have also been documented in land plants (Jones-Rhoades et al., 2006) and filamentous fungi (Lee et al., 2010), but also in green algae (Molnár et al., 2007), brown algae (Cock et al., 2010; Tarver et al., 2015) and social amoebae (Avešson et al., 2012). However, the miRNA pathway diverges between these lineages regarding miRNA biogenesis, with non-homologous protein effectors and different intracellular localization, as well as on the mechanism of target RNA recognition and repression. Nevertheless, these ~20 nt-long RNAs all share the common feature of being derived from hairpin-folded precursors that are successively cut to form duplexes, one strand of which mediates the repression of mRNAs with a complementarity site. The details of the differences in the miRNA pathway between these different species will not be described in this thesis, but readers can turn to the following reviews: Axtell et al., 2011 and Moran et al., 2017.

Currently, the miRNA pathway is considered to have emerged independently in those distinct lineages because of the different biogenesis, the lack of miRNA sequence homology and the different mode of repression (Axtell et al., 2011). Exactly, these divergent miRNA pathways would originate from a more ancient small RNA regulatory system such as the RNAi pathway (Ghildiyal et al., 2009), shared in all three eukaryotic kingdoms, plants, animals, and fungi, and based on endogenously produced small interfering RNAs (endo-siRNAs) which provide an additional layer of control on protein-coding gene expression or serve as a cellular defense system against foreign or deleterious nucleic acid, such as transposable elements (Claycomb, 2014).

Eventually, some viruses have been reported to express miRNA genes and exploit the infected cell miRNA machinery to regulate host and viral genes (Pfeffer et al., 2004).

## 1.2 microRNA Biogenesis in Metazoan

In the following section, we will present the features of miRNAs in metazoans, and particularly in humans. For ease of reading by a wider audience, gene and protein equivalents in the model organisms *C. elegans* and *D. melanogaster* will be specified whenever possible. For further information specific to miRNA biogenesis in these two model species, readers are redirected to the following reviews: Soleimani et al., 2020; Youngman et al., 2014.

### 1.2.1 Genomic Distribution and Annotation of microRNA Genes

According to the manually curated database MiRGeneDB (Fromm et al., 2020), the number of confidently identified miRNA loci is estimated to be around 500 in the human genome, between 400 and 500 for mammals and around 150 for *C. elegans* and *D. melanogaster*. MiRNA loci are either hosted by a long non-coding transcript whose usually the only known function is to produce miRNAs, or overlapping protein-coding gene introns or UTRs.

At the time of the identification of the first miRNAs, a nomenclature convention has been established for novel miRNAs that is still applied (Ambros et al., 2003). The maintenance of these rules and the monitoring of novel miRNAs assignment is currently the responsibility of the online database miRBase. The names in this registry are of the form hsa-mir-123 (Griffiths-Jones et al., 2008)<sup>1</sup>. **(i)** The first three letters specify the species, e.g. hsa for *H. sapiens*. **(ii)** Except for miRNAs identified by forward genetics that are named in accordance with the mutant phenotype (e.g. *let-7*, *lin-4* and *bantam*), the mature miRNAs are designated “miR” and the precursor hairpin loci “mir”, **(iii)** followed by an identifying number, in the order of their discovery. To note that miRNA orthologs – identical miRNA sequences in other species – are given the same number. **(iv)** Paralogous miRNAs – distinct precursor hairpin loci in the same species – that generate similar mature miRNAs are given the same number followed by a numeral, if fully identical, or letter suffix, if alike.

MiRNAs are grouped into “families” based on the identity of their seed region (nucleotides 2-7), which recapitulates their targeting properties as detailed in section 1.3.1.

Human miR-25 family		Murine miR-25 family		
hsa-miR-25	<b>CAUUGCA</b> CUUGUCUCGGUCUGA	mmu-miR-25	<b>CAUUGCA</b> CUUGUCUCGGUCUGA	) identical ) ) alike ) ) paralogos
hsa-miR-92a-1	<b>UAUUGCA</b> CUUGUCCCGGCCUGU	mmu-miR-92a-1	<b>UAUUGCA</b> CUUGUCCCGGCCUG	
hsa-miR-92a-2	<b>UAUUGCA</b> CUUGUCCCGGCCUGU	mmu-miR-92a-2	<b>UAUUGCA</b> CUUGUCCCGGCCUG	
hsa-miR-92b	<b>UAUUGCA</b> CUCGUCCCGGCCUCC	mmu-miR-92b	<b>UAUUGCA</b> CUCGUCCCGGCCUCC	
hsa-miR-363	<b>AAUUGCA</b> CGGUAUCCAUCUGUA	mmu-miR-363	<b>AAUUGCA</b> CGGUAUCCAUCUGUA	

orthologs

FIGURE 1.2: **The miR-25 family nomenclature.** The five members of the human miR-25 family present an orthologous miRNA in mice sharing the same name. Paralogos miR-92a-1 and miR-92a-2 arise from two different genome locations producing the same miRNA sequence, while miR-92b sequence differs slightly. MiRNA seeds are highlighted in bold.

<sup>1</sup>Plant and viral miRNA gene naming conventions differ slightly. Plant miRNA genes are of the form ath-MIR123 with letter suffixes specifying distinct loci expressing same mature miRNAs and numeric suffixes are not used. Viral miRNA names often relate to the locus from which the miRNA derives e.g. ebv-mir-BART1 from the Epstein Barr virus BART locus (Griffiths-Jones et al., 2008).

## 1.2.2 Transcription and Nuclear Processing of Canonical microRNAs

### Pri-miRNA transcription

MiRNA loci are transcribed by RNA polymerase II<sup>2</sup> as part of several kilobase-long 5′-end capped and 3′-end poly(A) tailed transcripts harboring one or multiple distinctive hairpin structures named “*pri-miRNA*” or *primary precursor for miRNA* within which the mature miRNA is embedded (Cai et al., 2004; Lee et al., 2004; Lee et al., 2002). Typically, a canonical pri-miRNA is formed of a 33-35 bp double-stranded stem – including Watson-Crick pairing (A-U and C-G) as well as some wobble base pairs (G-U, I-U, I-A, and I-C) – where a few mismatches or small bulges are tolerated, an unstructured terminal loop and unstructured single-stranded flanking segments (Han et al., 2006).

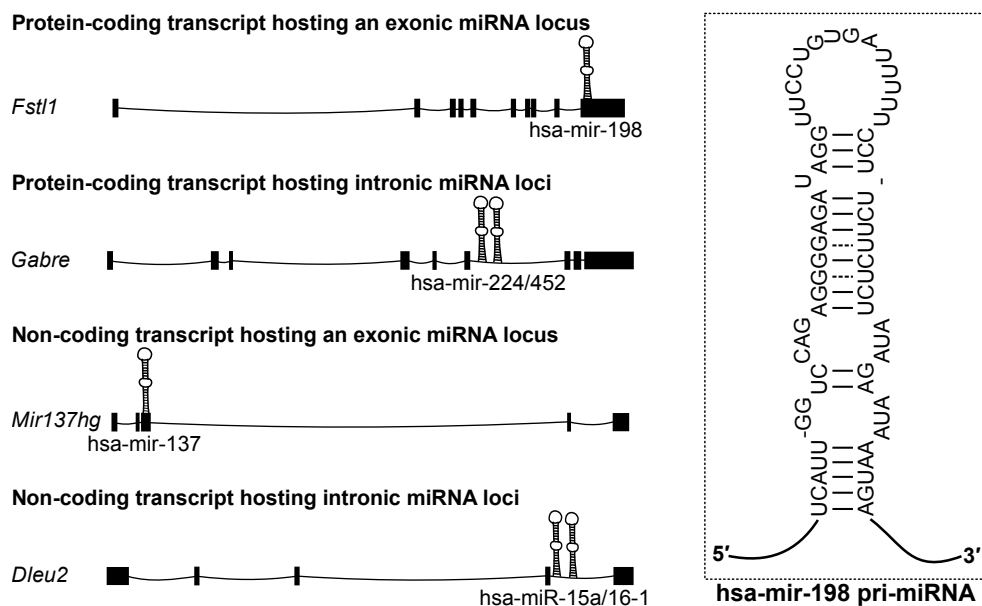


FIGURE 1.3: Examples of genomic location of miRNA loci and sequence of pri-miRNA. Left: miRNA loci can be categorized into four groups according to their genomic locations relative to exon and intron positions. Gene illustrations are not to scale, black boxes indicate exons and stem-loop structures represent miRNA loci. Right: Schematic of the pri-miRNA sequence of the human miR-198. Straight line represents Watson-Crick pairing and dotted line wobble pairing.

MiRNA loci are often clustered in a same host gene and consequently produce a single pri-miRNA with multiple hairpins, similar to a polycistronic transcriptional unit (Lee et al., 2002). If we consider that within 50 kb of each other, neighboring miRNAs exhibit highly correlated expression patterns (Altuvia et al., 2005; Baskerville et al., 2005), suggesting that they derive from a polycistronic transcript, we can estimate that 55% of conserved miRNAs in mammals form clusters<sup>3</sup> (Chiang et al., 2010). Commonly, there are two to three miRNA genes in a cluster but large clusters were also identified such as the conserved hsa-miR-17-92 cluster (six members) (Lagos-Quintana et al., 2001) or the imprinted human 14q32 cluster (forty

<sup>2</sup>If indeed most miRNA precursors are transcribed by RNA polymerase II, there are at least some viral miRNAs transcribed by Pol III (Bogerd et al., 2010; Diebel et al., 2010).

<sup>3</sup>The database miRBase defines a miRNA cluster within a maximal distance of 10 kb between miRNA genes. In actual fact, the term of a miRNA cluster has no clear definition.

members) (Seitz et al., 2004). For simplification, miRNAs generated from a polycistronic pri-miRNA are regarded as expressed from distinct miRNA loci in close proximity to each other. Consequently, ~500 miRNA loci in the human genome generate around ~200 pri-miRNAs (Chiang et al., 2010).

### **Maturation of pri-miRNA into pre-miRNA by the Microprocessor**

Following transcription, pri-miRNA hairpins are cropped by an heterotrimeric complex comprising one DROSHA protein associated with two DGCR8 proteins (for DiGeorge syndrome critical region 8 homolog; named PASHA in flies and nematode (Denli et al., 2004)), and known as Microprocessor complex (Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004; Lee et al., 2003). DROSHA is a RNase-III-type endonuclease, with one dsRNA-binding domain and two Rnase III domains. Its binding domain secures the recognition of the basal junction of the pri-miRNA stem while its RNase III domains act independently of each other: one cuts the 3' strand at ~22 bp away from the apical loop junction; the second cleaves the 5' strand ~11 bp away from the basal junction (Han et al., 2006; Nguyen et al., 2015; Zeng et al., 2005). DGCR8 proteins recognize the apical region of the pri-miRNA and interact with the stem through their two dsRNA-binding domains while stabilizing DROSHA-substrate interaction through their DROSHA-binding domain (Han et al., 2006; Nguyen et al., 2015; Yeom et al., 2006). Recently, novel stable components of Microprocessor have been identified: the Enhancer of Rudimentary Homolog (ERH) (Fang et al., 2020; Kwon et al., 2020) and the Scaffold Attachment Factor B2 (SAFB2) (Hutter et al., 2020). They are associated with DGCR8 and DROSHA respectively and both facilitate efficient processing of sub-optimal hairpins in polycistronic pri-miRNAs<sup>4</sup> (Fang et al., 2020; Hutter et al., 2020; Kwon et al., 2020). . Eventually, Microprocessor-induced cleavages generate a 2-nt 3' overhang hairpin with a 22 bp stem and 5'-phosphoryl and 3'-hydroxyl ends (Basyuk et al., 2003; Lee et al., 2003) named "*pre-miRNA*" abbreviated of *precursor for miRNA*.

Multiple sequence features also act as modules of interactions between Microprocessor and the pri-miRNA by emphasizing the orientation of pri-miRNAs and facilitating Microprocessor binding. The apical UGU motif is recognized by DGCR8 at the apical loop and indirectly favors the basal junction-DROSHA interaction by preventing DROSHA from binding to the apical junction (Nguyen et al., 2015). The basal UG motif is directly recognized by DROSHA at the basal junction and enforces accurate cleavage by DROSHA (Auyeung et al., 2013; Nguyen et al., 2015). The mismatched GHG motif in the stem has been shown to be involved in efficient pri-miRNA maturation but its mechanism remains unknown (Fang et al., 2015; Kwon et al., 2019). Finally, the CNNC motif at the 3' flanking region of pri-miRNA associates with SRSF3/SRp20, a splicing factor that recruits DROSHA to the basal junction of pri-miRNAs (Auyeung et al., 2013; Kim et al., 2018). All of these primary sequences are not necessarily

---

<sup>4</sup>This phenomenon named "cluster assistance" is detailed in section 1.2.5



required for proper pri-miRNA processing<sup>5</sup>. They act independently, to various extent, and enhance both the efficiency and the accuracy of Microprocessor.

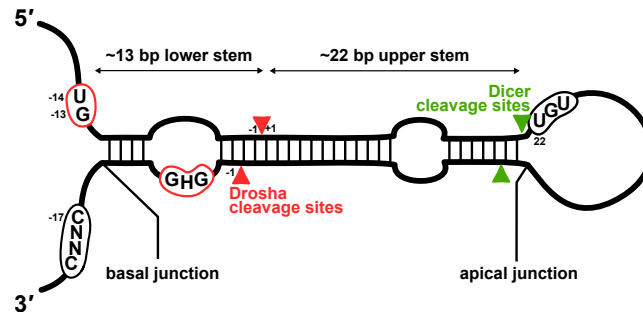


FIGURE 1.4: **Primary and secondary structures of canonical pri-miRNAs.** This schematic of a pri-miRNA shows the structural regions involved in miRNA biogenesis machinery recognition and the short motifs required for proper miRNA maturation in humans. Small numbers refer to established nucleotides position and arrows indicate DROSHA (red) and DICER (green) cleavage sites.

### 1.2.3 Cytoplasmic Maturation of Canonical microRNAs

#### Nuclear export of pre-miRNA

Once isolated, the pre-miRNA hairpin is exported into the cytoplasm by a transport complex, comprising the nuclear transport receptor protein exportin 5 (XPO5) with GTP-binding nuclear protein RAN-GTP (Bohnsack et al., 2004; Kim et al., 2016; Lund et al., 2004; Yi et al., 2003). This complex recognizes the 2-nt 3' overhang structure and the double-stranded stem of the pre-miRNA (Okada et al., 2009) and translocates it through a nuclear pore complex.

#### Processing of pre-miRNA into miRNA duplex by DICER

In the cytoplasm, the pre-miRNA is recognized by DICER and cleaved near the apical loop, releasing a ~22-bp miRNA duplex (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight et al., 2001). DICER is a RNase-III-type endonuclease harboring two RNase III domains, a dsRNA-binding domain and, unlike DROSHA, a DExD/H-box helicase domain, a DUF283 (Domain of Unknown Function) domain, and a Platform-PAZ-connector (PAZ for Piwi/Argonaute/Zwille) domain. The helicase domain recognizes the apical loop of the pre-miRNA (Gu et al., 2012; Tsutsumi et al., 2011) while the Platform-PAZ-connector domain binds the 2-nt 3' overhang and the 5' phosphate end (Macrae et al., 2006; Park et al., 2011; Tian et al., 2014). RNase III domains form a catalytic centre which cleaves pre-miRNA ~22-nt from the 5' end of the precursor to produce a mature miRNA duplex with ~2-nt overhang on both ends (Lee et al., 2003; Park et al., 2011; Zhang et al., 2004). Like DROSHA, human DICER is associated with double-strand RNA binding partners: TAR RNA-binding protein or TRBP (Chendrimada et al., 2005; Haase et al., 2005; Lee et al., 2006) (Loquacious

<sup>5</sup>The described motifs are specific to human pri-miRNAs and are shared among mammals. In nematodes, these features are unknown (Auyeung et al., 2013).

for DICER-1 et R2D2 for DICER-2 in flies (Förstemann et al., 2005; Saito et al., 2005)) regulates pre-miRNA cleavage accuracy (Kim et al., 2014; Wilson et al., 2015), and PACT, a human paralog of TRBP, (Lee et al., 2013a; Lee et al., 2006), has been reported to bind DICER and to influence pre-miRNA maturation but its exact role in miRNA biogenesis is still unknown and questioned in human cell lines (Kim et al., 2014).

### 1.2.4 **microRNA-Inducing Silencing Complex Formation**

Once generated, the miRNA duplex is loaded into an Argonaute protein to initiate the assembly of a ribonucleoprotein complex named *miRNA-induced silencing complex* or “*miRISC*” (Hammond et al., 2001; Mourelatos et al., 2002). MiRISC loaded with a duplex RNA becomes active miRISC after the removal of one of the two strands. The remaining strand is known as the guide strand of the duplex or else the mature miRNA, while the ejected strand is the passenger strand of the duplex formerly called miRNA\* or miRNA star.

The family of Argonaute proteins – named from the appearance of the *Arabidopsis thaliana* mutant which reminded the authors of a squid (Bohmert et al., 1998) – is divided into two subclasses: the ubiquitously expressed AGO subclass associated with miRNAs and siRNAs (*e.g.* AGO1 and AGO2 in flies<sup>6</sup> and AGO1, AGO2, AGO3 and AGO4 in mammals) and the germ-cell-specific PIWI subclass associated with piRNAs (*e.g.* PIWI, AUB and AGO3 in flies and HIWI, HIWI2, HILI and PIWIL3 in humans) (Ozata et al., 2019)<sup>7</sup>.

AGO proteins are small-RNA binding proteins with a three-dimensional bilobal structure: the N-terminal lobe is composed of an amino-terminal (N) domain and a PAZ (Piwi/ Argonaute/ Zwiille) domain, while the C-terminal lobe is formed of a MID (Middle) domain and a PIWI (P element-induced wimpy testis) domain (Song et al., 2004; Wang et al., 2008). The channel formed between the two lobes can accommodate a short single-stranded or double-stranded RNA, with the 3' hydroxyl group and 5' phosphate of the guide strand binding to the PAZ domain and to a pocket at the interface between the MID and PIWI domain respectively (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle et al., 2012). The PIWI domain adopts an endoribonuclease RNase H fold that can catalyze the cleavage of an RNA strand (Liu et al., 2004). Among the four human AGO proteins, this slicer activity has only been demonstrated for AGO2 (Meister et al., 2004).

In humans, the loading of a miRNA duplex into AGO requires the ATP-dependent chaperone activity of Hsc70/Hsp90 (Iwasaki et al., 2010) and the polarity of duplex loading is thought to be determined by AGO alone (Kobayashi et al., 2016). The 5'-end of the guide strand of the duplex is recognized by AGO's 5' nucleotide-binding pocket and is followed by the accommodation of the rest of the duplex into the channel. Because the binding of a 5'-phosphate end

---

<sup>6</sup>In flies, miRNA and siRNA duplexes are selectively loaded by AGO1 and AGO2 respectively (Tomari et al., 2007).

<sup>7</sup>In eukaryotes, a third subclass of AGO proteins could be defined: the worm-specific WAGO proteins associated with secondary siRNAs termed 22G-RNAs (Almeida et al., 2019).

bends the RNA strand and unpairs the first nucleotide of the 5′-end, the guide strand is distinguished by its relative thermodynamic stability: the strand with the less stable 5′ would be more easily anchored in AGO and consequently functions as the guide (Ma et al., 2005; Parker et al., 2005). Following, the miRNA duplex is partially unwound through active wedging by the N-domain of the 3′-end of the guide strand (Kwak et al., 2012; Wang et al., 2009). Thus, the guide strand is thoroughly anchored in the channel of AGO by both ends whereas the passenger strand has no specific interaction with the protein and is partially unpaired. This results in a passive ejection of the unstable strand due to structural constraints by various domains of the AGO protein (Kawamata et al., 2009; Kawamata et al., 2011; Yoda et al., 2010).

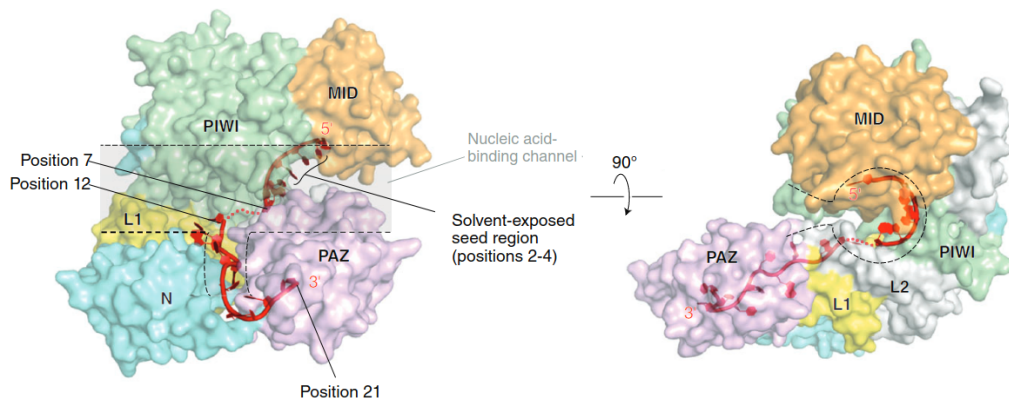


FIGURE 1.5: **Structures of human AGO2 associated with a guide RNA.** The nucleic acid-binding channels are highlighted with dotted lines. The guide RNA is colored in red and depicted as a ribbon model. The disordered parts of the guide are shown as dotted lines. Adapted from Nakanishi, 2016.

### 1.2.5 Non-Canonical microRNA Biogenesis Pathways

The microRNA biogenesis pathway described above generates the vast majority of miRNAs, also known as canonical miRNAs. Alternative mechanisms bypass Microprocessor or DICER processing and produce functional non-canonical miRNAs (Babiarz et al., 2008; Kim et al., 2016). It is remarkable that most non-canonical miRNAs – except miR-451 – are poorly conserved in evolution and much less abundant than canonical ones, which tempers their functional importance (Bartel, 2018).

#### Microprocessor-independent mirtrons

As we mentioned in section 1.2.1, miRNA loci are often nested in intronic regions of coding or non-coding genes. The expression of intronic miRNAs and their host gene mRNA is overall positively correlated (Baskerville et al., 2005), however the relationship between miRNA transcription-splicing-microprocessing is not completely understood, oscillating between independence (Kim et al., 2007), cooperation (Janas et al., 2011; Kataoka et al., 2009; Morlando et al., 2008) and competition (Agranat-Tamir et al., 2014). It appears that usually, miRNA cropping by Microprocessor occurs during ongoing transcription (Morlando et al., 2008) or after transcription and precedes splicing events (Bracht et al., 2004; Kataoka et al., 2009; Kim et al.,

2007), even facilitates intron degradation (Janas et al., 2011; Morlando et al., 2008) or trans-splicing<sup>8</sup> in flies and nematodes (Bracht et al., 2004; Kataoka et al., 2009).

A subclass of intronic miRNAs named “*mirtrons*” were identified first in flies and nematodes (Okamura et al., 2007; Ruby et al., 2007) then also observed in mammals (Berezikov et al., 2007). Mature miRNAs generated from mirtrons map to the end of short introns (> 100 nt) predicted to fold into a Microprocessor-insensitive hairpin structure since it comprises only the miRNA duplex region flanked by splicing junctions and the variable-length terminal loop, and lacks the lower stem of 11 nt required for Microprocessor recognition. In details, these pre-miRNA-sized short introns are not Microprocessor substrate, instead they are directly spliced and adopt a lariat structure in which the 5′ splice junction is covalently linked to the 3′ branch point. The lariat can be debranched by a lariat-debranching enzyme resulting in a 5′-end phosphate and 3′-end hydroxyl single-strand sequence (Padgett et al., 1984; Ruskin et al., 1985) which spontaneously folds into a hairpin structure with a 2-nt overhang, mimicking a canonical pre-miRNA. The obtained pre-miRNA-like species converges with the canonical miRNA pathway and is transferred to the cytoplasm by XPO5 where it is further processed by DICER to produce a functional mature miRNA.

5′ and 3′ tailed mirtrons have also been identified, located at one end of the intron instead of ending precisely with splice donor or splice acceptor sites. For example, with a 3′ tailed mirtron, its 5′-end matches the 5′ splice donor and its 3′-end is followed by an unstructured region. Such mirtron intermediates are subjected to tail resection by unclear mechanisms and thus converge with the canonical miRNA biogenesis (Wen et al., 2015).

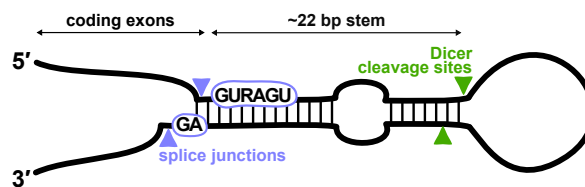


FIGURE 1.6: **Primary and secondary structures of mirtron pri-miRNAs.** This schematic shows the structural regions involved in splicing and miRNA biogenesis machineries recognition and the motifs required for splicing in humans. Arrows indicate the spliceosome (purple) and DICER (green) cleavage sites.

### Microprocessor and XPO5-independent 5′-capped miRNAs

For some miRNAs, including miR-320 and miR-484, transcription is coupled with early termination and results in a short capped 5′-end and RNA pol II-terminated 3′-end nascent transcript that folds into a pre-miRNA-like structure (Xie et al., 2013). Unlike canonical pre-miRNA, such miRNAs are 5′-capped thereby XPO5-insensitive. They are instead exported to the cytoplasm by the PHAX-XPO1 (Phosphorylated Adaptor for RNA export-dependent Exportin 1) pathway (Xie et al., 2013). Surprisingly, 5′-capped pre-miRNAs are efficiently processed by DICER, but the 5′-cap prevents AGO anchoring of the strand, thus mature miRNA

<sup>8</sup>Exons from two distinct pre-mRNA fragments are joined and spliced by the spliceosome (Dye et al., 2006; Lasda et al., 2011).

from such precursors are restricted to the 3'-strand (Babiarz et al., 2008; Chong et al., 2010; Kim et al., 2016; Xie et al., 2013).

### **DICER-independent miR-451**

MiR-451 is an erythropoietic miRNA highly conserved in vertebrates (Patrick et al., 2010; Rasmussen et al., 2010) produced from an unusual 41–42-nt pre-miR-451 hairpin with a highly structured 17-nt stem region, too short to be cleaved by DICER. Instead, pre-miR-451 is directly loaded into AGO proteins. When loaded into AGO2, the only AGO with an active endonucleolytic activity in vertebrates, pre-miR-451 hairpin is sliced in the middle of its 3'-strand, which produces a 30-nt intermediate RNA (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010). PARN, a Poly(A)-specific 3'-to-5' exonuclease, further trims the 3'-end of this intermediate to produce the mature 23-nt-long form of miR-451 (Yoda et al., 2013).

MiR-451 is transcribed from a polycistronic locus together with miR-144 and despite its poor structural features, miR-451 is one of the most highly expressed miRNAs in erythroblasts and erythrocytes, exceeding the abundance of miR-144, while this latter is produced from a canonical hairpin structure (Juzenas et al., 2017). An explanation could be a compensation mechanism called *cluster assistance* and described by Fang and Bartel (Fang et al., 2020). They demonstrated that defective miRNA hairpins (*e.g.* miR-451) could be efficiently processed if they are in the vicinity of a hairpin that can be efficiently processed on its own (*e.g.* miR-144). This mechanism would be conserved in mammalian and may require ERH and SAFB2 (Fang et al., 2020; Hutter et al., 2020; Kwon et al., 2020).

### **TUT-dependant miRNAs**

A few miRNAs, such as *let-7* and miR-105 family in vertebrates, are mainly generated from non-standard DROSHA-cleaved pre-miRNAs with a 1-nt 3' overhang. These shorter forms are not recognizable for the following maturation steps as seen, and are restored through 3'-end mono-uridylation catalyzed by Terminal Uridyl Transferases (TUTs), including TUT2, TUT4 and TUT7 (Heo et al., 2012; Kim et al., 2015).

1.2. *microRNA Biogenesis in Metazoan*

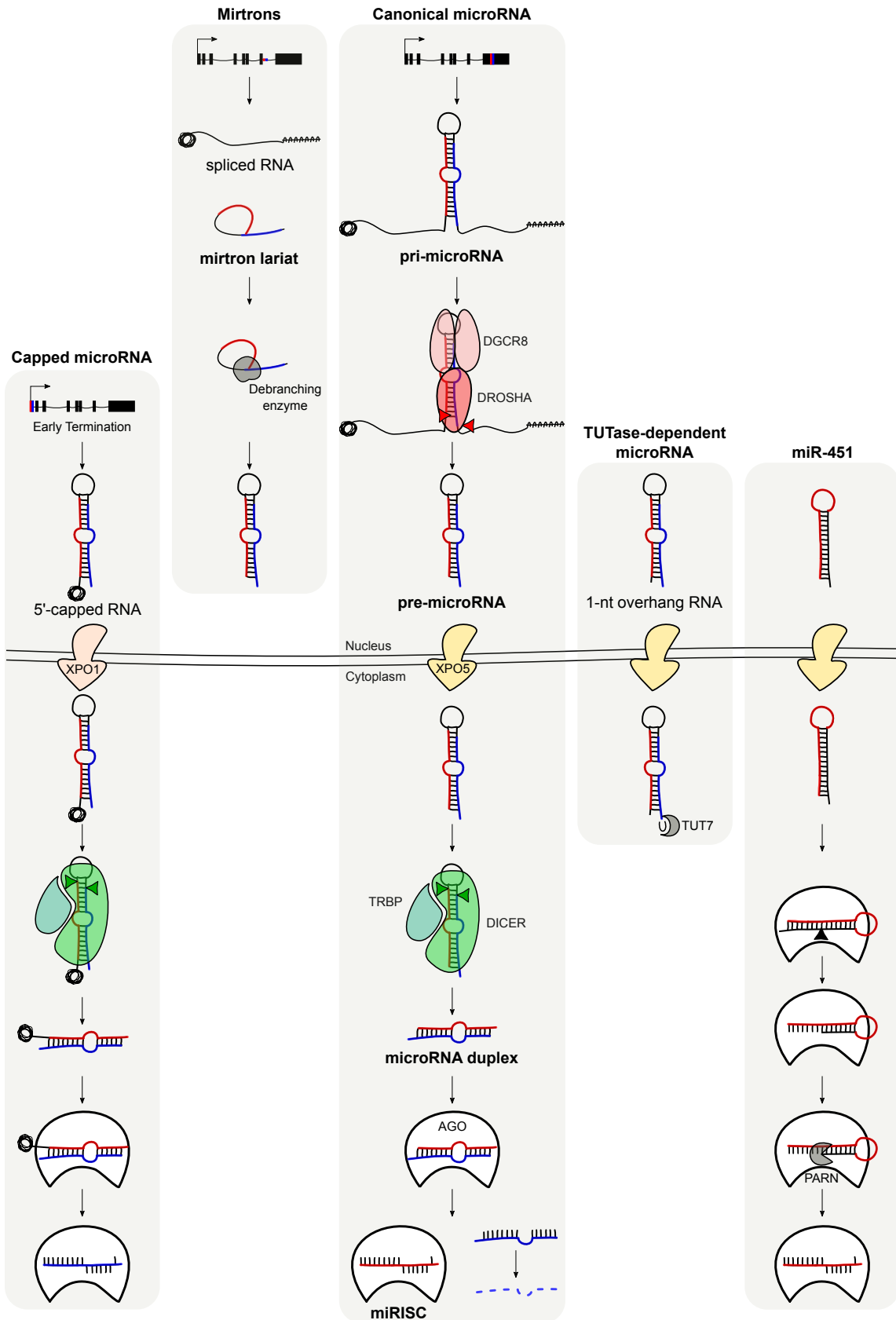


FIGURE 1.7: **Biogenesis of canonical and non-canonical miRNAs.** Most miRNAs are produced by the canonical miRNA biogenesis pathway, in the center of the figure, and the divergent steps of non-canonical miRNA synthesis pathways are shown alongside for comparison.

## 1.3 Molecular Mechanisms of microRNA-Mediated Gene Silencing

### 1.3.1 Principles of microRNA Targeting

As mentioned in section 1.1, it soon became apparent that miRNAs recognize target mRNAs by base-pairing with conserved regulatory elements mostly located in the 3'-UTR of transcripts. However, in metazoans, such interaction does not necessarily involve a perfect complementarity, as illustrated by lin-14:lin-4 and let-7:lin-41 bulged duplexes, which complexify the definition and the identification of genuine miRNA-responsive elements (MRE).

For further information, methods of miRNA targets identification are detailed in Mockly et al., 2019, introduced in section 3.1.1.

#### The notion of miRNA seed

Complementarity between miRNAs and known regulatory elements or validated miRNA targets located in 3'-UTR of transcripts revealed that target site sequences tend to perfectly base-pair with the 5'-end of the miRNA, especially with nucleotides 2 to 8 (Lai, 2002; Stark et al., 2003). This region of the miRNA was also shown to be the most conserved among homologous miRNAs (Lewis et al., 2003; Lim et al., 2003) and reciprocally, MRE residues complementary to nucleotides 2 to 8 at the 5'-end of the miRNA are strongly conserved in orthologous species (Stark et al., 2003). Altogether, the miRNA 5'-end seemed critical for miRNA targeting and since have been referred to miRNA residues 2 to 8 as the miRNA seed (Lewis et al., 2003). This finding has been reinforced by *in vivo* and *in cellulo* assays for target site regulation by artificial miRNAs or endogenous miRNA overexpression or knockdown (Brennecke et al., 2005; Giraldez et al., 2006; Lim et al., 2005; Rodriguez et al., 2007). Consistent with the pairing evaluation and evolutionary conservation, such experiments indicate that complementarity to the miRNA seed is generally sufficient to promote *in vivo* miRNA-mediated regulation and that mRNAs containing seed matches tend to either decrease upon miRNA addition or increase upon miRNA disruption.

Structural analysis of miRISC provides an explanation for the importance of the miRNA seed sequence in miRNA targeting. As described in section 1.2.4, AGO accommodates the guide miRNA in a central channel and thereby modulates the accessibility of the different regions of the miRNA. This conformation creates a seed chamber where nucleotides 2 to 7 at the 5'-end of the miRNA are arranged in a helical conformation with nucleotides 2 to 4 exposed and accessible for the initial stable target pairing (Nakanishi et al., 2012; Schirle et al., 2012; Schirle et al., 2014). It also form a supplementary chamber in which up to 5 nucleotides, preferentially around nucleotides 13 to 16, can pair an additional region of the mRNA target. These two chambers are separated by a narrow central gate which prevents targets from interacting with the miRNA central region, around nucleotides 9 to 11 (Sheu-Gruttadauria et al., 2019). Likewise, both extremities of the miRNA do not participate in target pairing because of threading through narrow clefts and their ends anchored within AGO (Nakanishi et al., 2012; Schirle et al., 2012). Like so, the miRNA sequence is compartmentalized into the following

subdomains: the 5' anchor (nt 1), the seed (nt 2-7), the center (nt 8-12), the 3' complementarity (nt 13-16) and the tail sequence (nt 17-end of the miRNA) (Wee et al., 2012). Interestingly, the term seed was given to reflect the idea that target pairing is initiated from the seed and could then spread to other segments of the miRNA (Bartel, 2004).

### Canonical and non-canonical miRNA target sites

The Bartel lab, which coined the term of miRNA seed<sup>9</sup>, has been particularly involved in the characterization of miRNA targeting. In particular, they provide a web resource, TargetScan (<http://www.targetscan.org/>), for MRE prediction based on a statistical model that consider multiple features (miRNA sequence, MRE sequence, MRE conservation of mRNA sequence around the MRE) (Agarwal et al., 2015). From the analysis of preferentially conserved miRNA-pairing motifs, they propose to distinguish three types of MREs (Bartel, 2009; Bartel, 2018; Friedman et al., 2009): canonical sites, atypical canonical sites and noncanonical sites.

Canonical miRNA target sites correspond roughly to perfect seed pairing with various flavors. In order of decreasing preferential conservation and efficacy:

- 8mer: Watson-Crick pairing to miRNA nt 2 to 8 + an A opposite to miRNA nt 1
- 7mer-m8: Watson-Crick pairing to miRNA nt 2 to 8
- 7mer-A1: Watson-Crick pairing to miRNA nt 2 to 7 + an A opposite to miRNA nt 1
- 6mer: Watson-Crick pairing to miRNA nt 2 to 7
- offset 6mer: Watson-Crick pairing to miRNA nt 3 to 8

The identity of the nucleotide opposite to miRNA nucleotide 1 (relative to its 5'-end) appears to be biased with a preference for A, regardless the miRNA sequence (Lewis et al., 2005). Nucleotides A at the first position of miRNAs are specifically bound by a narrow pocket of AGO, providing an additional anchor that favors miRISC interaction with seed-matched sites (Schirle et al., 2015).

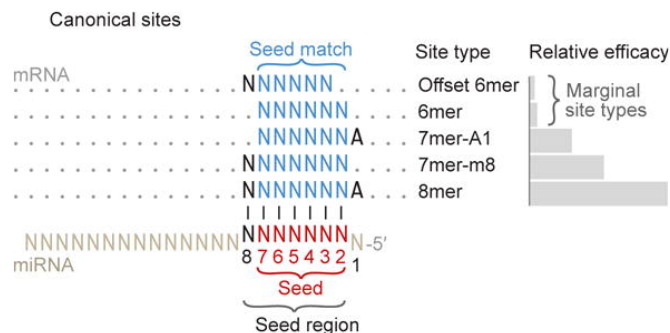


FIGURE 1.8: Canonical sites of mammalian miRNAs. From Bartel, 2018.

Atypical canonical miRNA target sites refer to perfect seed-matched sites associated with a 3'-supplementary Watson-Crick pairing centered on nucleotides 13 to 16 (Grimson et al., 2007). This additional pairing enhances the efficacy of seed-matched sites (Brennecke et al.,

<sup>9</sup>According to Bartel's terminology, the seed refers to nucleotides 2-7 at the 5'-end of the miRNA while the "seed region" or "extended seed" refers to nucleotides 2-8 (Bartel, 2018).



2005; Grimson et al., 2007). Noncanonical miRNA target sites include imperfect seed-matched sites associated with 3'-compensatory sites, also centered on nucleotides 13 to 16 (Friedman et al., 2009). An other type of noncanonical MRE are centered sites, that lack both perfect seed-match and 3'-compensatory pairing. Instead they exhibit at least 11 contiguous nucleotides that pair to a miRNA at positions 4-14 or 5-15 (Shin et al., 2010). Nevertheless, in term of abundance and selective conservation, atypical and non-canonical MREs are patently inconsequential (Friedman et al., 2009) as they constitute 5% and less than 1% of all MREs respectively (Bartel, 2009).

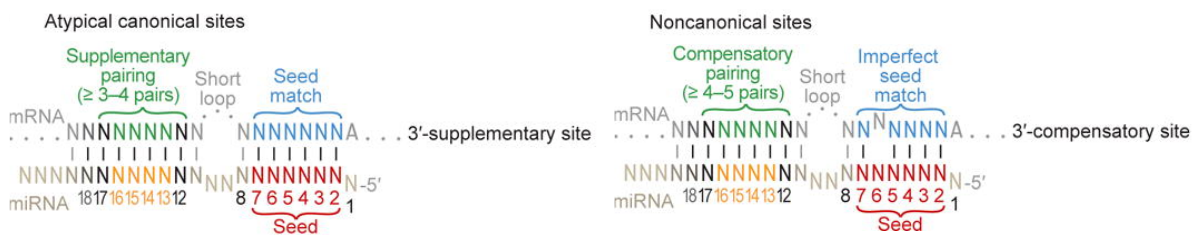


FIGURE 1.9: **Atypical and non-canonical sites of mammalian miRNAs.** From Bartel, 2018.

### MRE localization in transcriptional units

The search for novel MREs was first focused on 3'-UTRs based on initially reported miRNA targets. However, another point that justifies the preferential computational investigation within this region of mRNAs - by comparison with coding sequences - is that 3'-UTRs as a whole is not fully conserved and instead exhibit limited selectively conserved regions easily detectable from background. In this way, search for selective conservation can be combined with seed matching (Friedman et al., 2009; Krek et al., 2005; Lewis et al., 2003). Eventually, unsupervised computational and experimental identification of miRNA binding sites reveals that miRNAs interact more frequently or more effectively with 3'-UTRs than with other regions of mRNAs (Chi et al., 2009; Hafner et al., 2010; Lewis et al., 2005; Lim et al., 2005; Zisoulis et al., 2010). In addition, MREs have been shown to be more effective when located in not-translated regions, preferentially near both ends of the 3'-UTR but at least 15 nt after the stop codon (Grimson et al., 2007).

### 1.3.2 microRNA-Mediated Gene Silencing

In animals, two distinct mechanisms contribute to miRNAs' overall gene silencing effect: a cap-dependent repression of translation initiation and an mRNA decay following deadenylation and decapping. The relative importance of each effect depends on cellular context: development stage, cell type, growth condition, the concentration of miRISC components, number and proximity of MRE, or translational state of the target. Genome-wide measurements on protein and mRNA levels as well as ribosome profiling estimate that mRNA destabilization explains 66 to 90% of steady-state miRNA-mediated repression (Baek et al., 2008; Eichhorn

et al., 2014; Guo et al., 2010; Hendrickson et al., 2009; Selbach et al., 2008). As a consequence, target mRNA levels can provide a good approximation of miRNA-directed silencing.

In terms of dynamics, translational repression was shown to occur prior to mRNA deadenylation and decapping (Béthune et al., 2012; Djuranovic et al., 2012). Recent preprints present single-molecule visualizations of each step of miRNA-mediated gene silencing *in cellulo* and establish that miRISC induces the repression of target mRNA translation within 30 min after binding and mRNA decay within 60 min after binding (Cialek et al., 2021; Kobayashi et al., 2021).

#### **The interplay between components of the miRISC silencing machinery**

The minimal miRISC, composed of the mature miRNA loaded into AGO, establishes the recognition and stable interaction with MREs but is not able on its own to mediate poly(A)-dependent target destabilization or cap-dependent translation inhibition. Instead, AGO binds the protein TNRC6 (exists as three paralogs TNRC6A/B/C in mammals; named GW182 in flies and AIN1 and AIN2 in nematodes), which acts as a hub to recruit effector proteins (Ding et al., 2005; Jonas et al., 2015). TNRC6 proteins are characterized by an N-terminal AGO-binding domain and a C-terminal silencing domain, both unstructured domains displaying multiple tryptophan-containing motifs which mediate proteic interactions (Lazzaretti et al., 2009; Takimoto et al., 2009; Till et al., 2007). TNRC6 recruits a number of effector proteins including the poly(A)-Binding Protein (PABP) and the deadenylase complexes CCR4-NOT and PAN2-PAN3 (Jonas et al., 2015).

PABP proteins coat the mRNA 3' poly(A) tail and simultaneously bind to the 5' cap-associated eukaryotic initiation factor complex (eIF4F)<sup>10</sup> via its subunit eIF4G (Kahvejian et al., 2005), directing the interaction between the mRNA 3' poly(A) tail and its 5'-cap. This circularization of mRNAs is essential for the recruitment of ribosomes, and thus for the translation of mRNAs in the cytoplasm (Gingras et al., 1999). In a miRNA-independent manner, PABP is also directly involved in recruiting deadenylases to the mRNA by directly binding the PAN3 subunit of the deadenylase complex PAN-PAN3 and indirectly the CCR4-NOT complex (Mauxion et al., 2009; Siddiqui et al., 2007).

The interaction between TNRC6 and PABP increases accessibility of the poly(A) tail to deadenylase complexes recruited by TNRC6 by bringing them closer, but also by promoting the displacement of PABP (Moretti et al., 2012) thereby leaving the poly(A) tail accessible and disrupting mRNA circularization which may facilitate both translational repression and deadenylation (Fabian et al., 2009; Huntzinger et al., 2010; Zekri et al., 2013). However, if PABP proteins promote miRNA-mediated silencing, they seem to not be a prerequisite for miRISC activity since miRNAs can still silence mRNA reporters lacking poly(A) tails, at least in some models (Fukaya et al., 2011; Mishima et al., 2012).

---

<sup>10</sup>The eIF4F complex is composed of three subunits: the DEAD-box RNA helicase eIF4A, the cap-binding protein eIF4E, and the scaffold protein eIF4G which interact with both eIF4E and eIF4A (Gingras et al., 1999).

The CCR4-NOT complex is obviously involved in deadenylation of target mRNAs but also contributes to translational repression (Kuzuoğlu-Öztürk et al., 2016; Mathys et al., 2014) through the direct and indirect recruitment of proteins promoting mRNA decapping and cap-dependent translation inhibition (Jonas et al., 2013; Nishimura et al., 2015; Ozgur et al., 2015). The central interactions that have been deciphered for now are the direct binding of DDX6 by the CCR4-NOT complex (Chen et al., 2014a; Mathys et al., 2014; Rouya et al., 2014) which triggers the exclusive recruitment of the eIF4e-binding protein (4E-T) by DDX6 (Ozgur et al., 2015). 4E-T was supposed to compete with eIF4G for binding to eIF4E, and in this way, sequester eIF4E and block translation initiation. However, tethered 4E-T are still able to repress translation when eIF4E-binding is prevented by mutagenesis (Kamenska et al., 2014; Waghray et al., 2015) while a defection of miRNA-directed mRNA decay is observed (Nishimura et al., 2015). Thus, 4E-T may promote translation inhibition by displacing the eIF4F complex from the cap but through an eIF4E-independent mechanism, as well as mRNA decay by contributing to the proximity between the miRISC machinery and the mRNA 5' cap through its interaction with eIF4E.

To synthesize, miRNA-directed gene silencing involves an interplay between components of the miRISC machinery that is still extensively investigated. In this attempt to present the main characterized units of miRISC and their function, some recently identified effectors were eluded among which a novel identified TNRC6 interactor, GIGYF2 (Schopp et al., 2017), and a eIF4E homolog which interacts with 4-ET, 4EHP (Chapat et al., 2017; Chen et al., 2017). For further information about these additional factors of miRNA-mediated silencing and other models under investigation, the reader is advised to turn to the following review: Duchaine et al., 2019.

Despite grey areas, the miRISC-induced regulatory events can be simplified in four steps. First, miRISC inhibits the cap-dependent initiation of the translation via the CCR4-NOT complex–DDX6–4E-T interplay in a deadenylation and eIF4E-independent manner. Secondly, miRISC mediates mRNA deadenylation by the CCR4-NOT and PAN2-PAN3 deadenylase complexes. Thirdly, the shortening of the poly(A)tail induces mRNA decapping by multiple proteins recruited by the CCR4-NOT complex or TNRC6, including the decapping complex DCP1-DCP2 (Nishimura et al., 2015). Eventually, deadenylated and decapped transcripts are degraded by the 5'-to-3' exoribonuclease 1 (XRN1) (Chen et al., 2011).

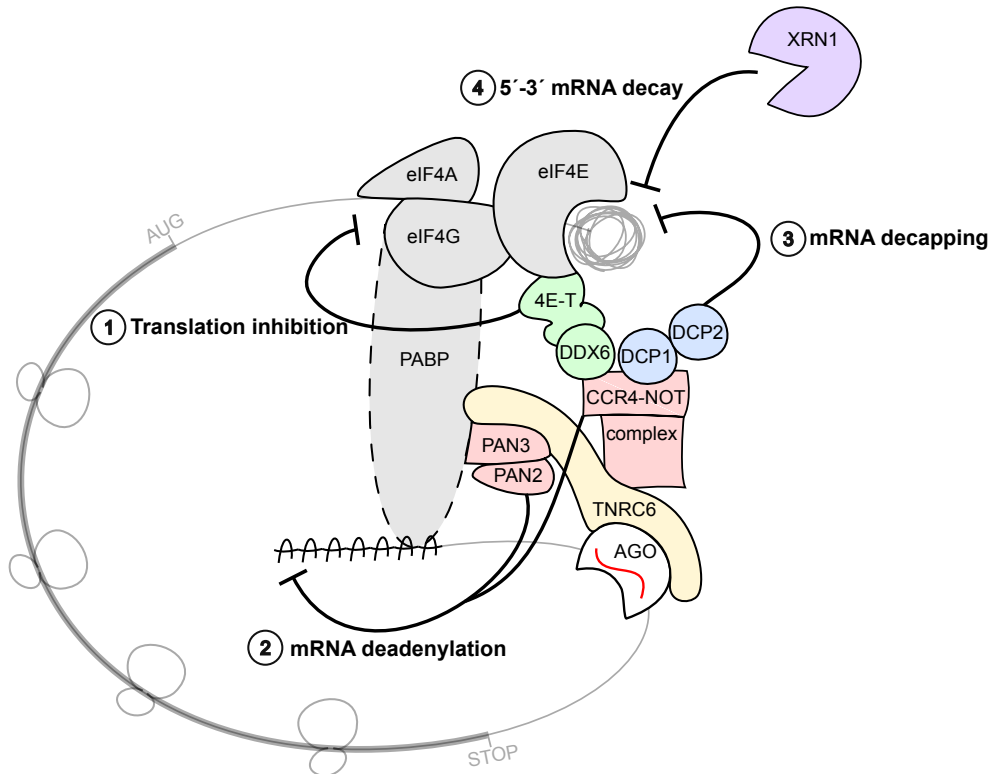


FIGURE 1.10: **Model of miRNA-mediated gene silencing.** The four steps of miRNA-mediated gene silencing with the involved protein actors are represented. A color code defines the primary function of represented proteins: cap destabilization in green; deadenylation in red; mRNA decapping in blue, 5'-to-3' decay in purple; miRISC central platform in yellow; cap- and tail-binding proteins in grey. PABP borders are dotted to illustrate its displacement from the poly(A) tail. AGO is not colored and a red strand represents the mature miRNA. The coding region of the mRNA is depicted with ribosomes between a start and an end codon. The cap structure is shown as a black roll and the poly(A) tail is indicated.

### RNAi-like target slicing

In metazoans, miRNAs can act as siRNAs in cases of highly or perfect complementarity (Salomon et al., 2015; Wee et al., 2012) and mediate slicing of the target transcript at a stereotypical position (Elbashir et al., 2001; Hutvágner et al., 2002). The endonucleolytic cleavage of the target strand occurs at the position facing nucleotides 10 and 11 of the miRNA and is catalyzed by an AGO protein with active endonucleolytic activity, typically AGO2 in mammals (Liu et al., 2004; Meister et al., 2004). Thus, target mRNAs perfectly complementary to miRNAs loaded into AGO2 in mammals are aimed to be sliced. But because such widely extensive interactions between miRNAs and their targets are not common compared with seed-matched pairings, very few examples of miRNA-directed cleavage targets have been reported in animals. In mammals, the only known examples are 20 cellular transcripts (Shin et al., 2010) of which *Hoxb8* (targeted by miR-126) (Yekta et al., 2004), *Rtl1* (targeted by *AntiPeg11*-hosted miRNAs) (Davis et al., 2005; Seitz et al., 2004); a circular RNA (Hansen et al., 2011) and few viral mRNAs (Barth et al., 2008; Sullivan et al., 2005).



## Chapter 2

# Dynamics and Physiological Relevance of microRNAs

### 2.1 Physiological Importance of microRNA Abundance and Function in Metazoans

Because the impact of miRNAs on biological pathways directly depends on their targets, the identification of every direct target mRNA of a miRNA should allow predicting its function. For this purpose, various experimental and computational methods have been developed over the years to detect or predict physical interactors for miRISC – reviewed in Mockly et al., 2019, *cf.* section 3.1.1. Because, among other things, the interaction between the miRISC and a mRNA is based on conserved short sequence pairing, such strategies report a myriad of potential miRNA-regulated transcripts. In general, it is admitted that over 60% of human protein-coding genes are conserved physical targets of miRNAs, from computational approaches (Friedman et al., 2009). Likely even more according to experimental approaches, which include non-canonical and non-conserved MREs (Grosswendt et al., 2014; Helwak et al., 2013), suggesting that miRNAs influence essentially all biological processes in animals.

#### 2.1.1 Global Effect of microRNAs on Development and Homeostasis

The importance of miRNAs in animal development and homeostasis has long been admitted since the first discovered miRNAs in *C. elegans* are required for proper developmental temporality. Nevertheless, to study the contribution of the miRNA-mediated silencing to other biological processes, deletion of the miRNA biogenesis factors DICER, DROSHA, DGCR8, and of the miRNA-associated protein AGO have been generated in various model organisms.

Interestingly, the zygotic repression of miRNA biogenesis during development does not equally impair all metazoans. Particularly in *C. elegans*, DROSHA- and DICER-deficient animals could reach the adult state and exhibit germline defects that lead to sterility (Denli et al., 2004; Ketting et al., 2001; Knight et al., 2001). While in *D. melanogaster*, zebrafish and mice, DGCR8 or DICER deletion causes embryonic or larval lethality (Bernstein et al., 2003; Giraldez et al., 2005; Martin et al., 2009; Wang et al., 2007; Wienholds et al., 2003). This irregular response of metazoans to global miRNA inhibition is explained by a differential maternal contribution which compensates zygotic protein depletion. Indeed, the maternal removal of DICER or

DROSHA/DGCR8 has been shown to exacerbate embryonic lethality in zebrafish (Giraldez et al., 2005) and to cause embryonic lethality in nematodes (Dexheimer et al., 2020; Lehrbach et al., 2012). Finally, the deletion of miRNA-loading *Ago* genes induce embryonic arrest during morphogenesis in every model animals, including nematodes, flies and mice (Alisch et al., 2007; Kataoka et al., 2001; Morita et al., 2007; Vasquez-Rifo et al., 2012). Altogether, these results agree with the importance of the miRNA activity in animal development. However, it should not be omitted that the inactivation of DICER, DROSHA, DGCR8 or AGO has other consequences in addition to impairing the miRNA biogenesis, including to affect the endo-siRNA pathway (Chung et al., 2008; Claycomb, 2014; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Tam et al., 2008; Watanabe et al., 2008) and the non-miRNA DROSHA substrate processing (Chong et al., 2010). However, this consideration is minimized by the overlap of phenotypes caused by the removal of these different components of the miRNA-pathway and collectively, they demonstrate the requirement of miRNAs for animal development.

Similarly, the extent of the post-developmental miRNA-mediated silencing in tissue homeostasis has been mainly investigated by conditional ablation of *Dicer*, *Drosha*, *Dgcr8* or *Ago* genes *in vivo* (Chong et al., 2008; Huang et al., 2012; Hébert et al., 2010; Kanellopoulou et al., 2005; Kobayashi et al., 2015). Recently, the Ventura lab has developed a genetically engineered mouse strain in which the miRISC complexing is reversibly disrupted based on the inducible expression of a TNRC6 competing peptide (La Rocca et al., 2021). In accordance with prior results, they found that the miRNA pathway is not essential for functionality in standard conditions for most adult tissues, except for the heart and the skeletal muscle. However, it could be required under exogenous stress, as demonstrated in the intestine and hematopoietic system, where it becomes essential for tissue regeneration following acute injury. These results emphasize the role of miRNAs as part of the cellular response to external and internal perturbations. They also suggest that, whether some miRNAs are necessary for proper morphogenesis, most miRNAs may be physiologically relevant only in sensitized genetic backgrounds or upon environmental perturbations at post-developmental stages, rather than in laboratory-controlled conditions, making their identification more challenging.

### 2.1.2 Physiological Contribution of Individual microRNAs

While miRNAs have been shown to be collectively essential in animals, the physiological function of individual miRNAs is more difficult to establish and requires the generation of collections of miRNA knockout animals. In nematodes, mutants for 95 miRNA loci (out of 151 according to MiRGeneDB) have been phenotypically characterized for obvious viability or development defects and have revealed that most miRNAs are not vital in a lab-controlled environment (Miska et al., 2007). Similarly, in flies, *in vivo* miRNA ablations of 130 individual miRNAs (out of 177) have demonstrated that only *bantam*, miR-1, miR-190 and miR-279/996 are essential for survival while most miRNA mutants leads to less evident phenotypic defects in tissue development and adult behavior (Chen et al., 2014b). Finally, in mice, the generation of conditional miRNA knockout mutants (Park et al., 2012) and independent studies

of miRNA-deficient animals (reviewed in Bartel, 2018) have revealed modest functions in a broad diversity of tissues and biological processes but primarily for the most conserved mammalian miRNA families, while individual miRNAs are not commonly indispensable. This observation could be explained by the functional redundancy among miRNA family members, *i.e.* miRNAs which share the same seed. Because the main feature for miRNA targeting is the seed sequence, members of a miRNA family would inevitably silence the same range of targets and compensate the loss of one family member. For example, in mice, the miR-34a family is composed of six members and while the deletion of the whole family leads to high postnatal mortality, with surviving animals displaying an array of phenotypes associated with defective ciliogenesis, mice carrying targeted deletions of single genes of the miR-34 family are viable and phenotypically normal (Song et al., 2014). However, in *C. elegans*, among the 23 known miRNA families, mutant animals for 12 families do not exhibit phenotypic defects even when subjected to a broad panel of phenotypical characterization (Alvarez-Saavedra et al., 2010). This suggests that most miRNA families, at least in nematodes, might have subtle functions and require appropriate models to be studied, such as genetically sensitized backgrounds (Brenner et al., 2010).

Unsurprisingly, the severity of miRNA knockouts seems correlated with the conservation of miRNA families in evolution (Bartel, 2018). For instance, the most conserved miRNAs among bilaterian animals are, for more than 80% of them, associated with severe abnormal knockout phenotypes in mice, including early lethality, reduced viability and developmental disorders. Additionally, in mammalian miRNAs conserved to fish, 62% have been associated with an abnormal knockout phenotype in mice. In the review Bartel, 2018, an enumeration of abnormal phenotypes observed after deletion is presented for ~100 murine miRNA families, and emphasizes the diversity of observable defects including: “*reduced viability, fatal neurological disorders, infertility, blindness, deafness, immune disorders or cancer*”. Furthermore, the murine model permits to address the importance of miRNAs for proper development or maintenance of specific organs (“*skeleton, teeth, brain, eyes, neurons, muscle, heart, lungs, kidneys, vasculature, liver, pancreas, intestine, skin, fat, breast, ovaries, testes, placenta, thymus, and each hematopoietic lineage*”), the regulation of cellular processes (“*axon sprouting, synapse formation and function, mitotic spindle orientation, polyploidization, ciliogenesis, and diverse functions in various hematopoietic lineages*”), physiological processes (“*cardiac conduction, blood pressure, lipid or cholesterol metabolism, insulin production, pituitary function, mobilization of glycogen, Abpeptide degradation, bone resorption, fibrosis, and the overall growth of embryos or pups*”), and finally, in the response to various diseases, infections and in the propensity to develop cancers.

Consequently, among the most studied animal organisms, *C.elegans*, *D. melanogaster* and *M. musculus*, most of the miRNA families exert a moderate regulation on cell- or tissue-specific phenotypes but are dispensable in standard conditions, while the very most conserved are required for animal development or viability.



### 2.1.3 Abnormal Regulation and Function of microRNAs in Cancer

Neoplastic transformation is caused by the accumulation of genetic lesions that ultimately impair signaling pathways (cell cycle checkpoint induction, DNA repair, induction of cell death...) and convert normal cells into tumor cells with uncontrolled proliferation and survival, unlimited replicative potential, and invasive growth (Hanahan et al., 2011). Because miRNAs are part of various developmental and homeostatic processes in a tissue-specific manner, their mutational status may have an impact on tumorigenesis. Their expression profiles are indeed altered in almost all types of cancer compared with healthy tissue (Lu et al., 2005; Volinia et al., 2006). Interestingly, miRNA expression profiles are particularly informative, by comparison with mRNA profiles, and reflect the tissue of origin and the differentiation state of the tumors (Ferracin et al., 2011; Lu et al., 2005; Rosenfeld et al., 2008). Therefore, miRNAs have emerged as potential biomarkers for cancer diagnosis.

An example of miRNAs presumed to act as a oncogenic miRNA is the miR-17~92 cluster, located at chromosome 13q31, a region amplified in the Burkitt's lymphoma, diffuse large B-cell lymphoma (DLBCL), follicular lymphoma, mantle cell lymphoma and lung cancer (Ota et al., 2004), and overexpressed in a large cohort of human B-cell lymphomas and multiple solid tumor types (He et al., 2005; Lu et al., 2005; Volinia et al., 2006). Furthermore, *in vivo* deletion of mir-17~92 in mice results in postnatal lethality with multiple developmental defects, including lung hypoplasia and ventricular septal defect, as well as in the direct derepression of the pro-apoptotic protein BIM (Ventura et al., 2008). It should be noted that the mir~17-92 cluster produces a polycistronic transcript that generates six individual mature miRNAs, including miR-17, 18a, 19a, 20, 19b, and 92a which belong to four distinct miRNA families, miR-17 (including miR-17 and 20), miR-18, miR-19 (including miR-19a and 19b), and miR-92 family. It is thus conceivable that the distinct biological effects of all four miRNA families collectively contribute to the oncogenic activity of the mir-17~92 polycistron.

Following the assumption that some miRNAs may directly contribute to neoplastic transformation through the derepression of pro-oncogenic factors, or on the contrary be involved in the activation of tumor-suppressor pathways, the cancer field early started investigating the potential of miRNAs therapeutics. More precisely, miRNA mimics and antimiRs are estimated as promising treatments to, respectively, replenish tumor-suppressor miRNAs, or suppress oncogenic miRNAs (Rupaimoole et al., 2017).

However, if miRNAs are rising biomarkers for the profiling of cancer subtypes and could be helpful for this matter, the function of individual miRNAs in cancer is more questionable. Specifically, studies querying miRNA functions by overexpressing or silencing specific miRNAs have yielded data often at odds with those collected from knockout models. An illustration of this issue is presented in section 3.3.2. To date, the miRNA literature is particularly contaminated by papers reporting the role of some miRNAs in some cancer models through the regulation of any tumor-suppressor or pro-oncogenic gene (Byrne et al., 2020). A simple way to observe this problematic phenomenon is to submit a random human miRNA name in

miRBase and inspect the word cloud illustrating the “Literature search”: the term “cancer” almost always appears.

## 2.2 Regulation Mechanisms of microRNA Abundance

As illustrated by the tightly regulated temporal expression of *lin-4* and *let-7*, most miRNAs have well-defined temporal and cell-type-specific expression patterns, while others tend to be more ubiquitous (Landgraf et al., 2007; Lee et al., 2008). In either case, because miRNA cellular abundance results from the balance between miRNA biogenesis and decay, spatial and temporal patterns of miRNA distribution depend on the regulation at both levels.

### 2.2.1 Dependence Between microRNA Levels and Function

Current high-throughput sequencing technologies are highly sensitive and can amplify even traces of miRNAs. Usually, hundreds of individual miRNAs are detected by cell type, and an unfortunate mistake is to consider that all of them are potentially functional. Instead, miRNA-mediated post-transcriptional repression is highly dependent on miRNA concentration and target abundance (Mullokanov et al., 2012). To present a general idea, according to Mullokanov and colleagues, the majority of miRNAs or miRNA families (>80% in the given example) mediate substantial regulation of a natural target when expressed above 1,000 reads per million (RPM)<sup>1</sup>. Consequently, for a given cell type, only the most highly expressed miRNAs are able to significantly induce RISC-directed gene silencing.

### 2.2.2 Dynamics of microRNAs

Recent papers provide accurate measurements of miRNA production, loading, and decay rates in *Drosophila* cells (S2 cells) and murine cells (contact-inhibited MEFs, dividing MEFs, and mouse embryonic stem cells) (Kingston et al., 2019; Reichholf et al., 2019). Both methods are based on metabolic labeling of newly synthesized RNAs followed by small RNA sequencing. Among all their common findings, they reported that miRNA steady-state abundance is generally well positively correlated with biogenesis rates, suggesting that miRNA production is the primary determinant for intracellular miRNA abundance. Besides, both mammalian and *Drosophila* miRNA biogenesis rates are exceptionally high, with a median of 18 copies per hour per cell in contact-inhibited MEFs and a maximum of 6,600 copies per minute per cell for miR-21a-5p in dividing MEFs. For comparison, mRNAs’ median transcription rates are about 2 molecules per hour, with the fastest reported mRNA produced at 500 copies per hour per cell (Schwanhäusser et al., 2011) – yet this comparison has to be taken with caution since miRNA biogenesis includes transcription *and* maturation to miRNA duplexes. This fast production of miRNAs is bottlenecked by the slow miRISC assembly, typically within 30 minutes in murine cells and over an hour in S2 cells. In *Drosophila* cells, this constraint was predicted to result

---

<sup>1</sup>RPM calculations are not detailed in the Material and Methods of this paper, nor those of the cited papers so it is possible that the normalization was performed from the total number of reads or else from the number of reads mapping miRNA sequences. Thus, this value should be taken cautiously.

in ~40% of all miRNAs degraded before AGO-loading (Reichholf et al., 2019). Such overproduction of miRNAs likely participates in the specific loading of miRNAs into AGO proteins to the detriment of others ncRNAs. Finally, as reported in former studies (Baccarini et al., 2011; Guo et al., 2015; Marzi et al., 2016), the vast majority of miRNA guide strands behaves as a stable molecule with a half-life over 24 hours in murine cells – while passenger strands exhibit a half-life ranged from <0.01 to 1 hour. Obviously, guide strands' higher stability is attributable to anchoring into AGO protein. Still, a subset of miRNA guide strands demonstrates fast turnover rates between 1 and 10 hours and likely participate in rapid environmental changes or signaling events. Among these short-lived miRNAs, a sequence-specific active degradation mechanism has been lately reported and is detailed in section 2.2.5.

### 2.2.3 Control of Global microRNA Biogenesis

Post-transcriptional control of miRNA biogenesis depends, for the whole miRNA pool, on the regulation of the maturation machinery proteins: DROSHA, DGCR8, DICER, TRBP and cofactors.

#### Regulatory loop of maturation factors

The Microprocessor requires correct subunit stoichiometry for proper assembly and processing: two molecules of DGCR8 for one molecule of DROSHA (Nguyen et al., 2015). The autoregulation between DROSHA and DGCR8 is a notable example of homeostatic control of miRNA biogenesis. DGCR8 mRNAs exhibit hairpin structures in their 5'-UTRs (Pedersen et al., 2006) which are cleaved by the Microprocessor, resulting in DGCR8 transcript destabilization. On the other hand, protein-protein interaction between DGCR8 and DROSHA stabilizes the DROSHA protein (Han et al., 2009; Triboulet et al., 2009). Another example of a regulatory loop regards the negative feedback loop between *let-7* and *Dicer*. The miRNA *let-7* directly targets *Dicer* mRNA through MREs located within its 3'-UTR and thus reduces DICER protein levels. Conversely, consequent downregulation of DICER results in a reduced global abundance of mature miRNAs, including *let-7* (Forman et al., 2008; Tokumaru et al., 2008).

#### Post-transcriptional modifications of maturation factors

Proteins associated with miRNA biogenesis are subjected to various post-transcriptional modifications that modulate their stability, nuclear localization, and processing activity. Particularly, the Mitogen-Activated Protein Kinase and Extracellular signal-Regulated Kinase (MAPK/ERK) signaling pathway (Guo et al., 2020) was reported to regulate miRNA biogenesis at multiple stages.

Regarding the Microprocessor, phosphorylation of DROSHA by the Glycogen synthase kinase 3  $\beta$  (GSK3 $\beta$ ) enhances DROSHA interaction with DGCR8 and cofactors, and pri-miRNA maturation, thus maintaining DROSHA nuclear localization (Fletcher et al., 2017; Tang et al., 2010; Tang et al., 2011). On the contrary, activated p38(MAPK) directly phosphorylates other DROSHA residues under stress conditions, inhibiting the DROSHA-DGCR8 interaction and promoting DROSHA translocation to the cytoplasm and degradation (Yang et al., 2015). Apart

from phosphorylation, acetylation of the N-terminal domain of DRISHA competes with its ubiquitination, thus preventing its degradation through the ubiquitin-proteasome pathway (Tang et al., 2013).

MAPK/ERK can also phosphorylate DGCR8, but in this case, the modification improves protein stability (Herbert et al., 2013). Deacetylation of the RNA-binding domains of DGCR8 by the Histone Deacetylase 1 (HDAC1) was also reported to increase DGCR8 interaction with pri-miRNAs and then miRNA processing (Wada et al., 2012).

Concerning DICER, its phosphorylation state appears crucial for *C. elegans* oocyte-to-embryo transition. During oocyte development, phosphorylation of DICER residues driven by MAPK/ERK triggers its nuclear relocalization, which prevents the maturation of pre-miRNA. Prior fertilization and embryogenesis, inactivation of MAPK/ERK induces rapid DICER dephosphorylation, and consequently proper localization and miRNA biogenesis (Arur et al., 2009; Drake et al., 2014).

Finally, TRBP can also be phosphorylated by either MAPK/ERK (Paroo et al., 2009) and the S6 kinase (Warner et al., 2016), which seem to improve the pre-miRNA processing complex stability.

### 2.2.4 Regulation of Individual microRNA Biogenesis

Control of individual miRNA abundance depends primarily on pri-miRNA expression regulation but also on structural features of the precursor sequence, detailed in section 1.2.2, and post-transcriptional modifications, which influence its affinity with the maturation and miRISC machineries.

#### Transcriptional regulation

MiRNA loci are hosted in protein-coding genes or expressed from non-coding genes with their own promoters (Saini et al., 2007). However, if the majority of miRNA loci located in introns of protein-coding genes shares the promoter of the host gene, others seems to be dependent of alternative promoters (Monteys et al., 2010; Oszolak et al., 2008). In any case, they are under Pol II transcriptional control, as indicated in section 1.2.2. Therefore, miRNA genes expression regulation is equivalent to that established of protein-coding genes, and depends on Pol II-associated transcription factors (TF), enhancers and silencers, promoter methylation status and epigenetic events such as DNA methylation or histone modifications (Krol et al., 2010). Among the best known examples of TF-activated miRNAs, the miR-17~92 cluster is induced by c-Myc (O'Donnell et al., 2005) and three members of the miR-34 family are direct transcriptional targets of p53 (He et al., 2007). It should be noted that in cases where mRNAs and miRNAs are produced from common transcripts, the latter often reaches higher cellular levels thanks to its high stability when recruited in miRISC (Baccarini et al., 2011; Bartel, 2018).

#### Precursor miRNA sequence modifications

Canonical biogenesis of a miRNA – as described in section 1.2 – supposes that one precursor sequence provides a single miRNA sequence, annotated as the mature reference miRNA in

miRBase. Actually, processing variability and end modifications can generate multiple miRNA sequences detectable by sequencing. Mature miRNA variants processed from the same precursor and slightly different in length or sequence are referred to as *isomiRs* (Morin et al., 2008).

About processing variabilities, DROSHA and DICER sequential cleavages direct the position of the miRNA 3' and 5' ends. Thus, any inaccuracy in cutting of the RNase III nucleases changes the miRNA sequence and creates 5' and 3' isomiRs (Kim et al., 2017). Especially, 5'-variant isomiRs result in a shifted seed; as a consequence, they do not belong to the same miRNA family and recognize a new range of target mRNAs.

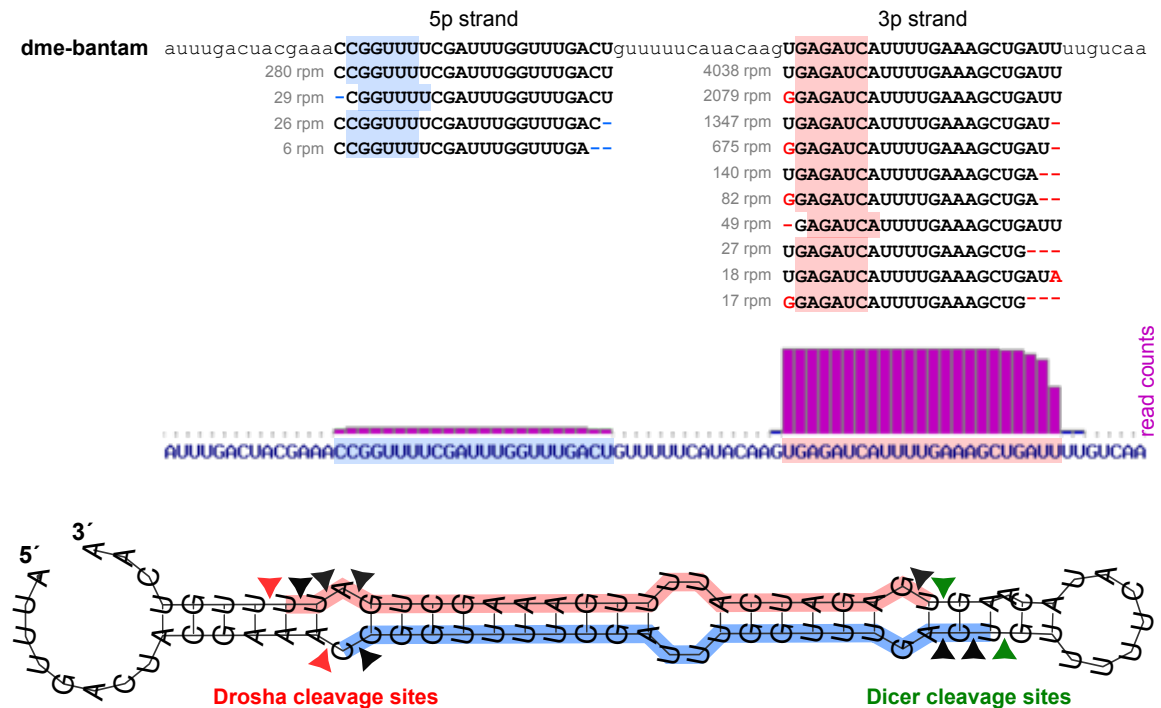


FIGURE 2.1: **The repertoire of *dme-bantam* isomiRs.** Top: A diverse variety of isomiRs were observed from wild-type *D. melanogaster* small RNA-seq data generated by our laboratory. The first line corresponds to the pre-miRNA *bantam* sequence with the reference mature *bantam*-3p and 5p highlighted in bold. IsomiRs are ranged below in order of decreasing abundance indicated by the number of reads per million for each variant (for each sequence, the number of reads was divided by the total number of reads mapping on the reference repertoire of miRNAs, then multiplied by one million). Colored rectangles highlight seed sequences of the reference miRNA and its isomiRs. Absence of nucleotides in isomiR sequences is illustrated by a colored dash and differences in sequence are colored (blue for *bantam*-5p and red for *bantam*-3p). Median panel is from miRBase. Bottom: Graphical representation of the minimum free energy structure prediction generated by RNAfold from the extended pre-miRNA *bantam* sequence. The reference mature *bantam*-5p and 3-p are colored respectively in blue and red. Arrows indicate DROSHA (red) and DICER (green) canonical cleavage sites or alternative miRNA ends (black).

Interestingly, there is less variability and extent in 5' isomiRs, with only single base additions or deletions, than in 3'<sup>2</sup> (Tan et al., 2014). This observation suggests other processing activities, notably, the action of 5'-to-3' polymerases, typically Terminal Nucleotidyl Transferase (TENTs) and 3'-to-5' exonucleases on either precursors or mature miRNAs. About 3'-end miRNA tailing, uridylation occurs mainly on pre-miRNAs and disturbs miRNA maturation –

<sup>2</sup>An easy method to recognize proper miRNA sequence is to create a deep-sequencing profile of matching reads and expect a sharp 5'-border and a blurry 3'-border.

except for the pre-let-7 monouridylation *cf.* section 1.2.5 – as illustrated below, while adenylation occurs on mature miRNAs and does not show a consistent effect on miRNA activity and stability as detailed in section 2.2.5 (Chiang et al., 2010; Newman et al., 2011).

For instance, the pre-let-7 uridylation by TUT4 and TUT7, in the presence of the stem cell-specific RNA-binding protein LIN28, is the most studied example of pre-miRNA polyuridylation (Hagan et al., 2009; Heo et al., 2008; Heo et al., 2009; Thornton et al., 2012). The addition of a short 3′-poly(U) stretch inhibits DICER recognition and pre-miRNA processing. Instead, the modified pre-miRNA is degraded by the 3′-to-5′ exoribonuclease DIS3L2 (Ustianenko et al., 2013). In the same way, in *D. melanogaster*, mirtron hairpins are specifically subjected to 3′ uridylation compared to canonical hairpins (Westholm et al., 2012) and the TUTase Tailor is necessary and sufficient for this mirtron modification. Because mirtrons are generated from spliced transcripts, their 3′-end is defined by the consensus 3′ splice site AG, *cf.* section 1.2.5. Consequently, mirtron hairpins are preferential substrates of Tailor – which exhibits a sequence preference for 3′-G –, relative to canonical pre-miRNAs (Bortolamiol-Becet et al., 2015; Reimão-Pinto et al., 2015). Finally, uridylation of mirtron hairpin 3′-ends impairs DICER recognition and processing and promotes their 3′-to-5′ exoribonucleolytic decay via the RNase II/R enzyme CG16940, an homolog of the human Perlman syndrome exoribonuclease Dis3l2 (Reimão-Pinto et al., 2016).

RNA editing such as adenosine to inosine (A-to-I) conversion occurs on some pri-miRNAs and influences miRNA biogenesis and activity (Luciano et al., 2004). A-to-I editing is catalyzed by the Adenosine Deaminases Acting on RNA (ADAR) in double-stranded RNAs, typically in pri-miRNA hairpin structures. Inosine behaves as guanosine and does not base-pair with uridine but with cytidine, creating bulges in the double-stranded structure (Nishikura, 2016). Consequently, A-to-I editing on pri-miRNA can interfere with various steps of miRNA biogenesis or targeting: **(i)** the binding of ADAR on miRNA precursors may compete with miRNA processing factors (Vesely et al., 2014); **(ii)** editing of the pri-miRNA sequence can alterate Microprocessor or DICER cleavages (Kawahara et al., 2008); **(iii)** editing the mature miRNA sequence can decrease the base-pairing stability with canonical target mRNAs and even reprogram the range of targets (Kawahara et al., 2007). Approximately 20% of pri-miRNAs are edited in the adult human brain (Kawahara et al., 2008), but only a low frequency of edition (>5%) is detected on mature miRNAs (Alon et al., 2012).

### **Influence of RNA-binding proteins**

After transcription, pri-miRNAs can be recognized by specific RNA-binding proteins (RBP) other than the miRNA maturation machinery. For instance, the stem cell factor LIN28, as mentioned above, identifies sequence elements in the loop of pre-let-7 miRNAs and recruits the terminal uridyltransferases TUT4 or TUT7 to pre-miRNAs (Heo et al., 2009; Thornton et al., 2012), resulting in their degradation. Treiber and colleagues recently investigated, at a large scale, RBPs involved in the direct miRNA biogenesis regulation (Treiber et al., 2017). They performed a mass-spectrometry-based screen for binders of immobilized pre-miRNAs (72 pre-miRNAs in 11 different cell line lysate) and identified ~180 RBPs that interact specifically with

individual pre-miRNAs *in vitro*. For functional validation of identified miRNA binders, they used combined transient RNAi knockdown or stable CRISPR/Cas9-mediated knockout experiments to analyze RBP-dependent changes in miRNA levels. Eventually, this study provides a repertoire of pre-miRNA interacting RBPs and demonstrates the tissue-specificity of this layer of miRNA abundance regulation.

### 2.2.5 Control of microRNA Stability

Despite their global high stability, individual miRNAs in specific cellular contexts can exhibit highly dynamic concentrations facilitated by accelerated decay. For instance, neuronal miRNAs exhibit a substantially faster turnover rate than in other cells (1 hour or less for murine neurons), dependent on neuronal activity (Krol et al., 2010; Rajasethupathy et al., 2009). Thus, during dark adaptation, light-dependent transcription inhibition of miR-183~96~182 cluster, miR-201, and miR-211 is sufficient to allow a sudden drop in their cellular levels (Krol et al., 2010). Similarly, in murine cell lines, the constitutive short half-life of several members of the miR-16 family permits rapid changes in their levels in response to biogenesis alteration. In this way, it allows precise and coordinated cell-cycle transitions through derepression of specific mRNAs (Rissland et al., 2011). Consequently, active miRNA degradation pathways are essential for inducing rapid and spatially localized changes in miRNA concentration, thus in target gene expression.

#### 3'-tailing and trimming impact on miRNA stability

The 3'-end trimming or addition of tails of miRNAs, and their composition, participate in determining the fate of the modified miRNAs. In flies, ~40% of AGO1-loaded miRNAs are trimmed at their 3'-end (Han et al., 2011), and notably the 3'-to-5' exoribonuclease Nibbler has been shown to mediate 3' trimming of ~60% of all mature miRNAs (Reichholf et al., 2019). Nibbler-directed trimming occurs after AGO1 loading and passenger strand removal, and preferentially targets atypically long and unstable miRNAs (Han et al., 2011; Liu et al., 2011; Reichholf et al., 2019). This shortening is supposed to allow proper miRNA anchoring into AGO1 and miRISC silencing efficiency.

In mammals, miR-122 is a conserved liver-specific miRNA that plays an important role in the regulation of hepatic function. Its expression in hepatocytes is tightly regulated and appears to be determined by the balance between tailing and trimming of its 3'-end (Chang et al., 2004). Mature miRISC loaded with miR-122 can associate *in vitro* with the RNA-binding protein QKI-7 which recruits the cytoplasmic poly(A) polymerase GLD2 (also known as TENT2) to AGO (Hojo et al., 2020). GLD2 directly adenylates a subpopulation of miRNAs including miR-122 for which such mono or poly(A) tailing induces miRNA stabilization (D'Ambrogio et al., 2012; Katoh et al., 2009). Once miR-122 has dissociated from miRISC, it is recognized by the CUG-binding protein 1 CUGBP1 which may recruit the poly(A)-specific ribonuclease PARN to induce miRNA degradation (Katoh et al., 2015).

The *Drosophila* ortholog of GLD2, WISPY, has also been reported to adenylate maternally inherited miRNAs in eggs and early embryos (Lee et al., 2014). Interestingly, ~15% of all

miRNA reads detected in eggs are adenylated against ~3% in *Drosophila* adult tissues (Berezikov et al., 2011). WISPY interacts directly with AGO1 and mono or polyadenylates miRNAs. Unlike GLD2-mediated stabilization of miR-122 in mammal hepatocytes, WISPY-mediated adenylation in flies embryo induces miRNA decay. Typically, adenylated miRNAs have a short average half-life of ~2 hours, implying the involvement of active miRNA decay and explaining the rapid clearance of maternal miRNAs during maternal-to-zygotic transition.

As a last example, during T lymphocyte (or T-cell) activation, a subpopulation of uridylylated miRNAs is specifically degraded. TUT4 has been reported to catalyze this 3' poly(U)-tailing in naïve T-cells which, in the context of T-cell activation, triggers miRNA degradation, likely by poly(U)-specific exoribonucleases (Gutiérrez-Vázquez et al., 2017).

Known mature miRNA tailings are reviewed in Yu et al., 2020. As illustrated with these examples, and confirmed by metabolic labeling of murine miRNAs: each miRNA differentially acquires terminal modifications and then differentially responds to these modifications. Interestingly, in MEFs, tailing and trimming are not necessarily correlated with changes in miRNA stability, which implies substantial independence between miRNA modification and degradation pathways (Kingston et al., 2019). This observation was reinforced by the findings about the active decay pathway detailed below.

### Target-directed miRNA degradation

Recently, a newly defined active sequence-specific miRNA degradation pathway has attracted much attention. Specifically, because it allows for discrimination between miRNAs sequences, it could explain the dynamics of individual miRNAs. However, before describing the related findings, we should begin with the very first observations of this mechanism.

Loss-of-function approaches support the study of miRNA function in model organisms, and genetic abolition of miRNA loci is state of the art in this matter. However, due to their difficulty of application before the development of CRISPR-Cas9, other approaches were preferred. Notably, the transient expression of antisense chemically modified single-stranded RNA analogs – also known as “antimirs” or “antagomirs”, depending on the chemical modifications. Interestingly, this approach was initially developed to hold miRNAs back through the formation of duplexes and thus to prevent miRNA activity (Hutvagner et al., 2004; Meister et al., 2004). However, unexpectedly, the most long-lasting antisense oligoribonucleotides can induce miRNA destabilization, resulting in an efficient miRNA abundance decrease (Krützfeldt et al., 2005; Xie et al., 2012).

In 2010, the Zamore lab confirmed that expression of reporter mRNAs bearing fully complementary MREs in their 3'-UTRs in *Drosophila* cells or the addition of antimirs in human cells, result in a decrease of cognate miRNAs – except for 2'-O-methylated small RNAs loaded into the siRNA-specific *Drosophila* AGO2<sup>3</sup>. *In vitro* dissection of the primary structural features required for this high complementary target-induced destabilization was performed by

---

<sup>3</sup>In flies, the RNA methyltransferase HEN1 methylates the 3'-end of AGO2-bound siRNA (Horwich et al., 2007; Pélisson et al., 2007).



introducing various antimirs against *let-7* in *Drosophila* embryo lysate followed by Northern-blotting of variable-length *let-7* species. They actually found out that extensive pairing between a target RNA and the 3'-part of a miRNA is associated with tailing (preferentially of A and U residues) or trimming of its 3'-end, and its decay (Ameres et al., 2010). Baccarini and colleagues verified that perfect or central bulge artificial miRNA targets accelerate the miRNA's rate of decay in mammalian cells in a manner that is dependent on target concentration and complementarity. In accordance with previous results, they also confirm an increase in 3'-end trimmed and uridylylated miRNA species during decay (Baccarini et al., 2011). Later, the Großhans lab interrogated the occurrence of the newly dubbed "*Target-Directed miRNA Degradation*" mechanism, abbreviated TDMD, in rodent neurons – since miRNA turnover is known to be exceptionally fast in neuronal cells (Krol et al., 2010). They confirmed that extensive 3'-end and central bulge pairings induce a decay of cognate miRNAs in rodent primary neurons in association with AGO-loaded miRNA trimming and tailing – with no substantial evidence that they are intermediates of the miRNA decay pathway. Furthermore, they quantified that an individual target molecule can induce the destabilization of more than one miRNA molecule and thus may act as an efficient multiple-turnover TDMD inducer. Interestingly, they observed that TDMD is significantly more efficient in primary neurons than in undifferentiated or differentiated neuroblastoma cells or other cell lines (de la Mata et al., 2015).

Examples of natural TDMD were first reported in mammalian cells infected by viruses. Primate T-cells transformed by *Herpesvirus saimiri* were found to express viral small non-coding RNAs called HSURs (for *H. saimiri* U-rich RNAs), including HSUR1 which is associated with reduction of the miR-27 family levels. Direct base-pairing between HSUR1 transcript and miR-27 through conserved central bulge complementary sites has been shown to direct miR-27 decay (Cazalla et al., 2010), and thus potentially alter host cell gene expression. Similarly, in mouse cells infected with the murine cytomegalovirus (MCMV), the viral coding transcript m169 exhibits a single central bulge complementary site to the orthologous miR-27 required for an equivalent specific miRNA degradation accompanied by 3' tailing and trimming (Libri et al., 2012; Marcinowski et al., 2012). Thus, conservation of host miR-27 decay suggests a functional benefit for these distinct virus families, and indeed, m169-directed decay of the miR-27 family was found to facilitate MCMV replication *in vivo* (Marcinowski et al., 2012). Eventually, a similar viral strategy was also observed in infected cells with the human cytomegalovirus but did not concern miR-27, instead of miR-17 and miR-20a, which are members of the miR-17-92 cluster and belong to the miR-17 family. In the same manner, viral-induced decay of the miR-17 family through TDMD is shown to be associated with higher virus production rates (Lee et al., 2013b).

Recently, the deciphering of TDMD pathway made a huge step forward with the simultaneous report of two endogenous TDMD events *in vivo*, in zebrafish and mice, with demonstrated physiological relevance. From the identification of long intervening non-coding RNAs, abbreviated lincRNAs, in zebrafish *Danio rerio* with conserved genomic locations and required for proper embryonic development including brain morphogenesis, two lincRNAs hold the

attention since they exhibit conserved TDMD-compatible MREs (Ulitsky et al., 2011). The lincRNA called *libra* – for “lincRNA involved in behavioral alterations” – shares orthologous sequences with the 3′-UTR of an ORF-encoding transcript named *Nrep* – for “neuronal-regeneration-related protein” – in mammals (Ulitsky et al., 2011). Both transcripts are brain-specific and involved in behavioral regulation (Taylor et al., 2008; Ulitsky et al., 2011). Interestingly, they contain a deeply conserved near-perfect miRNA-binding site for the miR-29 family (*dre-miR-29a*, *dre-miR-29b* and *mmu-miR-29a*, *mmu-miR-29b*, *mmu-miR-29c*) defined by 3′ and 5′ extensive complementarity with a 3 nt-long central mismatch for miR-29b. The Shkumatava lab discovered that *libra/Nrep* specifically defined the spatial expression of mature miR-29b in the cerebellum by inducing its degradation through TDMD. More importantly, this regulation is conserved between zebrafish and mice and is required for proper behavior in both model organisms (Bitetti et al., 2018). The second lincRNA called *Cyrano* is broadly conserved in vertebrates and expressed in the nervous system during embryonic development, and is enriched in the brain of adult mice. The Bartel lab demonstrated that *Cyrano* contains a highly complementary site to the miR-7 family – 3′ and 5′ extensive complementarity with a 2 nt-long central mismatch for *mmu-miR-7* – which mediates TDMD of the cognate miRNA. In response to miR-7 decay through *Cyrano* interaction, a circular RNA<sup>4</sup> exhibiting numerous MREs to miR-7 and named *Cdr1as*, for Cerebellar degeneration-related protein 1 antisense, can accumulate properly in mice cerebellum (Kleaveland et al., 2018).

Eventually, another endogenous TDMD inducer has been reported by the Nicassio lab – but without any reported physiological function – and will be detailed in section 4 (Ghini et al., 2018).

The correlation between miRNA 3′-end tailing-trimming and full-length degradation was early supposed to result from the independent action of both TENTs and exoribonucleases on the exposed miRNA 3′-end. This assumption was based on results from eubacterial AGO proteins demonstrating that extensive pairing with a target RNA causes guide RNA torsion and 3′-end releasing from the PAZ pocket (Wang et al., 2009).

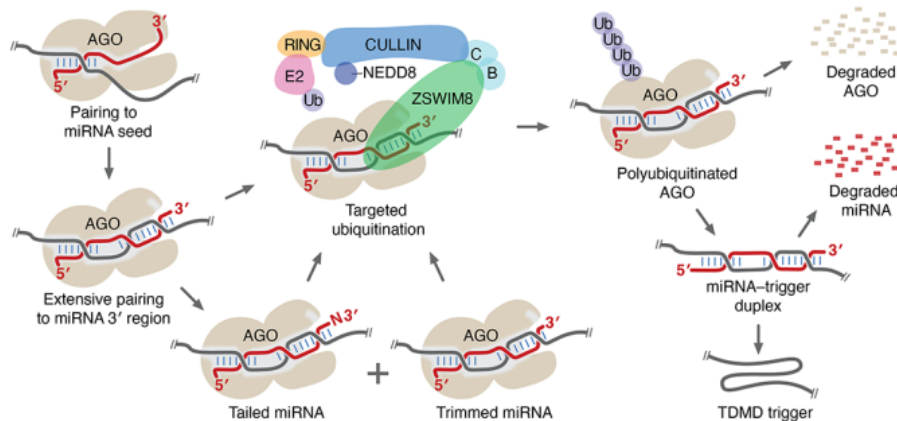
Recently, the MacRae lab reported crystal structures of human AGO2 loaded with miRNA-TDMD target duplexes. They observed that the miRISC complex adopts an unforeseen conformation in which the AGO2 central gate is wide open, and the miRNA and its target are wrapped around each other in a bent duplex partially underwound at its center. Importantly, this structure confirms that, through such pairing, the miRNA 3′-end is released from its PAZ binding pocket (Sheu-Gruttadauria et al., 2019).

Eventually, the Mendell and Bartel labs recently identified concurrently trans-acting factors that mediate miRNA decay through TDMD and thus described this mechanism with more details. They performed genome-wide CRISPR screening in the nearly diploid human cell line HCT-116 or CRISPRi screening in the hematopoietic cell line K562 combined with a reporter gene to miR-7 activity in order to identify proteins involved in *Cyrano*-directed miR-7 decay. Knock-down of *ZSWIM8* and associated proteins of the Cullin-RING E3 ubiquitin ligase

---

<sup>4</sup>Circular RNAs or circRNAs are covalently closed circular transcripts without 5′ caps and 3′ tails, which are mainly formed from pre-mRNAs through exon back-splicing (Memczak et al., 2013).

(CRL) complex were found to impair miR-7 TDMD. Interestingly, the CRL complex has been reported to target specific proteins for ubiquitin-dependent degradation by the 26S proteasome (Enchev et al., 2015). They finally demonstrated, by epistasis study, that ZSWIM8 lies downstream Cyrano in the TDMD pathway, and with the use of 3'-end modified miR-7 oligoribonucleotides, that target-directed miRNA tailing and trimming (TDTT) is not required for target-directed miRNA-degradation (TDMD). Instead, TDTT is likely enabled either by the guide RNA torsion described by the MacRae lab or by its releasing mediated by the CRL complex. Eventually, the TDMD mechanism model proposed by these two studies is the following: AGO structural changes following the pairing of the guide RNA with an unusually complementary target induce the creation of a platform for ZSWIM8 association followed by other CRL partners. Then, CRL-catalyzed ubiquitination of AGO residues triggers AGO degradation by the proteasome, releasing the target RNA – stabilized by cap and its poly(A) tail – and the miRNA – fully exposed to cytoplasmic exoribonucleases (Han et al., 2020; Shi et al., 2020). Additionally, this model characterized in human cell lines may be extended to mammals, flies and worms as Shi et al., 2020 also identified ZSWIM8-sensitive miRNAs displaying short half-lives in various models: human cell lines (HCT-116, HeLa, MCF7, A549); murine cells (NIH-3T3, MEFs, neurons induced from mESC); *Drosophila* S2 cells; adult nematodes.



**FIGURE 2.2: The ZSWIM8/Cullin-RING E3 ubiquitin ligase model of TDMD and its interplay with TDTT.** Following AGO conformation changes due to extensive pairing between the miRNA and its target, ZSWIM8 can interact with AGO. ZSWIM8 is the substrate receptor of the Cullin-RING E3 ubiquitin ligase complex. The recruitment of this complex – adaptor proteins Elongin B (B) and Elongin C (C); Cullin protein; Ubiquitin-conjugating enzyme (E2); and RING-finger protein (RING) – induces ubiquitination of the AGO. Polyubiquitinated AGO is then degraded by the proteasome with release of the intact miRNA:target duplex. From Shi et al., 2020.

## 2.3 Physiological Relevance of microRNA Targets

Since miRNAs are extensively involved in many diverse biological processes, any dysregulation of miRNA expression could lead to defective phenotypes (*cf.* section 2.1). To date, 500 miRNAs have been confidently annotated in humans and both computational and experimental analyses indicate that most human protein-coding genes are potential targets of one or more miRNAs (Friedman et al., 2009). Consequently, functional miRNA analyses aim at identifying genes efficiently targeted by a given miRNA, in other words presumed phenotypically relevant miRNA targets. To this end, most studies rely on computational prediction followed by experimental validation with 3' UTR reporter assay or more relevant genetic invalidation of miRNA:target interaction. In this part, we introduce the different miRNA-mediated gene regulatory effects reported in the literature, as well as the two divergent theories concerning the range of genuine phenotypically relevant miRNA targets.

### 2.3.1 Classes of microRNA-Guided Regulatory Effects

#### Developmental genetic switch

Sometimes corroborated with reciprocal gain-of-function, loss-of-function strategies revealed the importance of some miRNAs for controlling developmental timing in a switching manner. Typically, the deletion of the miRNA, or its target site, prevents the cell from moving on to the next development stage, while ectopic miRNA levels prematurely induce the next state or cause cells to overcommit to this state. As *C. elegans* heterochronic genes, *lin-4* and *let-7* are obvious binary switch miRNAs. Likewise, *Drosophila* mutants lacking *let-7* cluster – including *let-7*, the *lin-4* homolog miR-125, and miR-100 – exhibit temporal misregulation of specific metamorphic processes. The endogenous activation of *let-7* expression during metamorphosis was found to regulate the abundance of the *abrupt* (*ab*) gene and to silence it to nearly undetectable protein levels during this time window (Caygill et al., 2008). The *ab* gene is widely expressed during fly development and is required for diverse functions, including the timing of neuromuscular junctions maturation (Hu et al., 1995). *let-7* therefore acts by switching *ab* expression from ON to OFF and aid the transition from pre-mature muscular cells to mature muscular cells during *Drosophila* metamorphosis.

#### Fine-tuning and noise buffering

To evaluate the regulatory effects of a miRNA in a large-scale manner, quantitative-mass-spectrometry-based approaches using metabolic labeling were performed in mammalian cells after changes in miRNA expression. The influence of ectopic levels (Baek et al., 2008; Selbach et al., 2008) or disruption of specific miRNAs (Baek et al., 2008) on protein output was investigated. These approaches reported that each MRE directs modest repression, usually less than 50% and often less than 20%.

This result is in accordance with another paradigm of miRNA regulation effect: the fine-tuning of a target mRNA or protein level. In this case, the sudden activation – or inactivation

– of the miRNA does not switch off – or on – the target expression. Instead, the miRNA is locally coexpressed with its target and sets a minimal and maximal limit to its expression, according to the cellular state or type. For instance, this mode of action was reported *in vivo* between miR-8 and its functional target, the *atrophin* gene in *Drosophila*. Following the definition presented above, miR-8 and the mRNA *atrophin* are coexpressed in the brain, and both overexpression and reduction of *atrophin* below the levels normally set by miR-8 are detrimental, and cause neurodegeneration or viability defects, respectively (Karres et al., 2007). Thus, miRNAs can buffer target levels to prevent potentially detrimental excess expression while allowing required expression of the target.

Interestingly, comparing the expression of the miRNAs with that of their predicted targets indicates that the conserved mRNA targets tend to be expressed higher in tissue that lacks the cognate miRNA, albeit still present at lower levels in tissues that express the miRNA (Farh et al., 2005; Shkumatava et al., 2009; Sood et al., 2006). Consequently, miRNAs likely act more frequently as rheostats than switches, even during development.

Similarly, it has been postulated that miRNA activity could confer robustness to biological processes against environmental and endogenous perturbations (Ebert et al., 2012; Posadas et al., 2014). For example, miR-7 is required to maintain normal gene expression and sensory organ fate determination under fluctuating temperature conditions in *Drosophila* (Li et al., 2009). Arguments in favor of miRNAs as convincing noise buffers comprised (i) the redundancy inside miRNA families, which could ensure repression in case of localized miRNA failure; (ii) miRNAs are part of extensive feedback and feedforward loops within gene regulatory networks aimed to adapt to gene expression fluctuations; and (iii), mathematical modeling from synthetic systems and single-cell reporter assay data suggest that miRNAs decrease endogenous stochastic fluctuations in gene expression at protein levels for lowly expressed genes (Schmiedel et al., 2015; Siciliano et al., 2013). In this way, miRNAs may buffer the biological noise and reinforce cellular identity by reducing phenotypic variability among individual cells (Raj et al., 2008).

Furthermore, it has also been proposed that protein response to miRNA-mediated silencing depends on target mRNA levels. More precisely, target protein production would be highly repressed below a threshold level of target mRNA abundance – *i.e.*, at low mRNA levels –, but would respond weakly to transcription above this threshold – *i.e.*, at high mRNA levels (Mukherji et al., 2011). Thus, a miRNA can behave as a switch in the target expression regime below the threshold, and as a fine-tuner in the sensitive transition between the threshold and the minimal repression regime at high mRNA levels. This would explain how an individual miRNA could exhibit divergent effects on its different targets and according to the cellular context.

### Neutral interactions

Nevertheless, it appears inevitable that the least dose-sensitive targets can tolerate the miRNA-guided repression. In these cases, it is admitted that the physical interaction between miRISC

and its mRNA targets does induce modest repression of the mRNA levels but is either buffered by homeostatic mechanisms or offset by feedback loops. Such MREs are not supposed to be under selective pressure and should be lost during evolution. Interestingly, miRNA knock-out experiments have shown that many non-conserved sites mediate repression *in vivo* (Baek et al., 2008; Giraldez et al., 2006), which may represent neutral interactions.

Furthermore, miRNA efficiency obviously depends on the miRNA availability. Logically, miRNAs with low miRNA:target ratios confer minimal repression (Arvey et al., 2010; Garcia et al., 2011) and are not biologically relevant. Therefore, the relative cellular concentration of the miRNA and the total target pool are essential to estimate the functionality of a miRNA.

#### 2.3.2 The “Many” and “Few Targets” Hypotheses

As mentioned above, the evaluation of miRNA impact on protein levels moderates the robustness of miRNA silencing. Namely, an individual miRNA can directly repress hundreds of genes, but each to a modest degree around twofold (Baek et al., 2008). Such low-extent miRNA-mediated silencing generally results in inconsequential effects but in some cases could reach biological effects to act as binary switch or rheostat of gene expression. These variant outputs of miRNA-guided silencing may demonstrate the range of miRNA target types or the establishment of complex regulatory networks between miRNAs and their targets. And consequently, it is assumed that a miRNA can act differently in a different cellular context and according to the target transcript.

##### “Many Targets”

The current dogma asserts that miRNAs function by coordinately and modestly regulating a large number of targets to have phenotypic consequences: the “Many targets hypothesis” (Bartel et al., 2004). This assumption is distinctly based on the identification of hundreds of conserved targets per miRNA – and the modest repression mediated by each of them – which suggest that animal fitness is impacted by the precise leveling of the overall proteome. Thus, functional MREs are conserved in a biological purpose and cannot be necessarily revealed by individual genetic disruption. Indeed, they may be part of a complex regulatory network with bifurcating pathways and feedback controls that enable robust response despite a defective edge in the network. Typically, some targets could be functional only in very subtle contexts as intermittent environmental or genetic stress conditions. Therefore, only comparative sequence analysis can help distinguish such functional sites fixed during the evolution.

Furthermore, cooperation between MREs was reported to be mediated by the multivalent binding of TNRC6 with AGO proteins (Briskin et al., 2020). Indeed, multiple MREs in the same 3′-UTR act cooperatively, instead of additively, when close to each other (<40 nt) (Grimson et al., 2007; Saetrom et al., 2007). This finding reinforces the idea that miRNA activity can take advantage of the multiplicity of target sites.

Eventually, in the case of the “many targets” theory, miRNAs produce a complex layer of gene expression regulation in a modular and combinatorial regulatory manner. Precise genetic dissection of highly connected regulatory networks appears particularly difficult to achieve, and could be also complexified by intrinsic robustness. As a consequence, examples of “many targets” would be exceptionally difficult to describe and thus, this theory would be laborious to demonstrate and would stay primarily conceptual.

### “Few Targets”

However, a most straightforward point of view proposes that the vast majority of weakly repressed targets are noises, while few critical targets – named functional or phenotypic targets – can exert phenotypic effects in specific cellular contexts: the “Few targets hypothesis”. For instance, the Rajewsky lab demonstrated with gain- and loss-of-function experiments that the phenotypic effect of miR-150, a miRNA selectively expressed in mature, resting B, and T cells in mice and human, is broadly explained by one functionally critical target in the *in vivo* context: the transcription factor c-Myb (Xiao et al., 2007). It could be assumed that this example results from exceptionally high repression of c-Myc by miR-150, in the same way that binary switch miRNAs. Instead, miR-150 modulates the concentration of C-MYB over a relatively narrow range ( $\sim 30\%$  of protein decrease), indicating that lymphocyte phenotype is exquisitely sensitive to small changes of C-MYB levels. Thus, the Few target hypothesis advantages the regulation of dose-sensitive targets, such as haplo-insufficient genes, and supports the strategy of promoting repression of key targets in place of spreading the repression effect across numerous mRNAs.

Indeed, the dose-sensitivity appears to be crucial to define physiologically relevant miRNA targets. Because gene expression typically fluctuates by 2-fold when comparing two individuals in the human population (Cheung et al., 2003) or two phenotypically identical mouse embryonic neural stem cells (Subkhankulova et al., 2008), the Seitz lab compared inter-individual variability in miR-223 expression among wild-type mice neutrophils to miR-223-guided gene repression quantified from miR-223-deficient mice neutrophils (Pinzón et al., 2017). MiR-223 targets that are not functionally affected should be weakly repressed, *i.e.* at a lower extent than inter-individual variability. They found that most predicted targets of miR-223 in neutrophils fall into this category of miRNA-insensitive targets, raising the concern of the conservation of such sites. To confirm that dose-sensitive genes are more convincing functional miRNA targets, they also compared the conservation scores of MREs predicted in known human haploinsufficient genes (Huang et al., 2010) to MREs in other genes. Results corroborate that haploinsufficient genes tend to bear the most highly conserved miRNA binding sites, leading to the conclusion that miRNA target predictions are contaminated with large numbers of false positives.

Therefore, false positives are conserved pseudotargets for which the miRISC interaction could induce weak molecular repression, likely buffered by homeostatic mechanisms, without inducing efficient repression at the cellular scale. Nevertheless, such sites are found to be

conserved in the evolution and consequently are likely biologically relevant, but in a miRNA-silencing-independent manner.

Competing endogenous RNAs, abbreviated ceRNAs, have been proposed to act as “sponges” to titrate miRNAs away from their other regular targets. In this view, as target pool abundance increases, target mRNAs titrate away miRNAs available for repression. Therefore, ceRNAs have to be sufficiently expressed to supplant the whole target pool, and only activities of low expressed miRNAs can be altered by ceRNAs (Bosson et al., 2014; Denzler et al., 2014). Some molecular titrator candidates have been proposed among pseudogenes (Poliseno et al., 2010) and circRNAs (Hansen et al., 2013) bearing numerous MREs. However, these examples were recently questioned since the results of Poliseno and colleagues were not validated by the Cancer Biology Reproducibility Project (Kerwin et al., 2020), and those of Hansen and colleagues are more likely explained by TDMD of miRNAs instead of titration (Kleaveland et al., 2018; Piwecka et al., 2017). For that matter, TDMD-inducing MREs are convincing pseudotargets with demonstrated biological relevance and broadly conserved during evolution. Eventually, since RNA-binding protein (RBP) have a large number of 3'-UTR-binding sites (Kim et al., 2021), MREs may overlap RBP motifs (Lebedeva et al., 2011; Mukherjee et al., 2011).

Recent genome editing technologies now allow precise dissection of genetic regulatory networks. Especially, the use of CRISPR-Cas9 in *C. elegans* by the Großhans lab demonstrated that *lin-41* is the only target that *let-7* needs to silence to prevent vulval bursting and death (Ecsedi et al., 2015). For this purpose, they studied the mutation of the *let-7* miRNA locus – precisely its seed sequence – which is lethal in nematodes, and verified that restoration of *lin-41* repression – with a compensatory seed match mutation in the *lin-41* locus – rescue that lethality phenotype. Such *in vivo* genetics demonstration permits a rigorous assessment of the contribution of individual putative targets to the *in vivo* phenotype, and it can be expected that it will be used more frequently. The limitation is, however that subtle phenotypes may remain unnoticed or hard to measure precisely.





## Chapter 3

# Functional Relevance of microRNAs on Cellular Phenotypes

One of the interests of this thesis project was the study of phenotypical impact of miRNAs and, primarily, to provide original methods to characterize miRNA targets according to their translation into a macroscopic phenotype. In this chapter, we introduce in the first part a review of canonical miRNA target identification methods, as well as the outline of a novel experimental identification strategy of miRNA targets involved in cell proliferation phenotypes. However, the practical application of this method on miR-34a brought out discrepancies with the known function of this miRNA on cellular growth. Therefore, in the second part, we present our results regarding revising the assumed phenotypic role of miR-34a as a general tumor-suppressor miRNA.

### 3.1 Functional microRNA Targets Identification Methods

In collaboration with my supervisor Hervé Seitz, we wrote a chapter on “Inconsistencies and limitations of current miRNA target prediction methods” for a volume on “microRNA Target Identification”, published in the lab protocol series *Methods in Molecular Biology*, edited by Springer Nature. This work gives an overview of the current methodologies for experimental and computational miRNA target identification. More precisely, we detailed each strategy with its experimental limits and common over- or mis-interpretations. We also discussed the conceptual misconception surrounding the definition of functional or phenotypical miRNA targets. My contribution was the writing of the third chapter – about high-throughput experimental methods – and the review of the whole manuscript.

#### 3.1.1 Review: *Inconsistencies and limitations of current miRNA target identification methods*

# Inconsistencies and limitations of current microRNA target identification methods

Sophie Mockly<sup>1</sup> and Hervé Seitz<sup>1,2</sup>

<sup>1</sup> IGH (CNRS and University of Montpellier), Montpellier, France

<sup>2</sup> Corresponding author; telephone: (+33)434359936; email: herve.seitz@igh.cnrs.fr

## Published in:

MicroRNA Target Identification. Methods in Molecular Biology, vol 1970, p 291-314  
PMID: 30963499; doi: 10.1007/978-1-4939-9207-2\_16

## Abstract

microRNAs and their Argonaute protein partners constitute the RISC complex, which can repress specific target mRNAs. The identification of microRNA targets is of central importance, and various experimental and computational methods have been developed over the last 15 years. Most experimental methods are based on the assumption that mRNAs which interact physically with the RISC complex constitute regulatory targets and, similarly, some computational methods only aim at predicting physical interactors for RISC. Besides specific limitations, which we discuss for each method, the mere concept of assuming a functional role for every detected molecular event is likely to identify many deceptive interactions (*i.e.*: interactions that really exist at the molecular scale, but without controlling any biological function at the macroscopic scale).

In order to select biologically important interactions, some computational tools interrogate the phylogenetic conservation of microRNA/mRNA interactions, thus theoretically selecting only biologically relevant targets. Yet even comparative genomics can yield false positives.

Conceptual and technical limitations for all these techniques tend to be overlooked by the scientific community. This review sums them up, emphasizing on the implications of these issues on our understanding of microRNA biology.

## Key words:

microRNA targets, CLIP, comparative genomics, biological functionality

## 1 Introduction

MicroRNAs (“miRNAs”) are loaded on members of the “Ago” subfamily of the Argonaute protein family, and the resulting ribonucleoprotein complex represses specific target mRNAs. Specificity is due to the miRNA, that base-pairs to the target RNA, while the effector repressive activity is due to the Ago protein, that can repress mRNAs by various mechanisms. If the Ago protein bears an active endonucleolytic site (1, 2) and if the miRNA/mRNA double helix is perfectly paired around the middle of the duplex (3–7), then the target mRNA undergoes endonucleolytic cleavage at a stereotypical position. Otherwise, it is degraded by exonucleases, and translationally repressed — with exonucleolytic decay apparently contributing most of the repressive effect (8–11).

The first miRNA target identification efforts date back to the discovery of miRNAs themselves. It had been known for some time that the *Cænorhabditis elegans lin-4* gene is a repressor of the *lin-14* gene (12). When the functional product of the *lin-4* gene was found to be a small RNA (the first discovered miRNA), it was also realized that the 3′ UTR of *lin-14* contains phylogenetically

conserved regions of imperfect complementarity to the *lin-4* miRNA, which are necessary for *lin-4*-mediated repression (13, 14).

Based on that single example, it was quickly assumed that miRNAs in animals recognize their targets by imperfect base-pairing to 3' UTR's, and when additional miRNA/mRNA interactions were searched, potential targets were screened for imperfectly complementary segments in their 3' UTR's only (15–19). Such a strategy could have led to an observation bias: 5' UTR's and coding sequences could have hosted functional binding sites, which were simply ignored. But in some cases, it was indeed shown that the developmental control of target regulation depends on the 3' UTR (20), suggesting that additional instances of miRNA-mediated regulation indeed require base-pairing to the target's 3' UTR. More recent, unsupervised experimental identification of miRNA binding sites indicated that 3' UTR's are disproportionately frequently bound by miRNAs (21–23). Together with the observation that miRNA binding sites in 3' UTR's repress target expression more efficiently than in other locations (22, 24, 25), these considerations imply that restricting target searches to 3' UTR's may be a good approximation.

Another issue had to be clarified after the initial reports of miRNA/mRNA interactions: the exact definition of a functional “imperfect” complementarity. Data from *Drosophila* genetics (26), from the analysis of sequence conservation within miRNA sequences (27) and from statistical analysis of known miRNA binding sites or conserved putative sites (18, 28, 29), suggested the importance of a perfect pairing between the target and the 5'-most 7 or 8 nucleotides of the miRNA. The concept has then been refined, and it is now clear that the best predictor for miRNA binding is a perfect match to the “seed” (nucleotides 2–7 or 2–8 depending on the authors), or to nucleotides 3–8 (“offset 6mer sites”), of the miRNA (30). Exceptions exist, with 3' UTR's exhibiting perfect seed matches without being repressed (for example, because of steric hindrance by an RNA-binding protein; 31), and with imperfect seed matches mediating target repression (for example, with extensive pairing to the 3' end of the miRNA, compensating for seed mismatches; 4, 16), but for the most part, prediction algorithms favoring seed-matched candidates are in good agreement with experimental data (32).

Various experimental and computational methods have been developed to identify miRNA targets. Each of these methods relies on expected features, more or less explicitly assumed, of miRNA targets:

- their expression level (assessed in terms of mRNA abundance, mRNA occupancy by ribosomes, or protein product abundance) is repressed in a miRNA-dependent fashion;
- loss of their interaction with the miRNA triggers a macroscopic phenotype;
- they interact physically with Ago proteins;
- they exhibit sequence complementarity to the miRNA (e.g., a seed match);
- their interaction with the miRNA is phylogenetically conserved because it is biologically beneficial.

This chapter will review current miRNA target identification methods, with a special emphasis on their conceptual or technical limitations. Interrogating the theoretical grounds underlying each of these techniques raises important questions, which are frequently overlooked, regarding the definition of *biological functionality*.

## 2 Low-throughput experimental methods

### 2.1 In vivo genetics

The discovery of the first miRNA target was made possible by the availability of 3' UTR deletion mutants in *C. elegans*: two gain-of-function mutants of the *lin-14* target turned out to be deleted in

the *lin-14* 3' UTR, suggesting that the UTR contains repressive elements (33). Epistasis experiments involving these UTR deletions, as well as the observation that phylogenetically conserved elements in the 3' UTR are imperfectly complementary to the *lin-4* miRNA, paved the way to the first description of an miRNA recognizing a target 3' UTR by base-pairing (14, 34).

That approach has the huge advantage of interrogating directly the role of a miRNA binding site on macroscopic *in vivo* phenotypes. It has the disadvantage of requiring the availability of viable mutants where the binding sites of interest have been deleted. As the analyzed UTR deletions can be large, there is also a risk that the observed phenotypes are not due to the loss of the identified miRNA binding sites, but to other, unidentified sequence elements in the deletion.

The recent development of genome editing techniques should allow the generation of precise, targeted mutations of miRNA binding sites with greater ease. This technology has already permitted a rigorous assessment of the contribution of individual putative targets to the *in vivo* phenotype controlled by the *let-7* miRNA in worms (35), and it can be expected that it will be used more frequently in the near future for the *in vivo* investigation of the biological function of individual candidate target sites. The obvious limitation of that technique is that subtle phenotypes may remain unnoticed, or hard to measure precisely — hence their occurrence, or their rescue by compensatory miRNA mutations, may be more or less easily scored, possibly resulting in divergent experimental reports by various laboratories.

## 2.2 Quantification of artificial reporter expression in cultured cells

The most popular method for miRNA target assessment has been the quantification of reporter (e.g., luciferase) expression in transfected cultured cells. The 3' UTR of interest being cloned downstream of the reporter coding sequence, such constructs can be transfected in a cell line that expresses the miRNA of interest (or it can be co-transfected with a synthetic miRNA or with a plasmid directing miRNA expression, if the miRNA's expression in that cell line is not thought to be sufficiently high).

miRNA transfection results in an uncontrolled over-expression of miRNAs, which can lead to various types of confounding effects (Figure 1): **(i)** Supra-physiological amounts of transfected small RNAs can titrate components of the endogenous miRNA machinery, thus relieving repression of its mRNA targets (36). **(ii)** Activation of unspecific responses by exogenous small RNAs has been reported in various cellular contexts (37 and references therein). **(iii)** In general, it appears that targets are in large excess relatively to cognate miRNAs (38–40). Such imbalance results in partial occupancy of the targeted mRNAs, even for high-affinity targets when miRNA/(total target) ratios are close to 1 (39). Introduction of additional miRNA copies increases target occupancy, hence achieving higher-than-natural target repression. **(iv)** For a given miRNA, some targets exhibit a higher affinity than others, and binding sites on these high-affinity targets are populated in priority when miRNA levels are low. Increasing miRNA/(total target) ratio will progressively affect novel, low-affinity targets that are not repressed with lower miRNA concentrations (39).

A more convincing experimental verification involves miRNA inhibition, rather than overexpression. Transfection of cultured cells (41, 42) (or intravenous treatment of animals; 43) with chemically stable oligonucleotides efficiently inhibits complementary miRNAs. Instead of measuring the consequences of an increased (potentially supra-physiological) miRNA concentration as in an overexpression experiment, such miRNA inhibition assesses the consequences of decreasing miRNA activity, thus eliminating several biases of overexpression experiments. Yet, undesired, miRNA-independent effects of these inhibitory oligonucleotides have also been reported (44).

Besides issues with miRNA perturbation experiments, detection of reporter expression itself can be problematic. The most widely used reporters are Luciferase proteins, e.g., Firefly luciferase being cloned upstream of the 3' UTR of interest, and *Renilla* luciferase serving as a transfection normalizer. Worrying reports suggest that the measured miRNA-guided repression depends on the transfection protocol (45), on the identity of the promoter driving reporter expression (46), or on various other genic features (47), and such considerations are usually neglected by experimenters. Heterogeneous

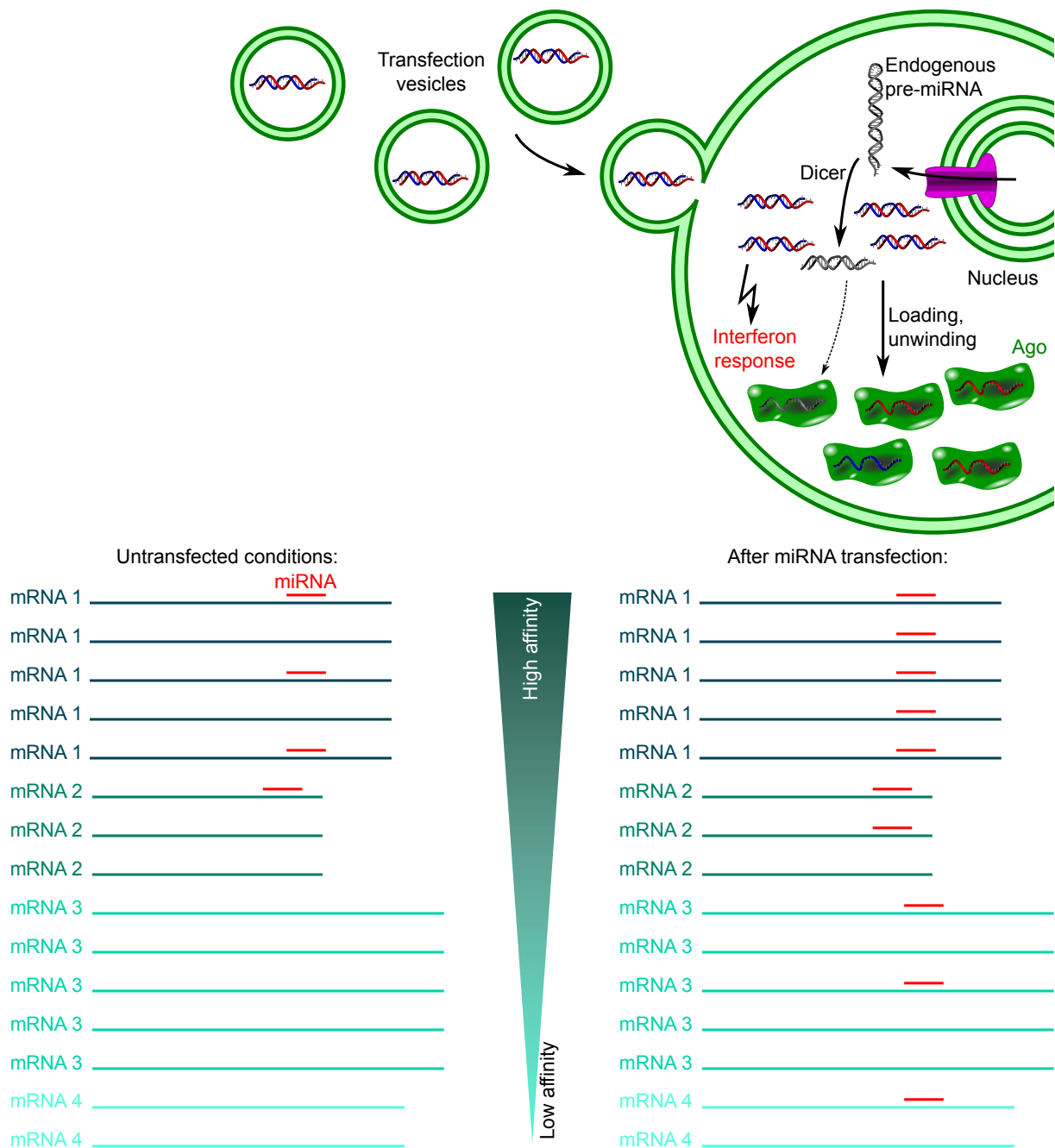


Figure 1: **Confounding effects in miRNA overexpression experiments.** Top panel: The introduction of high, supra-physiological amounts of miRNA in a cell can trigger global responses (most notably, the interferon response) and perturbs endogenous miRNA action by competing for common factors (e.g., Ago proteins). Bottom panel: when miRNA concentration increases, endogenously-repressed targets become more strongly repressed (e.g., mRNAs 1 and 2 here) and novel, artifactually-repressed targets become repressed (e.g., mRNAs 3 and 4 here).

proteolysis could also constitute a source of artifacts. The intracellular half-life of the commonly used Firefly and *Renilla* luciferase ranges from 3 to 4.5 hours (48), which is in the same order of magnitude than the duration of a typical luciferase experiment (including lysis, plate loading, substrate injection and measurement): sample-to-sample variability in proteolysis is conceivable if many samples are to be analyzed together.

## 2.3 Measurement of reporter repression *in vivo*

Precisely manipulating miRNA binding sites was very tedious before the improvement of genome editing methods. Hence it was common practice to use an artificial reporter (whose expression could be easily measured:  $\beta$ -Gal or GFP for example), cloned upstream of the 3' UTR of interest, and transgenically expressed *in vivo* (17, 49). Comparing reporter expression in tissues or animals where the miRNA is mutated or over-expressed, it is then possible to measure the miRNA-guided repression conferred by the UTR in an *in vivo* setting.

Not only the usage of artificial reporters can induce poorly-understood artifacts, as seen in cultured cells (see subsection 2.2), but additional issues arise when using reporter genes *in vivo*. The expression pattern of the reporter may be unrelated to that of the natural target, suggesting functional interaction between RNA molecules that are actually not co-expressed *in vivo* (see for example the overexpressed reporter for the proapoptotic *hid* gene in *Drosophila* imaginal wing discs, while apoptosis is very limited in that organ; 17, 50).

Setting aside such technical problems, it is also important to realize that reporter experiments will only, at best, demonstrate that a particular 3' UTR confers miRNA-mediated repression. It will not test whether such repression translates into a macroscopic phenotype: changes in molecule abundances do not necessarily trigger changes on integrated phenomena, because homeostatic mechanisms may buffer them before they reach the macroscopic scale (see section 3.8). Reporter genes are not subjected to the same feedback loops than the endogenous targets they are supposed to emulate — in fact, reporters have been *purposely designed* to reflect as directly as possible a regulation of interest, uncoupled from any potential feedback loop. Measured fold-changes on reporter expression is thus quantitatively different from that of endogenous genes that belong to interconnected regulatory networks.

## 3 High-throughput experimental methods

### 3.1 Measurement of differential expression after miRNA perturbation

High-throughput identification of miRNA targets was pioneered with the exploitation of proteomics in 2005, using a protocol named “Difference Gel Electrophoresis” (DIGE) associated with Mass Spectrometry (MS). With the assumption that the level of protein products of miRNA-regulated genes would be higher in the absence of mature miRNAs, the proteome from wild-type *Drosophila* oocytes was compared to the proteome of *Dicer-1* deficient oocytes (*Dicer-1* is essential for *Drosophila* miRNA biogenesis). That experiment identified a range of potential miRNA-regulated genes during *Drosophila* oocyte maturation (51). This approach was limited by the sensitivity of protein detection, and it likely captured indirect miRNA targets, whose expression would also vary in the absence of *Dicer-1*. Similarly, transcriptomics and proteomics experiments (8, 52), as well as ribosome profiling experiments (measuring ribosome occupancy transcriptome-wide; 10) have been used to measure gene expression changes after perturbation of miRNA activity in various biological samples, but computational analyses were required for the annotation of direct vs. indirect targets .

In order to identify direct miRNA targets, various methods have then been developed to detect physical interactions between miRNAs and their target RNAs.

### 3.2 HITS-CLIP

High-throughput mapping of protein-RNA interactions *in vivo* became possible with the development of Cross-Linking and Immunoprecipitation (CLIP) strategies (53, 54). The CLIP method is based on the properties of UV irradiation to generate a covalent bond between proteins and ribonucleic acids in close proximity (a few Å apart) (55). Such cross-linked protein-RNA complexes can then be purified by immunoprecipitation using an antibody for the protein of interest, then its RNA partners

are identifiable by sequencing after being trimmed with exonucleases, flanked with adapters, reverse-transcribed and amplified. Recently, this method has been upgraded by including next-generation sequencing to identify *in vivo* protein-RNA interactions in high throughput and was named High-Throughput Sequencing of RNA isolated by CLIP (HITS-CLIP) (56).

Crystallographic study of the Ago-miRNA-mRNA ternary complex structure confirmed that Ago proteins make adequately close protein-RNA contacts to allow Ago HITS-CLIP (57), which can identify both Ago-bound miRNAs and their direct RNA targets (21). In this method, proteins and RNAs are cross-linked by short-wavelength UV irradiation (254 nm) and Ago-RNA complexes are immunoprecipitated in order to purify both Ago-miRNA and Ago-target RNA complexes. Partial RNA digestion is then performed to shorten captured RNAs into fragments that can be sequenced by high-throughput sequencing after partial removal of cross-linked Ago proteins, 5' and 3' adapter ligation, reverse transcription and amplification (Figure 2).

Thereby, the HITS-CLIP method identifies physical interactions between miRNAs and direct targets by an experimental approach, but it does not allow the identification of the miRNA which recruited one particular target: the experiment captures all the target mRNAs which have been recognized by any expressed miRNA. To resolve this issue, computational methods are required to match immuno-precipitated mRNAs to abundant miRNAs in order to find the specific target mRNAs for each miRNA. Thus, standard Ago HITS-CLIP is unable to elude bioinformatic predictions and consequently, it does not provide a genuine experimental identification of target mRNAs for a given miRNA.

Furthermore, technical biases are expected to affect miRNA target identification by HITS-CLIP. It has been shown that UV-induced Ago-RNA cross-linking preferentially occurs at uridines (58), leading to the conclusion that CLIP efficiency varies according to the sequence of miRNA targets, yielding false negatives (32). The extent of such biases is disputed, with other studies suggesting that HITS CLIP-identified miRNA binding sites have little context preference (21, 59).

It is also important to realize that CLIP efficiency is greatly reduced when opaque biological samples are to be analyzed because UV light cannot penetrate thick samples. Hence this technique is intrinsically limited to thin samples, *e.g.*, cell cultures, dissected mouse neocortex, worms (21, 23).

### 3.3 PAR-CLIP

An alternative CLIP-based technique was developed in order to improve the efficiency of the cross-link reaction: Photoactivatable Ribonucleoside-enhanced CLIP (PAR-CLIP) (22). This method involves incorporation of 4-thiouridine (4-SU), a photoactivatable ribonucleoside analogue, into newly synthesized transcripts by supplementing culture media with 4-SU, as first demonstrated *in vivo* with *Escherichia coli* (60) and *ex vivo* with CV-1 cells (61). Transcripts labeled with photoreactive nucleosides are efficiently cross-linked by long-wavelength UV irradiation (365 nm) (Figure 2), achieving higher cross-linking coverage than classic 254 nm-UV cross-linking (100 to 1000-fold improvement in RNA recovery; 22).

An other characteristic of the PAR-CLIP method is that crosslinked 4-SU tends to behave like a C nucleotide during reverse transcription, thus allowing a precise localization of the RNA-protein cross-link after sequencing. The mapping of these mutations on CLIP cDNA has been presented as a major enhancement since it allows to distinguish non-cross-linked noise from formal cross-linked RNA sites, and to increase the resolution of Ago-RNA interaction localization. Nevertheless it was later demonstrated that similar cross-link-induced mutations are also observed with classical 254 nm-UV cross-linking, offering the same advantages to HITS-CLIP (62).

PAR-CLIP suffers from a main technical limitation : the poor rate of 4-SU incorporation. The amount of 4-SU uptaken by a cell in the PAR-CLIP protocol should be adjusted to maximize cross-linking while avoiding cytotoxicity and changes in gene expression due to 4-SU labeling. The PAR-CLIP method therefore involves 12 to 16 hours of cell culture in 100  $\mu$ M 4-SU (a dose that was shown to avoid cytotoxic effects or noticeable changes in gene expression in HEK293 cells; 22).



However, the rate of incorporation determined by HPLC analysis is low: 2.5% (12 hour incubation) (22) or 4% (16 hour incubation) in HEK293 cells and 1 to 4% in other cell lines (63). In practice, and as an example, a typical PAR-CLIP experiment on HeLa cells requires 60 to 100 15 cm Petri dishes (64).

### 3.4 iCLIP

In CLIP-based methods, protease treatment on cross-linked RNA/proteins results in RNA molecules with covalently bound residual protein fragments. Such fragments constitute an obstacle for the reverse-transcriptase: during reverse transcription, the majority of cDNAs are truncated immediately before the cross-linked site (65). Consequently, they are not flanked by both 5' and 3' adapters, and they cannot be amplified by PCR for sequencing. Only the cDNA molecules that could be successfully polymerized despite the cross-linked amino-acids can be sequenced, greatly limiting the efficiency of CLIP library sequencing.

The individual-nucleotide resolution CLIP (iCLIP) approach exploits this limitation to identify cross-linked sites at a nucleotide-resolution (66). The iCLIP protocol differs from HITS-CLIP after the immunoprecipitation of protein-RNA complexes: an adapter is only ligated to the RNA 3' end. Cross-linked proteins are then partially digested and reverse transcription is performed with a primer complementary to the 3' adapter. cDNAs are then circularized and re-linearized by cutting at a site within the primer sequence so as to produce a cDNA with both 5' and 3' adapters. The obtained linear cDNAs are thus amplifiable regardless of where reverse transcription ended (Figure 2).

Comparison of HITS-CLIP and iCLIP libraries showed that 80% of cDNA molecules are truncated at the cross-link site and are thus absent in HITS-CLIP data; the position of the cDNA 3' end indicates precisely the cross-linking site (58). Thereby, the iCLIP approach improves library recovery while providing a single-nucleotide resolution of cross-linked site localization. However, this method requires numerous and diverse enzymatic reactions and purification steps, making the protocol technically challenging. This could explain why iCLIP has not been so heavily used for the identification of Ago-bound mRNAs. To our knowledge, only two laboratories have used this method in published articles : in cultured cells (39) and in *C. elegans* (67, 68).

### 3.5 CLASH

In all the methods described above, the experiment does not identify the miRNA responsible for recruiting a detected target, and further computational analyses are required to try to predict which was the recruiting miRNA. In order to resolve this problem experimentally, a method was developed for the simultaneous identification of the target and the miRNA: Cross-linking, Ligation, And Sequencing of Hybrids (CLASH), a high-throughput method for the characterization of intramolecular and intermolecular RNA-RNA interactions (69), that has been adapted to allow direct miRNA and target identification (70).

In this approach, cells expressing tagged Ago (6×His tag) are irradiated with short-wavelength UV to cross-link protein-RNA complexes. Tagged Ago's are then immunoprecipitated, and Ago-bound RNAs are partially digested by RNases. Ago-RNA complexes are then eluted, immunoprecipitated a second time and RNA 5' ends are phosphorylated. Such RNA modification prepares the keystone step of this protocol: target mRNA-miRNA ligation. Indeed, because of their short length, miRNAs tend not to be trimmed by RNases, so miRNA ends are not enzymatically modified (they keep their 5'-phosphate and 3'-OH) in contrast to the trimmed mRNA ends (5'-OH and 3'-phosphate) (71). Thus, after 5' end phosphorylation, the target mRNA's 5' phosphate and the native miRNA 3'-OH can be enzymatically ligated, assuring the formation of target mRNA-miRNA hybrids with the miRNA sequence positioned 5' to the mRNA fragment. This stage is directly followed by 3' adapter ligation and Ago-RNA complex elution. Ago proteins are then partially digested, 5' adapters

are ligated and finally, target RNA-miRNA hybrids are reverse-transcribed, amplified and sequenced (Figure 2).

However, the efficiency of the RNA-RNA ligation is very low, usually resulting in less than 2% of chimeric reads in all sequenced reads (70). Finally, among tens of millions of reads, only a few tens of thousands correspond to miRNA-mRNA interaction sites, which is far from covering all the actual miRNA interactions and does not enable comparative studies of target profiles. Even in the optimised CLASH protocol, chimeras identifying mRNA targets constitute a fraction of a percent of the total library (Table 1).

It has been reported that regular Ago HITS-CLIP also generates miRNA-target chimeras, probably as a result of an endogenous ligase activity (72), so that specific advantages of the CLASH protocol are not obvious. Further improvements are underway, but the efficiency of the ligation reaction remains problematic (73).

Library:	SRR959756	SRR959757
Reads with barcode:	8,688,014	30,848,119
Reads containing miRNAs:	166,735 (1.92%)	422,865 (1.37%)
Chimeric reads with $\geq$ 15 nt target:	33,364 (0.38%) (5' miRNA: 31,818)	105,831 (0.34%) (5' miRNA: 85,516)
Target does not match rRNAs or tRNAs:	27,401 (0.32%)	75,934 (0.25%)
Target matches RefSeq mRNA:	17,824 (0.21%)	42,819 (0.14%)

**Table 1: Statistics in typical CLASH experiments.** In the original description of the CLASH method, two libraries were prepared with the optimised protocol: SRA accession numbers SRR959756 and SRR959757 (70). “Reads with barcode” are the reads where the 5' barcode CACAGC is found. “Reads containing miRNAs” are the barcoded reads containing miRNA sequences (tolerating up to 3 trimmed nucleotides on the 3' end). “Chimeric reads with  $\geq$  15 nt target” are the reads containing both a miRNA sequence and a fragment of target RNA whose length exceeds 15 nt. The CLASH protocol was designed to generate chimeric reads where the miRNA lies 5' of the target fragment, but some chimeras are ligated backwards (the number of chimeras in the expected orientation is given as “5' miRNA” in the table); note that in the following rows, every chimeric read is considered, regardless of its orientation. “Target does not match rRNAs or tRNAs” are the reads whose  $\geq$  15 nt target fragment does not match nuclear or mitochondrial rRNA or tRNA genes in the sense orientation. “Target matches RefSeq mRNA” are the reads whose  $\geq$  15 nt target fragment matches human sense RefSeq mRNAs without matching sense rRNAs or tRNAs.

### 3.6 Choice of negative controls

Regardless of the CLIP method being used (HITS-CLIP, PAR-CLIP or CLASH), common issues can be avoided if negative controls are carefully chosen. Because of the sensitivity of high-throughput sequencing, RISC-independent background signal originating from the affinity of RNA for antibodies or beads can contaminate the list of identified targets. Repeating CLIP with two different antibodies and comparing the results can help reducing experimental noise (21, 74).

More rigorously, when possible it is advisable to design negative controls which are immunoprecipitated with the same antibodies and beads than the experimental replicates. Such controls include (i) Ago-null cells (or organisms), immunoprecipitated with the same anti-Ago antibody than the wild-type samples (23, 67); (ii) cells expressing wild-type (untagged) Ago, immunoprecipitated

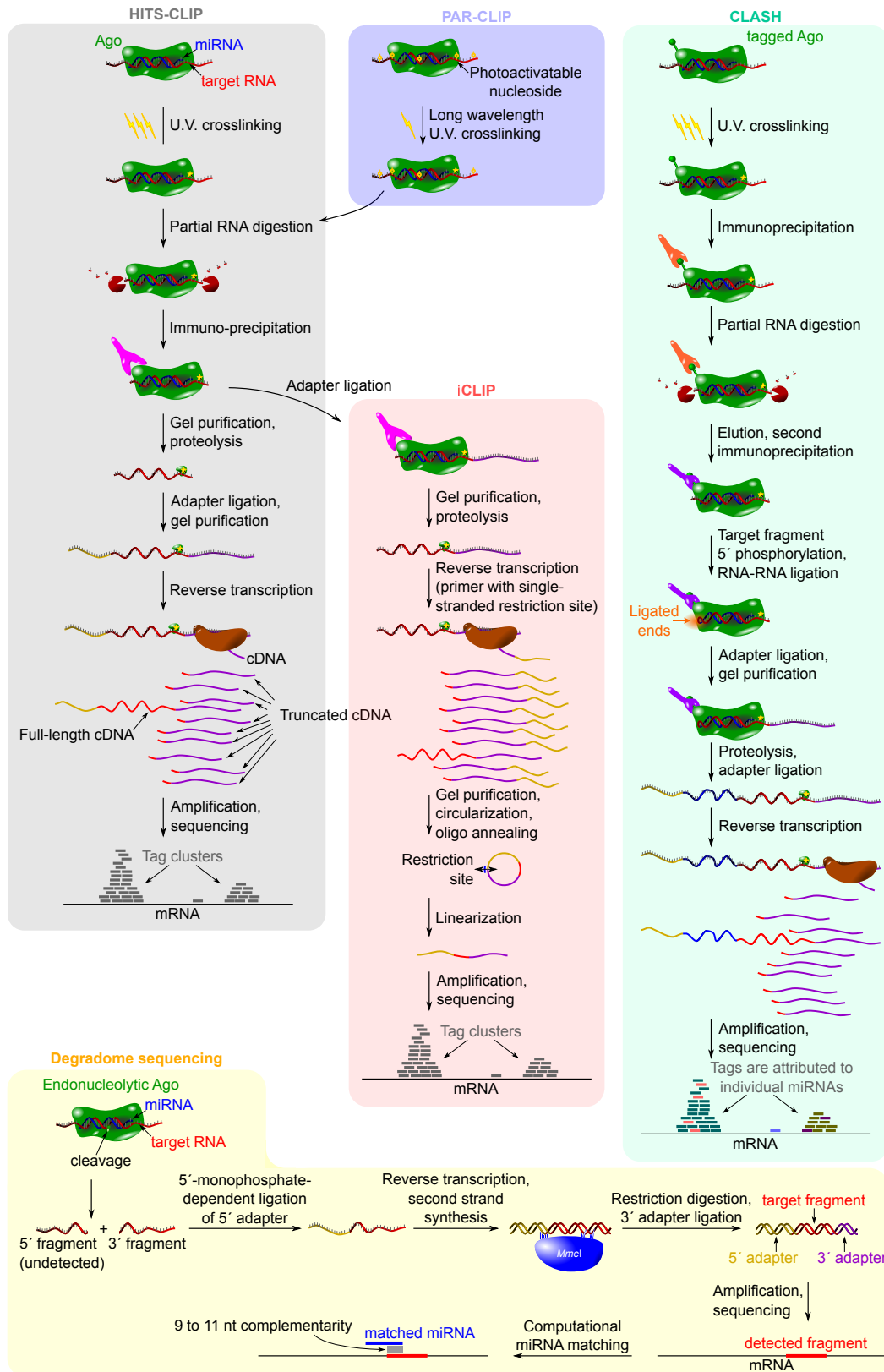


Figure 2: **High-throughput methods for miRNA target identification.** Grey panel: in HITS-CLIP, cross-linked Ago-target complexes are immunoprecipitated, then proteolysed prior to NGS library preparation. Specific steps are modified in PAR-CLIP (blue panel; photoactivatable nucleosides increase cross-linking efficiency) and iCLIP (pink panel; abortive reverse-transcription events are detected). Green panel: the CLASH protocol allows a simultaneous identification of the target and the miRNA. Yellow panel: degradome sequencing identifies 3' fragments of miRNA-guided cleavage.

with the same anti-tag antibody than the cells expressing tagged Ago (39, 75, 76); (iii) cells (or organisms) where a miRNA of interest has been inhibited, immunoprecipitated with the same anti-Ago antibody than the wild-type samples (40, 77, 78).

This latter negative control also provides an experimental way to identify RNA targets for a given miRNA of interest. This technique, termed “differential CLIP” (dCLIP), discerns miRNA-dependent interactions from background interactions. Statistical tests are used to identify RNA targets which are significantly more abundant in Ago immunoprecipitations when the miRNA is active than when it is inhibited (77).

### 3.7 Degradome sequencing

In various eukaryotes, some Ago proteins possess an efficient endonucleolytic activity, that can cleave their RNA targets (1, 2, 79–81). Target cleavage occurs at a stereotypical position (between the nucleotides facing nucleotides 10 and 11 of the small guide RNA), provided that the RNA double helix is regular around the cleavage site. This property implies that the small RNA and the target have to be perfectly complementary around positions 10 and 11 of the guide, in addition to the complementarity requirements for stable RISC/target binding (e.g., a perfect seed match is strongly favored) (3–5, 82, 83). Ago-mediated cleavage produces a 5′-phosphorylated 3′ product and a 3′-hydroxyl 5′ product (84, 85).

Together, these features of RISC-catalyzed target cleavage provide the basis for a method for the detection of endonucleolytically-cleaved miRNA targets: “degradome sequencing” (86) (Figure 2). Endogenous 5′-monophosphorylated RNAs (including the 3′ products of RISC-mediated cleavage) are ligated to a 5′ adapter containing a site for the *MmeI* restriction enzyme (*MmeI* cleaves ≈20 bp downstream of its recognition sequence). Following reverse-transcription, second strand DNA synthesis and digestion with the restriction enzyme, double-stranded DNA products are ligated to a 3′ adapter, then deep-sequenced. This protocol enriches efficiently for RISC cleavage products: neither 5′-capped, nor 5′ polyphosphorylated RNAs (e.g., primary transcripts), nor 5′ hydroxyl RNAs (which are commonly produced by many types of endogenous RNases; see 87 and references therein) are detected. To distinguish miRNA target fragments from other cellular 5′ monophosphorylated RNAs, sequences of the detected fragments are then mapped to the transcriptome, and the presumptive target site is compared to the sequences of known miRNAs: if the presumptive target is complementary enough to a miRNA to allow endonucleolytic cleavage, and if the 5′ nucleotide of the detected fragment faces the 10<sup>th</sup> nucleotide of the miRNA (+/- 1 nt, to account for observed heterogeneities in miRNA 5′ ends and cleavage position), then the detected fragment is assumed to derive from miRNA-guided cleavage of the target.

Degradome sequencing, when applied to mammalian samples, reveals a class of targets which undergo miRNA-guided endonucleolytic cleavage (6, 7). Many of these targets do not exhibit a full seed match to the miRNA, probably explaining why, for a given target RNA, that repressive mode is very unfrequent (currently, only two seed-matched sites allowing target cleavage are known in mammals; 5, 88, 89). Even when tolerating seed mismatches, miRNA binding sites allowing cleavage are very rare in mammalian transcriptomes (7, 89).

Because degradome sequencing is restricted to the identification of cleaved targets, it is fundamentally limited to a minor subset of miRNA targets, at least in mammals. It uncovers more targets in other clades, where miRNA-guided cleavage is more widespread (86, 90, 91).

It is also important to realize that this protocol only detects cleaved RNA fragments, and it fails to quantify how frequently such cleavage occurs. If a given mRNA is rarely cleaved (say, less than 1 molecule out of 1,000), such repression is very likely to be functionally inconsequential — yet it will be detected in a degradome sequencing experiment, provided that the mRNA is abundant enough.

### 3.8 The issue with molecular assays

Low-throughput as well as high-throughput molecular assays face a fundamental limitation: they may identify molecular interactions that really take place *in vivo*, but the mere existence of these interactions does not imply that they have a biological function. Molecular events may be real, but without triggering any beneficial macroscopic consequences. In other words, the identified targets may be real from a molecular point of view, without being *functionally* targeted: their repression by the miRNA does not control any macroscopic phenotypes, that could be subjected to natural selection (40).

It would be misleading to assume that every observed molecular event necessarily translates into a biological phenotype. Luckily, with the development of genome-editing techniques, it is becoming easier to probe the functional relevance of molecular interactions. miRNA/target interactions that had been very clearly established using molecular biology may very well fail to control any obvious phenotype (compare for example 92 with 93).

## 4 Computational methods

### 4.1 Basic principles

miRNA target prediction programs rely on the established molecular rules for RISC/target binding. Because the miRNA complement is probably well known in most model organisms, applying these rules to transcriptome sequences should identify all the possible binding sites for miRNAs. But several complications make the task non-trivial:

1. Initially, the molecular determinants for RISC binding were very imprecisely known. The first identified miRNA/target pair (the *lin-4/lin-14* pair in *Cænorhabditis elegans*; 14) showed that the complementarity does not have to be perfect — and the definition of an *imperfect yet functional* complementarity remained vague for several years.
2. Prediction accuracy has to be evaluated, which means that a control list of established, faithful miRNA binding sites has to be available as well as, ideally, a negative control set, which is not repressed by the miRNAs of interest.
3. Even if a given mRNA has the intrinsic ability to bind a miRNA-programmed RISC with a high affinity, it may not actually bind it *in vivo* (the site may be hidden by a bound protein, or the miRNA and the mRNA may not be expressed in the same cells).

Because of point #1, early target prediction tools had to rely on disputable assumptions. The first published prediction program was optimized using its own prediction of the *bantam/hid* miRNA/target pair in *Drosophila*, creating a circular reasoning (compare 28 with 17). The first developed programs also ranked candidate targets by their predicted affinity to RISC, which was approximated by the estimated free folding energy of the target with the protein-free miRNA sequence (18, 19, 28). But further investigations revealed that GC-rich target contexts are actually less efficient than AU-rich contexts, despite being predicted to bind more tightly to the miRNA sequence in the absence of protein partners (24).

Based on the few known targets at the time, each of these three initial computational programs (18, 19, 28) also tended to favor binding sites with a good complementarity to the 5′-most sequence of the miRNA. The importance of that subsequence (the “seed” of the miRNA) was indeed confirmed later (reviewed in 30), and it is now used by most modern prediction programs.

Adopting a different strategy, some authors attempted to explore the efficiency of various imperfect miRNA/target duplexes in order to derive an experimentally-validated set of computational rules (94). Such an effort is inevitably hindered by the infinite number of possible mismatches and bulges in an RNA duplex.

To address point #2, it is natural to compare the list of predicted targets to experimentally-validated miRNA targets, and the power of such an approach increases when more targets are experimentally confirmed (compare 19, 28 with 8, 52). Alternatively, it was proposed that a carefully controlled set of shuffled miRNA sequences could provide a negative control to estimate signal/noise ratios in target prediction (18). That approach has to assume that every predicted binding site for such artificial, shuffled miRNA sequences is inefficient, which remains to be demonstrated (it is thus likely to under-estimate the signal/noise ratio).

Many different metrics could estimate prediction sensitivity and specificity, and they may not agree with each other. For instance, the program named “rna22” does not appear to predict miRNA-guided mRNA repression better than simple random picking (32), even though its performance seemed to be carefully controlled when the algorithm was published (95).

Point #3 can be addressed computationally, using comparative genomics: functionally important sequence elements are assumed to be conserved in evolution. Because 3′ UTR's tend to be poorly conserved overall, any conserved sequence in a 3′ UTR (e.g., a miRNA binding site) will be particularly obvious and easily detectable. Selection of phylogenetically conserved matches to miRNA seed sequences is thus a common strategy in target prediction programs (96). Phylogenetic conservation of a miRNA binding site is seen as a proof that the miRNA and the target *do interact* (implying that they are co-expressed in some cells, and that the target site is accessible for miRNA hybridization), therefore resolving the issue raised in point #3.

Importantly, phylogenetic conservation of an miRNA seed match is not only perceived as a proof of an RNA/RNA physical interaction: it is also interpreted as an indication of the functional importance of the interaction (30). One major limitation of molecular assays is that they cannot demonstrate the biological relevance of miRNA/target binding (see subsection 3.8): this obstacle is overcome if there is evidence of a selective pressure on the interaction.

## 4.2 Refinements

### 4.2.1 Predictors of molecular interaction

While most target prediction programs tend to look for phylogenetically conserved miRNA seed matches (with various definitions and scoring systems for the notions of “match” and “conservation”), they differ in the way they incorporate additional information. The program using most additional information is named “TargetScan”. It benefits from a long, sustained development, with incremental improvements through the incorporation of successive predictors and with an optimized weighting system to combine their contributions (see 32 for the latest version to date).

The most efficient features that TargetScan uses to predict miRNA binding include, among others:

- local AU content (efficient binding sites tend to be AU-rich);
- position of the site within the 3′ UTR (with efficient binding sites avoiding the middle of long UTR's);
- total number of predicted binding sites in every 3′ UTR (miRNAs with many possible binding sites tend to be titrated by the transcriptome, hence they are less efficient on each individual site).

Whenever that information is available, it is also useful to consider 3′ UTR polymorphism. Specific UTR isoforms can be expressed in a cell type-specific manner, thus modulating target responsiveness to miRNAs (97).

### 4.2.2 Phylogenetic conservation

Comparative genomics has become an important component of most target prediction programs. With the multiplication of genome sequencing projects, pre-computed whole genome alignments involving tens of species are now available, facilitating the identification of conserved UTR segments (98).

Because phylogenetic conservation cannot help identifying targets for poorly-conserved miRNAs (e.g., clade-specific miRNAs and viral miRNAs), a few computational programs were developed without the help of comparative genomics (99–102), or where target site conservation does not contribute much (103). When assessed against experimental data, these programs tend to perform less well than conservation-based predictors (32). Also note that, in some cases (for poorly-conserved miRNAs), the program TargetScan makes no use of phylogenetic conservation to predict miRNA targets, and it only uses context features (local AU richness, ...).

In theory, it should be possible to predict miRNA targets without the need for comparative genomics: interacting molecules obviously do not have any notion of whether they are conserved in other species and yet, by definition they only bind their real interactors. Phylogenetic conservation is used by computational programs to compensate for our lack of information on the molecular determinants of the interaction (and to estimate the physiological impact of these molecular events). It is thus possible that, in the future, refinements of computational predictions with enough molecular predictors will make phylogenetic conservation dispensable for an accurate prediction of the mRNA interactors of a given miRNA, perhaps even for an accurate prediction of the physiological functionality of individual miRNA/mRNA interactions.

### 4.3 False positives in comparative genomics

Conserved genomic elements may be conserved for miRNA-independent reasons, e.g., they could constitute protein binding sites on the DNA locus or the RNA transcript. The odds are high that they could be complementary to a miRNA seed by chance: according to the current version of the miRBase repository (version 21, dated June 2014; 104), 1,508 different seeds (defined as nt 2 – 7) can be found in the 2,588 listed human miRNAs, and 1,228 different seeds can be found in the 1,915 listed murine miRNAs. Because there are only  $4^6 = 4096$  possible hexanucleotides, a large fraction of all the possible  $\geq 6$  nt-long motifs matches an miRNA seed. It is thus expected that many conserved genomic elements are complementary to miRNA seeds, even if their phylogenetic conservation is not due to the miRNA.

Two types of methods have been proposed to estimate the number of such false positives. Considering short sequence motifs, excluding those that are complementary to miRNA seeds (but selecting those motifs that behave similarly to miRNA seed matches, in terms of sequence composition features and other genomic features), then measuring their phylogenetic conservation, it was proposed that mammalian miRNAs exert a selective pressure on most coding genes (105). According to that analysis, more than 60% of human coding genes are miRNA targets.

Alternatively, the number of false positives could be estimated by scoring the number of conserved miRNA seed matches in phylogenetic clades where the miRNA does not exist: seed matches which are more conserved than their cognate miRNA itself are probably conserved for a miRNA-independent reason (40). Here, miRNA-independent conservation is measured on the seed-matching motifs themselves; a drawback of that method is that deeply conserved miRNAs are hard to analyze (there are not so many sequenced genomes where that miRNA is absent), probably under-estimating false positive rates for the most conserved miRNAs (40, 106). According to that analysis, several tens of percent of computationally-predicted targets are false positives: they are phylogenetically conserved for miRNA-independent reasons, and they are only fortuitously complementary to miRNA seeds.

Seed matches to poorly-conserved miRNAs (e.g., Hominidae-specific miRNAs) appear to be very

frequently more conserved than their cognate miRNAs (40). Yet some computational programs consider phylogenetic conservation of their seed matches as an evidence of biological functionality (e.g., “miRanda” and “DIANA-microT”; 107, 108). Another program, TargetScan, does not use phylogenetic conservation as far as poorly-conserved miRNAs are concerned (see subsection 4.2.2): it only attempts to predict physical interactions between the miRNA and the mRNA, and thus cannot predict physiological relevance (see subsection 3.8). As for the “PicTar” prediction program, it simply does not attempt to predict targets for the least conserved miRNAs.

## 5 Conclusion

Various experimental and computational methods are currently used to identify miRNA targets. Most of them tend to exhibit both technical and conceptual limitations, which are frequently ignored. Some of these caveats can affect profoundly the interpretation of many published studies.

Low-throughput reporter assays are commonly used for the validation of miRNA/target regulatory interactions. These tests frequently involve miRNA overexpression, which is particularly prone to the generation of artifactual responses. Moreover, and because the repressive action of miRNAs is usually very modest, the amplitude of experimental noise frequently falls in the same order of magnitude than the biological effect to be measured. More generally, reporter genes face a fundamental weakness: they cannot recapitulate complex genetic interactions (regulatory feedbacks, indirect phenotypic effects), implying that their repression by miRNAs is quantitatively different from that of endogenous genes.

Yet high-throughput methods for the quantification of miRNA-guided repression of endogenous genes are themselves polluted by numerous biases. The interpretation of such experimental results can also be affected by the quality of computational analyses. For example, several studies have suggested the existence of novel classes of non-canonical miRNA binding sites, which were believed to be functional. Yet a detailed re-analysis of these cases shows that the observed repressive effect seems to be due to neighboring, canonical miRNA binding sites (32).

It is also important to keep in mind that every physical interaction between RISC and an mRNA may not trigger functional consequences. It is therefore wrong to equate “physical interactors” with “targets”. Computational analyses could, in principle, assess the biological importance of individual miRNA/target interactions by measuring their phylogenetic conservation. But these methods are themselves contaminated with false positives. The large number of available computational prediction programs also contributes to the inflation of false positives, because each individual interaction is more likely to be predicted by a program if more programs are available.

Finally, the most convincing targets identified so far tend to be the ones that were identified before the development of most of these techniques (compare for example 35 with 16; also see 14). This could actually be very meaningful: the initial studies were based on genetics, where the observation of an *in vivo* phenotype is a prerequisite. The availability of many experimental or computational methods for high-throughput miRNA target identification could have promoted a new vision of regulatory interactions, where molecular responses were considered as biological phenotypes. Considering that these interactions constitute actual functional regulation is certainly risky if not confirmed *in vivo*. In the future, targeted mutation of miRNA binding sites *in vivo* will probably be a great help for the functional validation of candidate targets (35, 93).

## Computational code accessibility

All the instructions, scripts and intermediary data used for the preparation of Table 1 can be downloaded from [https://www.igh.cnrs.fr/images/microsite/herve-seitz/files/Mockly\\_and\\_Seitz\\_2018\\_Table1.tar.bz2](https://www.igh.cnrs.fr/images/microsite/herve-seitz/files/Mockly_and_Seitz_2018_Table1.tar.bz2) or from [https://github.com/HKeyHKey/Mockly\\_and\\_Seitz\\_2018](https://github.com/HKeyHKey/Mockly_and_Seitz_2018).



## Acknowledgements

The authors wish to thank Dr. Isabelle Busseau and Dr. Séverine Chambeyron for critical reading of the manuscript.

## References

- [1] Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi *Science* **305**, 1437–1441.
- [2] Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, and Tuschl T (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs *Mol Cell* **15**, 185–197.
- [3] Hutvagner G and Zamore PD (2002) A microRNA in a multiple-turnover RNAi enzyme complex *Science* **297**, 2056–2060.
- [4] Yekta S, Shih Ih, and Bartel DP (2004) MicroRNA-directed cleavage of *HOXB8* mRNA *Science* **304**, 594–596.
- [5] Davis E, Caiment F, Tordoir X, Cavaillé J, Ferguson-Smith A, Cockett N, et al. (2005) RNAi-mediated allelic trans-interaction at the imprinted *Rtl1/Peg11* locus *Curr Biol* **15**, 743–749.
- [6] Karginov FV, Cheloufi S, Chong MM, Stark A, Smith AD, and Hannon GJ (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases *Mol Cell* **38**, 781–788.
- [7] Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, and Bartel DP (2010) Expanding the microRNA targeting code: functional sites with centered pairing *Mol Cell* **38**, 789–802.
- [8] Baek D, Villén J, Shin C, Camargo FD, Gygi SP, and Bartel DP (2008) The impact of microRNAs on protein output *Nature* **455**, 64–71.
- [9] Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, et al. (2009) Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA *PLoS Biol* **7**, e1000238.
- [10] Guo H, Ingolia NT, Weissman JS, and Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels *Nature* **466**, 835–840.
- [11] Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, et al. (2014) mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues *Mol Cell* **56**, 104–115.
- [12] Ambros V (1989) A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans* *Cell* **57**, 49–57.
- [13] Lee RC, Feinbaum RL, and Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14* *Cell* **75**, 843–854.
- [14] Wightman B, Ha I, and Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans* *Cell* **75**, 855–862.
- [15] Moss EG, Lee RC, and Ambros V (1997) The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the *lin-4* RNA *Cell* **88**, 637–646.
- [16] Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans* *Nature* **403**, 901–906.
- [17] Brennecke J, Hipfner DR, Stark A, Russell RB, and Cohen SM (2003) *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila* *Cell* **113**, 25–36.

- [18] Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, and Burge CB (2003) Prediction of mammalian microRNA targets *Cell* **115**, 787–798.
- [19] Enright AJ, John B, Gaul U, Tuschl T, Sander C, and Marks DS (2003) MicroRNA targets in *Drosophila Genome Biol* **5**, R1.
- [20] Lin SY, Johnson SM, Abraham M, Vella MC, Pasquinelli A, Gamberi C, et al. (2003) The *C. elegans hunchback* homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target *Dev Cell* **4**, 639–650.
- [21] Chi SW, Zang JB, Mele A, and Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps *Nature* **460**, 479–486.
- [22] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP *Cell* **141**, 129–141.
- [23] Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, et al. (2010) Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans* *Nat Struct Mol Biol* **17**, 173–179.
- [24] Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, and Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing *Mol Cell* **27**, 91–105.
- [25] Gu S, Jin L, Zhang F, Sarnow P, and Kay MA (2009) Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs *Nat Struct Mol Biol* **16**, 144–150.
- [26] Lai EC (2002) MicroRNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation *Nat Genet* **30**, 363–364.
- [27] Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, et al. (2003) The microRNAs of *Caenorhabditis elegans* *Genes Dev* **17**, 991–1008.
- [28] Stark A, Brennecke J, Russell RB, and Cohen SM (2003) Identification of *Drosophila* microRNA targets *PLoS Biol* **1**, E60.
- [29] Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. (2005) Combinatorial microRNA target predictions *Nat Genet* **37**, 495–500.
- [30] Bartel DP (2009) MicroRNAs: target recognition and regulatory functions *Cell* **136**, 215–233.
- [31] Kedde M, Strasser MJ, Boldajipour B, Oude JA Vrieling, Slanchev K, le Sage C, et al. (2007) RNA-binding protein Dnd1 inhibits microRNA access to target mRNA *Cell* **131**, 1273–1286.
- [32] Agarwal V, Bell GW, Nam JW, and Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs *Elife* **4**, doi: 10.7554/eLife.05005.
- [33] Wightman B, Burglin TR, Gatto J, Arasu P, and Ruvkun G (1991) Negative regulatory sequences in the *lin-14* 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development *Genes Dev* **5**, 1813–1824.
- [34] Arasu P, Wightman B, and Ruvkun G (1991) Temporal regulation of *lin-14* by the antagonistic action of two other heterochronic genes, *lin-4* and *lin-28* *Genes Dev* **5**, 1825–1833.
- [35] Ecsedi M, Rausch M, and Großhans H (2015) The *let-7* microRNA directs vulval development through a single target *Dev Cell* **32**, 335–344.
- [36] Khan AA, Betel D, Miller ML, Sander C, Leslie CS, and Marks DS (2009) Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs *Nat Biotechnol* **27**, 549–555.

- [37] Sioud M (2005) Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization *J Mol Biol* **348**, 1079–1090.
- [38] Denzler R, Agarwal V, Stefano J, Bartel DP, and Stoffel M (2014) Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance *Mol Cell* **54**, 766–776.
- [39] Bosson A, Zamudio J, and Sharp P (2014) Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition *Molecular Cell* **56**, 347–359.
- [40] Pinzón N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, et al. (2017) microRNA target prediction programs predict many false positives *Genome Res* **27**, 234–245.
- [41] Hutvágner G, Simard MJ, Mello CC, and Zamore PD (2004) Sequence-specific inhibition of small RNA function *PLoS Biol* **2**, E98.
- [42] Meister G, Landthaler M, Dorsett Y, and Tuschl T (2004) Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing *RNA* **10**, 544–550.
- [43] Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, et al. (2005) Silencing of microRNAs in vivo with 'antagomirs' *Nature* **438**, 685–689.
- [44] Sarvestani ST, Stunden HJ, Behlke MA, Forster SC, McCoy CE, Tate MD, et al. (2015) Sequence-dependent off-target inhibition of TLR7/8 sensing by synthetic microRNA inhibitors *Nucleic Acids Res* **43**, 1177–1188.
- [45] Lytle JR, Yario TA, and Steitz JA (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR *Proc Natl Acad Sci USA* **104**, 9667–9672.
- [46] Kong YW, Cannell IG, de Moor CH, Hill K, Garside PG, Hamilton TL, et al. (2008) The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene *Proc Natl Acad Sci USA* **105**, 8866–8871.
- [47] Cottrell KA, Szczesny P, and Djuranovic S (2017) Translation efficiency is a determinant of the magnitude of miRNA-mediated repression *Sci Rep* **7**, 14884.
- [48] Thorne N, Inglese J, and Auld DS (2010) Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology *Chem Biol* **17**, 646–657.
- [49] Johnston RJ and Hobert O (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans* *Nature* **426**, 845–849.
- [50] Milán M, Campuzano S, and García-Bellido A (1997) Developmental parameters of cell death in the wing disc of *Drosophila* *Proc Natl Acad Sci USA* **94**, 5691–5696.
- [51] Nakahara K, Kim K, Sciulli C, Dowd SR, Minden JS, and Carthew RW (2005) Targets of microRNA regulation in the *Drosophila* oocyte proteome *Proc Natl Acad Sci USA* **102**, 12023–12028.
- [52] Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, and Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs *Nature* **455**, 58–63.
- [53] Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, and Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain *Science* **302**, 1212–1215.
- [54] Ule J, Jensen K, Mele A, and Darnell RB (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells *Methods* **37**, 376–386.
- [55] Greenberg JR (1979) Ultraviolet light-induced crosslinking of mRNA to proteins *Nucleic Acids Res* **6**, 715–732.
- [56] Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing *Nature* **456**, 464–469.

- [57] Schirle NT, Sheu-Gruttadauria J, and MacRae IJ (2014) Structural basis for microRNA targeting *Science* **346**, 608–613.
- [58] Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, et al. (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions *Genome Biol* **13**, R67.
- [59] Darnell RB (2010) HITS-CLIP: panoramic views of protein-RNA regulation in living cells *Wiley Interdiscip Rev RNA* **1**, 266–286.
- [60] Favre A, Bezerra R, Hajnsdorf E, Lemaigre Y Dubreuil, and Expert-Bezan A con (1986) Substitution of uridine in vivo by the intrinsic photoactivable probe 4-thiouridine in *Escherichia coli* RNA. Its use for *E. coli* ribosome structural analysis *Eur J Biochem* **160**, 441–449.
- [61] Favre A, Moreno G, Blondel MO, Kliber J, Vinzens F, and Salet C (1986) 4-Thiouridine photosensitized RNA-protein crosslinking in mammalian cells *Biochem Biophys Res Commun* **141**, 847–854.
- [62] Zhang C and Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data *Nat Biotechnol* **29**, 607–614.
- [63] Ascano M, Hafner M, Cekan P, Gerstberger S, and Tuschl T (2012) Identification of RNA-protein interaction networks using PAR-CLIP *Wiley Interdiscip Rev RNA* **3**, 159–177.
- [64] Lebedeva S, Jens M, Theil K, Schwanhäusser B, Selbach M, Landthaler M, et al. (2011) Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR *Mol Cell* **43**, 340–352.
- [65] Urlaub H, Hartmuth K, and Lührmann R (2002) A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles *Methods* **26**, 170–181.
- [66] König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, et al. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution *Nat Struct Mol Biol* **17**, 909–915.
- [67] Broughton JP and Pasquinelli AE (2013) Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP *Methods* **63**, 119–125.
- [68] Broughton JP, Lovci MT, Huang JL, Yeo GW, and Pasquinelli AE (2016) Pairing beyond the Seed Supports MicroRNA Targeting Specificity *Mol Cell* **64**, 320–333.
- [69] Kudla G, Granneman S, Hahn D, Beggs JD, and Tollervey D (2011) Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast *Proc Natl Acad Sci USA* **108**, 10010–10015.
- [70] Helwak A, Kudla G, Dudnakova T, and Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding *Cell* **153**, 654–665.
- [71] Helwak A and Tollervey D (2014) Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH) *Nat Protoc* **9**, 711–728.
- [72] Grosswendt S, Filipchuk A, Manzano M, Klironomos F, Schilling M, Herzog M, et al. (2014) Unambiguous identification of miRNA:target site interactions by different types of ligation reactions *Mol Cell* **54**, 1042–1054.
- [73] Moore MJ, Scheel TK, Luna JM, Park CY, Fak JJ, Nishiuchi E, et al. (2015) miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity *Nat Commun* **6**, 8864.
- [74] Guo YE, Oei T, and Steitz JA (2015) Herpesvirus saimiri MicroRNAs Preferentially Target Host Cell Cycle Regulators *J Virol* **89**, 10901–10911.
- [75] Stefani G, Chen X, Zhao H, and Slack FJ (2015) A novel mechanism of LIN-28 regulation of let-7 microRNA expression revealed by in vivo HITS-CLIP in *C. elegans* *RNA* **21**, 985–996.

- [76] Eckenfelder A, Ségéral E, Pinzón N, Ulveling D, Amadori C, Charpentier M, et al. (2017) Argonaute proteins regulate HIV-1 multiply spliced RNA and viral production in a Dicer independent manner *Nucleic Acids Res* **45**, 4158–4173.
- [77] Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, et al. (2012) Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting *Mol Cell* **48**, 760–770.
- [78] Luna JM, Barajas JM, Teng KY, Sun HL, Moore MJ, Rice CM, et al. (2017) Argonaute CLIP defines a deregulated miR-122-bound transcriptome that correlates with patient survival in human liver cancer *Mol Cell* **67**, 400–410.
- [79] Okamura K, Ishizuka A, Siomi H, and Siomi MC (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways *Genes Dev* **18**, 1655–1666.
- [80] Baumberger N and Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs *Proc Natl Acad Sci USA* **102**, 11928–11933.
- [81] Irvine DV, Zaratiegui M, Tolia NH, Goto DB, Chitwood DH, Vaughn MW, et al. (2006) Argonaute slicing is required for heterochromatic silencing and spreading *Science* **313**, 1134–1137.
- [82] Wee LM, Flores-Jasso CF, Salomon WE, and Zamore PD (2012) Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties *Cell* **151**, 1055–1067.
- [83] Salomon WE, Jolly SM, Moore MJ, Zamore PD, and Serebrov V (2015) Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides *Cell* **162**, 84–95.
- [84] Llave C, Xie Z, Kasschau KD, and Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA *Science* **297**, 2053–2056.
- [85] Martinez J and Tuschl T (2004) RISC is a 5′ phosphomonoester-producing RNA endonuclease *Genes Dev* **18**, 975–980.
- [86] Addo-Quaye C, Eshoo TW, Bartel DP, and Axtell MJ (2008) Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome *Curr Biol* **18**, 758–762.
- [87] Peach SE, York K, and Hesselberth JR (2015) Global analysis of RNA cleavage by 5′-hydroxyl RNA sequencing *Nucleic Acids Res* **43**, e108.
- [88] Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, et al. (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene *Nat Genet* **34**, 261–262.
- [89] Bracken CP, Szubert JM, Mercer TR, Dinger ME, Thomson DW, Mattick JS, et al. (2011) Global analysis of the mammalian RNA degradome reveals widespread miRNA-dependent and miRNA-independent endonucleolytic cleavage *Nucleic Acids Res* **39**, 5658–5668.
- [90] Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R, and Axtell MJ (2009) Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome *RNA* **15**, 2112–2121.
- [91] Moran Y, Fredman D, Praher D, Li XZ, Wee LM, Rentzsch F, et al. (2014) Cnidarian microRNAs frequently regulate targets by cleavage *Genome Res* **24**, 651–663.
- [92] Becam I, Rafel N, Hong X, Cohen SM, and Milán M (2011) Notch-mediated repression of bantam miRNA contributes to boundary formation in the *Drosophila* wing *Development* **138**, 3781–3789.
- [93] Bassett AR, Azzam G, Wheatley L, Tibbit C, Rajakumar T, McGowan S, et al. (2014) Understanding functional miRNA-target interactions in vivo by site-specific genome engineering *Nat Commun* **5**, 4640.

- [94] Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, et al. (2004) A combined computational-experimental approach predicts human microRNA targets *Genes Dev* **18**, 1165–1178.
- [95] Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. (2006) A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes *Cell* **126**, 1203–1217.
- [96] Friedman RC and Burge CB (2014) MicroRNA target finding by comparative genomics *Methods Mol Biol* **1097**, 457–476.
- [97] Nam JW, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, et al. (2014) Global analyses of the effect of different cellular contexts on microRNA targeting *Mol Cell* **53**, 1031–1043.
- [98] Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, et al. (2016) The UCSC Genome Browser database: 2016 update *Nucleic Acids Res* **44**, D717–725.
- [99] Thadani R and Tammi MT (2006) MicroTar: predicting microRNA targets from RNA duplexes *BMC Bioinformatics* **7 Suppl 5**, S20.
- [100] Kertesz M, Iovino N, Unnerstall U, Gaul U, and Segal E (2007) The role of site accessibility in microRNA target recognition *Nat Genet* **39**, 1278–1284.
- [101] Elefant N, Altuvia Y, and Margalit H (2011) A wide repertoire of miRNA binding sites: prediction and functional implications *Bioinformatics* **27**, 3093–3101.
- [102] Gumienny R and Zavolan M (2015) Accurate transcriptome-wide prediction of microRNA targets and small interfering RNA off-targets with MIRZA-G *Nucleic Acids Res* **43**, 1380–1391.
- [103] Betel D, Koppal A, Agius P, Sander C, and Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites *Genome Biol* **11**, R90.
- [104] Kozomara A and Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data *Nucleic Acids Res* **42**, D68–73.
- [105] Friedman RC, Farh KK, Burge CB, and Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs *Genome Res* **19**, 92–105.
- [106] Seitz H (2017) Issues in current microRNA target identification methods *RNA Biol* **14**, 831–834.
- [107] Betel D, Wilson M, Gabow A, Marks DS, and Sander C (2008) The microRNA.org resource: targets and expression *Nucleic Acids Res* **36**, D149–153.
- [108] Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows *Nucleic Acids Res* **41**, W169–173.

## 3.2 Identification of microRNA:Target Interactions Involved in Cell Proliferation

### Rationale of the strategy

The evaluation of the biological relevance of a miRNA target permits to distinguish between molecular and functional targets. However, to allow this distinction, it is necessary to study a quantifiable phenotype to assess the involvement of an individual target in establishing this phenotype. Specifically, deletion of a functional target site should partially recapitulate the phenotype induced by loss of miRNA expression. To interrogate the phenotypic involvement of many molecular targets of a miRNA on a large scale, we turned to a genetic screen in cultured cells. Among measurable biological responses in this model, cell proliferation is particularly suitable to interrogate a quantitative phenotype using mathematical modeling. Cell proliferation is a sensitive readout since cell number increases exponentially with time as long as environmental conditions are approximately constant. Additionally, a genetic screen on miRNA:target interactions is a powerful way of identifying molecular mechanisms underlying proliferation control. Especially, it could provide quantitative information allowing the precise estimation of individual and combinatorial effects of molecular events on proliferation through mathematical modeling.

We thus conceived a CRISPR-mediated screen where all the MREs for a miRNA of interest would be mutated and then cells would be allowed to grow for several days to measure the relative growth rate of individual mutant clones. This approach would allow measuring the proliferative effect of individual miRNA:target interactions, as well as that of combinations of interactions. Furthermore, because it is experimentally easier to detect cellular clones that over-accumulate in the cell population rather than clones that under-accumulate, an anti-proliferative miRNA would particularly fit this strategy. We opted from the miR-34 family described in the literature as an anti-proliferative miRNA in mouse and human cell lines and more broadly as a general tumor-suppressor miRNA (Rupaimoole et al., 2017). This work would identify miRNA:target interactions that do control a proliferative phenotype. Moreover, it would be the first example in which miRNA:target interactions are screened on the phenotype they regulate, rather than from theoretical inferences.

### Methodology description

In more detail, the strategy consists in using a CRISPR library that targets every conserved – at least between human and mouse – miR-34 binding site in 3' UTRs of diploid (HCT-116) and haploid (HAP1) human cell lines. These cell lines are particularly suitable for the observation of phenotypes after mutation of only two or a single allele. The CRISPR library and Cas9 will be transduced by lentiviral vectors to ensure a homogeneous cellular delivery, and the multiplicity of infection (MOI) will be optimized in order to perform two independent screens: on either (i) unique mutations per cellular clone to probe the effect of individual miR-34:target

### 3.2. Identification of microRNA:Target Interactions Involved in Cell Proliferation

interactions or (ii) multiple mutations to probe the effect of combinations of miR-34:target interactions. After transduction, cells will be allowed to grow exponentially while trypsinizing and splitting cells regularly to ensure that they never reach confluency. Genomic DNA will be extracted from a sample of cells over time for high-throughput sequencing of enriched loci of interest using the SureSelect Target Enrichment System (Agilent) and quantification of the proportion of mutant subclones in the cell population. Such a time-course experiment will allow to detect overgrowing mutants independently of sgRNAs' efficiency since even mutant clones generated in latter days of the time course will still be tracked by measuring their accumulation rate. The CRISPR screen will be conducted independently on wild-type and miR-34-deficient cells to identify and exclude 3' UTR mutations affecting cell proliferation in a miR-34-independent manner – e.g. by perturbing the recruitment of RNA-binding proteins or other UTR-dependent miRNA-independent mechanisms.

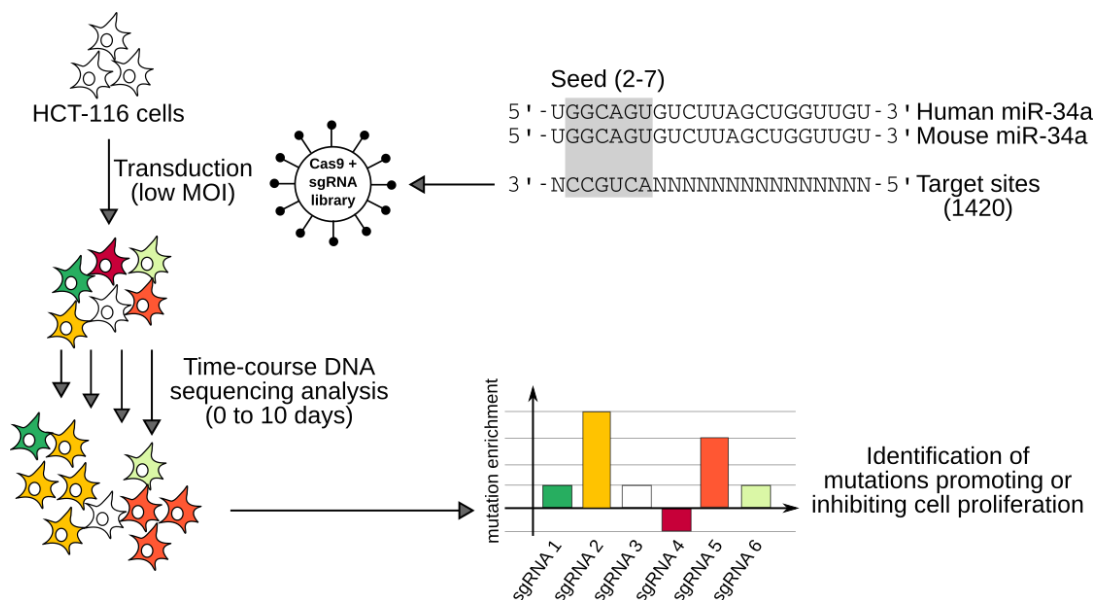


FIGURE 3.1: **Overview of a unbiased CRISPR screen for the identification of miRNA targets involved in cell proliferation control.** Among all potential miR-34a binding sites in the human genome, 1420 are conserved with the mouse. We planned to design guide RNAs to impair these sites, including for each site one guide targeting the *Streptococcus pyogenes* Cas9 PAM-compatible CCG sequence within the seed match. The CRISPR library would be transduced at low MOI, and a time course population genomic DNA sequencing will be performed to follow the enrichment or depletion of mutant subclones.

#### Quantification of the global proliferative effect of miR-34a

Prior to performing the CRISPR screen in cultured cells, it was first necessary to confirm the effect of the miR-34 family on cell proliferation, and to provide an accurate quantification of its global effect as a reference value for the evaluation of individual target contributions in our mathematical model.



The screen was supposed to be performed first in HCT-116 cells, a human colon carcinoma cell line, for two reasons: **(i)** this cell line is essentially diploid thus highly responsive to CRISPR/Cas9-mediated genome editing, and **(ii)** the miR-34 family has been reported to be anti-proliferative in HCT-116 in various studies (Chang et al., 2007; He et al., 2007; Navarro et al., 2015; Tazawa et al., 2007; Yamakuchi et al., 2008; Zhao et al., 2013). The miR-34 family is composed of six members but miR-34a is the major member of this family in most tissues and cell lines (Bommer et al., 2007; Concepcion et al., 2012; Okada et al., 2014; Song et al., 2014), including in HCT-116 cells. Therefore, we generated miR-34a-null HCT-116 clones by CRISPR/Cas9-mediated knock-out of the *miR-34a* locus as well as wild-type clones that have undergone the same transfection and single-cell sorting processes to be compared with. Indeed, the whole selection procedure could modify the proliferation behavior of cells or favorize the most robust clones, thereby the parental wild-type cell line would differ from clones not only in the expression of miR-34a but also in its systemic state.

To delete the *miR-34a* locus, multiple guide RNAs (sgRNAs) were designed using CRISPOR (<http://crispor.tefor.net/>) (Concordet et al., 2018) to target each side of the human genomic sequence expressing the pre-mir-34a sequence, and cloned into an expression plasmid for *Streptococcus pyogenes* Cas9 (pSpCas9(BB)-2A-GFP plasmid, Addgene 48138). The targeting efficiency of each plasmid was estimated by Sanger sequencing of the targeted locus in transfected HCT-116 cells, and analyzed with the Synthego ICE Analysis online tool (<https://ice.synthego.com/>). The mutagenesis was performed using two sgRNAs, one for each end of the targeted locus, with a high edition score. HCT-116 cells were grown until 80% confluency and transfected with the two plasmids following the protocol for Lipofectamine 2000 Transfection Reagent (ThermoFisher Scientific). After 24 hours, Cas9-GFP-expressing single cells were isolated in 96-well plates by flow cytometry on a BD FACSMelody (Becton Dickinson), then grown for 10 days. Homozygous wild-type and mutant clones were first tested by discriminative PCRs and eventually validated by Sanger sequencing of their *miR-34a* locus.

Eventually, we compared the proliferation rate of four mutant clones with four wild-type clones. Each cell line was seeded in 96-well plates in four replicates at low confluency, and from 24 hours after cell seeding to three days later, the number of living cells was determined twice a day by CellTiter-Glo Luminescent Cell Viability Assay (Promega), based on quantitation of the ATP present as an indicator of metabolically active cells. The luminescent signal is thus directly proportional to the number of viable cells and was log-transformed for linear regression relative to time and genotype to measure doubling time and estimate the significance of the effect of genotype.

In contradiction with previous papers based for the majority on over-expression experiments and for one of them on a knockout model, we did not observe any significant difference in the growth rate of mutant and wild-type clones (Figure 3.2.A), suggesting that miR-34a is not involved in cell proliferation control in this cell line. We repeated this approach on a second cell line, also suitable for a CRISPR screen, HAP1, a near-haploid cell line derived from

### 3.2. Identification of microRNA:Target Interactions Involved in Cell Proliferation

the chronic myelogenous leukemia. Once again, we did not observe the expected increase in cell proliferation for miR-34a-deficient clones, but instead a slight decrease (Figure 3.2.B).

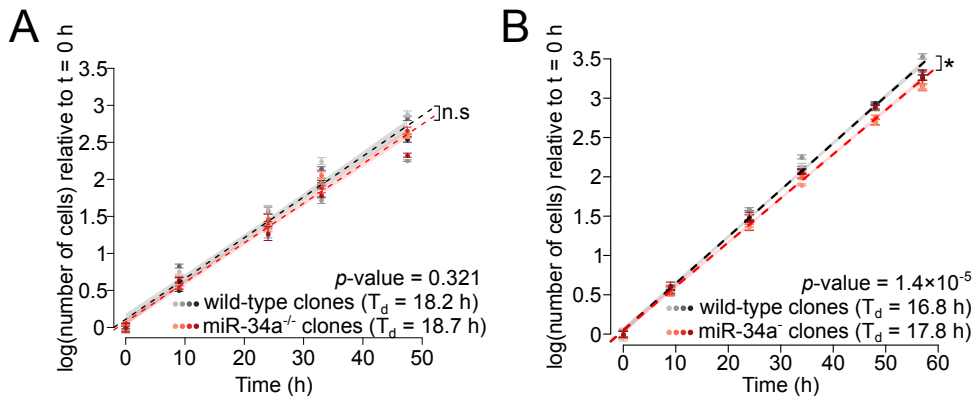


FIGURE 3.2: **Proliferative rate of wild-type and miR-34a-deficient clones in HCT-116 and HAP1 cells.** Four wild-type and four miR-34a mutant clones of **A**: HCT-116 and **B**: HAP1 were grown in sub-confluent conditions. Means and standard errors of 4 biological replicates are represented by dots and error bars. Linear modeling of log-transformed cell counts relative to time was used to measure doubling time ( $T_d$ ), and to estimate the significance of the effect of genotype (p-value is given in the inset). Shaded areas represent the 95% confidence interval for theoretical future measurements.

At this stage of the project, we gave up on the feasibility of the CRISPR screen on miR-34a targets since this miRNA does not exhibit an endogenous anti-proliferative activity in these two CRISPR-friendly cell lines. Nevertheless, the discrepancy between our results and published results, in particular compared with the knockout experiment in HCT-116 from Navarro et al., 2015 (Figure 3.3), led us to interrogate the genuine impact of miR-34a on cell proliferation. More broadly, we also questioned its supposed tumor-suppressor role, which had been based on the assumption of its anti-proliferative property.

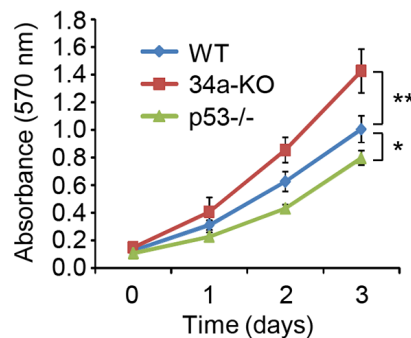


FIGURE 3.3: **Proliferative rate of wild-type and miR-34a-deficient clones in HCT-116 from Navarro et al., 2015.** Proliferation of wild-type, p53-KO and miR-34a-KO cells by MTT cell proliferation assay. Adapted from Navarro et al., 2015.

### Generation of miR-21-deficient HAP1 cells

Because miR-34a does not endogenously affect cell proliferation in HCT-116 or HAP1 cells, we assessed an other miRNA involved in cell growth regulation to perform the CRISPR-mediated screen. In place of selecting a well established pro-proliferative or anti-proliferative miRNA from the literature only, we prioritized miRNAs highly expressed in HCT-116 and HAP1 with the assumption that the most expressed miRNAs would be more likely functional (Mullokan-dov et al., 2012). We selected from published small RNA-seq data miR-21, an established pro-proliferative miRNA (Pfeffer et al., 2015), highly expressed in both cell lines (Figure 3.4).

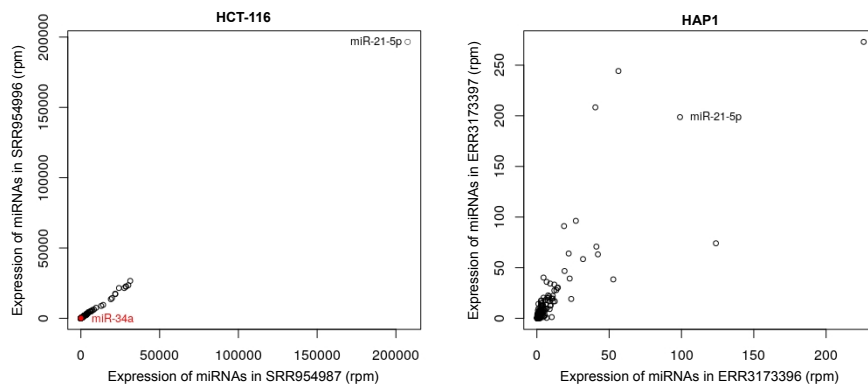


FIGURE 3.4: **Quantification of miRNAs expressed in HCT-116 and HAP1 cells from two biological replicates of small RNAseq.** The expression of human miRNAs was quantified from two published small RNA sequencing datasets (available on NCBI SRA) and represented by the respective number of reads per million (the number of reads mapping on a miRNA was divided by the total number of reads mapping on the reference repertoire of miRNAs, then multiplied by one million).

We attempted to generate miR-21-deficient HAP1 cells, following the exact same protocol as for miR-34a-deficient HCT-116 and HAP1 clones. Surprisingly, we tested about thirty sgRNAs, and none of them had an edition score reaching a similar score to those tested for mir-34a deletion. We went through with the mutagenesis process anyway and while the mir-34a knock-out in HAP1 cells was so potent that we struggled to obtain non-mutated clones, we did not obtain a single miR-21 knock-out clone. Instead, the only generated mutant clone exhibited an insertion of 542 bp from chromosome 4, while the miR-21 locus is on chromosome 17. The predicted secondary structure of this variant suggests that it could not be recognized as a Microprocessor substrate and thus could not produce mature miR-21 (Figure 3.5). That said, we did not verify by Northern Blot or RT-ddPCR the expression level of miR-21, and thus we cannot conclude whether this mutant is miR-21-null. This mutagenesis method was repeated with various sgRNAs with no better outcomes. Therefore, it is likely that miR-21 is required for cellular survival in HAP1, and consequently, miR-21-deficient cells die out during clone selection.

### 3.2. Identification of microRNA:Target Interactions Involved in Cell Proliferation

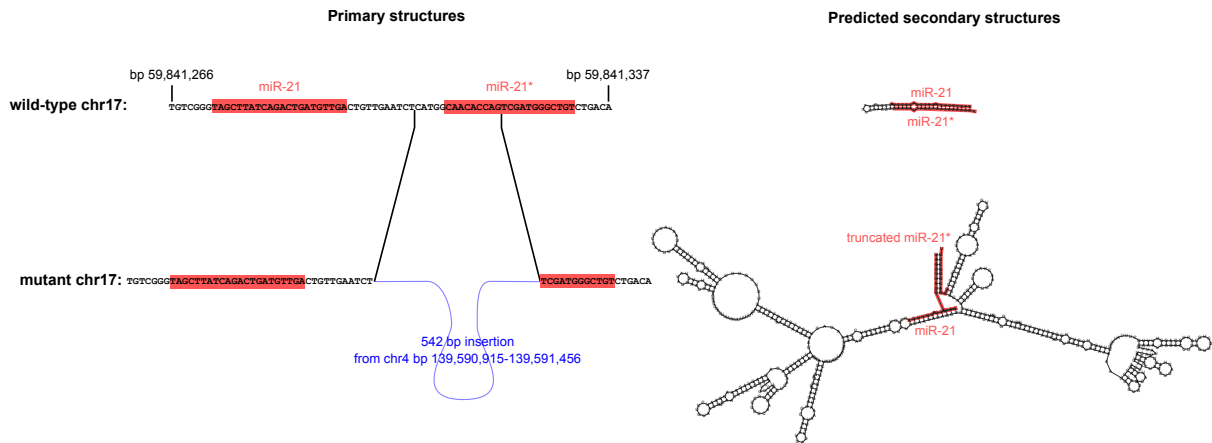


FIGURE 3.5: **Schematic of the single miR-21-mutant clone successfully isolated.** Left: The sequence of the miR-21 mutant clone presents an insertion of 542 bp at the expected editing position of one of the sgRNA used. In blue is illustrated the insertion of the 542 bp-long sequence which is based on chromosome 4 as a template. Right: This insertion disrupts the predicted secondary structure of pre-miR-21 by RNAfold (with default parameters). In red are highlighted the mature miR-21 sequence and its complementary strand, miR-21\*.

This project was halted at this stage and we did not investigate further the applicability of our CRISPR screen for the identification of functional miRNA targets involved in cell proliferation regulation. Instead, we explored the discrepancies between our results and the literature about the anti-proliferative activity of miR-34a, and questioned its definition as a general tumor-suppressor miRNA.

### 3.3 Revisiting the Anti-Proliferative Function of miR-34a

#### 3.3.1 The Master Tumor-Suppressive Role of miR-34a in the Literature

##### Description of the miR-34 family

The miR-34/449 miRNAs constitute a conserved family in vertebrates known as the miR-34 family and comprising three genomic loci, mir-34a, mir-34b~34c and mir-449c~449b~449a, which express six homologous miRNAs miR-34a, 34b, 34c, 449a, 449b and 449c. In humans, the miR-34a loci is located on chromosome 1p36.22 within the second exon of its non-coding host gene, the miR-34b~c cluster on chromosome 11q23.1, and the miR-449a~b~c cluster is originated from chromosome 5q11.2 and more specifically in the second intron of the cell division cycle 20b (*cdc20b*) gene.

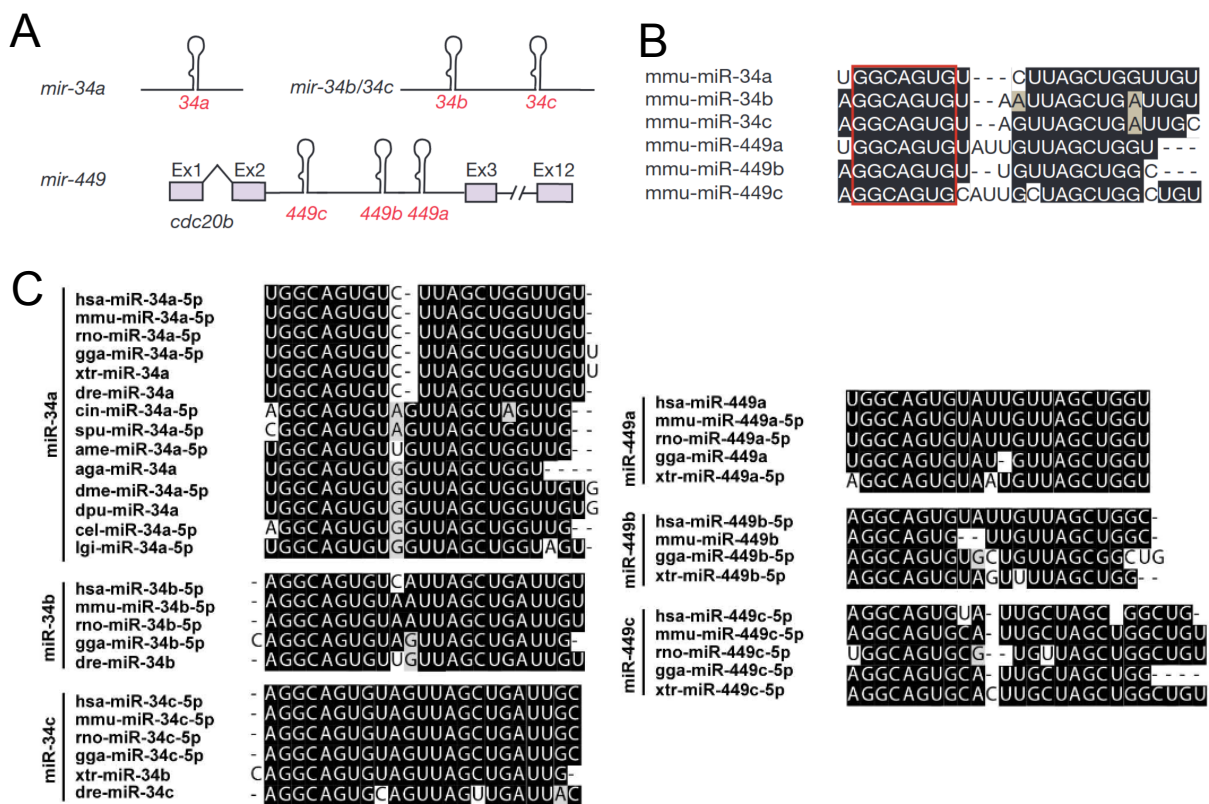


FIGURE 3.6: Structure and conservation of the miR-34 family in vertebrates. **A:** Gene structure of murine miR-34/449 miRNAs. **B:** Sequence alignment of murine miR-34/449 miRNAs. The red box indicates seed sequences. **C:** The miR-34 family is evolutionarily conserved with extensive sequence homology across many species. miR-34a has a more ancient evolutionary history compared to the rest of miR-34/449 miRNAs. miR-34a is conserved in Deuterostome, Ecdysozoa and Lophotrochozoa, yet the rest of miR-34/449 miRNAs have only vertebrate homologues. Adapted from Song et al., 2014.

In mice, this family of miRNAs is mainly detectable in the brain, lungs, testes, ovaries, and fallopian tubes (Bommer et al., 2007; Concepcion et al., 2012; Okada et al., 2014; Song et al., 2014). More precisely, miR-34/449 miRNAs are enriched in multiciliated cells, such as in airway mucociliary epithelial cells (lung) (Marcet et al., 2011) or efferent ductules (testis) (Yuan et al., 2019). Genetic ablation of the miR-34 family has been reported to result in ~60% of

postnatal mortality, ~50% of growth attenuation at the adult stage, severe respiratory distress and infertility of male and female mice (Song et al., 2014). These defects are overall explained by ciliogenesis dysfunction in the tissues of expression (Marcet et al., 2011; Song et al., 2014; Yuan et al., 2019).

#### **miR-34a, a candidate tumor-suppressor miRNA**

In 2007, seven papers identified the three members of the miR-34 family miR-34a, miR-34b and miR-34c as *bona fide* transcriptional targets of the tumor-suppressor protein p53 (Bommer et al., 2007; Chang et al., 2007; Corney et al., 2007; He et al., 2007; Raver-Shapira et al., 2007; Tarasov et al., 2007; Tazawa et al., 2007). Both miR-34a and miR-34b~c loci were found to be localized downstream predicted p53-binding sites, and the miR-34a/b/c expression profiles have been shown to be correlated with the p53 status in various cancer cell lines or induced by p53-mediated response to genotoxic stress.

The transcription factor p53 acts as a cellular central control hub that governs the cell decision to proliferate or, alternatively, to activate senescence and apoptotic programs in response to intrinsic and extrinsic cellular stresses resulting in DNA damage. In broad terms, if the degree of damage to the genome becomes excessive, or if the levels of nucleotide pools, growth-promoting signals, glucose or oxygenation conditions are suboptimal, p53 can induce cell-cycle progression halt until conditions go back to standard. Alternatively, in case of overwhelming genomic damage, p53 can trigger apoptosis (Hanahan et al., 2011). The primary target genes of the p53-mediated transcriptional activation differ according to the stress signal and activate a cascade of regulators providing a wide variety of cellular responses. Thus, the p53 pathway is redundant in how it mediates cell cycle control or cell death. Eventually, mutations in the *p53* gene are the most common genetic alteration in human cancers (Levine, 2020).

The three miRNAs, miR-34a, miR-34b and miR-34c, were found to be directly induced upon p53 activation in multiple murine and human cell types (Bommer et al., 2007; Chang et al., 2007; Corney et al., 2007; He et al., 2007; Raver-Shapira et al., 2007; Tarasov et al., 2007; Tazawa et al., 2007) and thus have been proposed to be part of the tumor-suppressor activity of the p53 network. In mice, miR-34a is detectable by Northern Blot and RT-qPCR in most tissues, while miR-34b/c are predominantly expressed in lung and testis along with miR-449a/b/c (Bommer et al., 2007; Concepcion et al., 2012; Okada et al., 2014; Song et al., 2014). The nearly ubiquitous expression of miR-34a made it a potential general actor of the p53 pathway. This assumption was corroborated by the fact that the *miR-34a* locus has been associated with fragile genomic sites in human cancer (Bagchi et al., 2008; Cole et al., 2008) and is often down-regulated in tumors and cancer cell lines (Adams et al., 2016; Bommer et al., 2007; Chang et al., 2007; Li et al., 2013; Liu et al., 2011; Tazawa et al., 2007; Wang et al., 2015a; Welch et al., 2007; Wiggins et al., 2010), suggesting that inactivation of miR-34a is involved in tumorigenesis and that other miR-34 family members could not compensate for this loss. Following, pieces of evidence had multiplied that ectopic expression of miR-34a in various cell types antagonizes cell proliferation (reviewed in Bader, 2012). Overexpression experiments

were performed in various ways: delivery of synthetic mature miR-34a, ectopic expression of miR-34a precursors, or else p53 induction with a genotoxic agent; all resulting in cell proliferation reduction due to either cell cycle arrest, apoptosis or senescence according to the cell type. Altogether, the affiliation of miR-34a in the p53 network, its loss of expression in various tumor samples or transformed cell lines, as well as its anti-proliferative activity at ectopic levels gave rise to the conclusion that miR-34a is a component of the tumor-suppressive p53 pathway and, more generally, is a putative general tumor-suppressor miRNA.

### **miR-34a replenishment therapeutics and their limits**

However, whereas these arguments are largely detailed in reviews about tumor-suppressor miRNAs, or miR-34a in particular, their limits are not or barely mentioned. First, the statement that miR-34a is located on a genomic region frequently deleted in cancer is based on the 1p36 cytogenetic band status evaluation which is 28 Mb-long (Cole et al., 2008; Mosse et al., 2007). Evidently, thousands of other genes are located within this region and, among them, potential tumor-suppressor genes. For instance, the chromodomain helicase DNA binding domain 5 (Chd5) gene, a tumor-suppressor candidate (Bagchi et al., 2008), is co-localized with miR-34a in the 3.5 Mb-long 1p36.22 band. Furthermore, the multiple evaluations of miR-34a expression in cancer tissues compared with adjacent or healthy tissue are penalized by (i) the low number of biological samples, (ii) the RT-qPCR detection limit for low-expressed miRNAs, as well as the nearly null endogenous level of miR-34a in control tissue making relative quantification obsolete, (iii) the visible heterogeneity in miR-34a expression profiles between samples which is not statistically investigated.

Moreover, contrary to the severe phenotypical defects observed in miR-34 family triple knock-out mice, miR-34a and miR-34b/c double knock-out mice do not exhibit apparent survival or growth defects. More importantly, and in contrast with p53-deficient mice, they do not display defects in p53-dependent proliferation control or increased susceptibility to spontaneous, irradiation-induced or cMyc-initiated tumorigenesis (Concepcion et al., 2012). Similarly, miR-34a/b/c-deficient MEFs do not show significant alteration in proliferation rate or replicative senescence in comparison with wild-type MEFs, while previous results would expect a p53-like phenotype with indefinite proliferation or higher proliferation rate. Interestingly, this study reports that while the expression of the miR-34 family is particularly high in testes, lungs and brains, it is largely p53-independent (Concepcion et al., 2012). Such investigation of the tumorigenesis susceptibility has not been published on triple knock-out miR-34 mice, yet they have been generated since 2014 (Song et al., 2014) and would be valuable to conclude on the role of the miR-34 family in tumorigenesis by taking into consideration the functional redundancy within members of a miRNA family.

Despite these findings, miR-34a was considered as a putative general tumor-suppressor worthy of becoming a miRNA-based therapeutics against cancer. MiRNA replacement therapy aimed at reintroducing a tumor-suppressor miRNA to reactivate a multiplicity of cellular pathways that drive a “therapeutic” response. Synthetic miR-34a mimics were delivered with

various vehicle systems in different solid tumors in murine models and were found to reduce solid tumor size and to increase the efficiency of already used therapeutic chemotoxic agents (Kasinski et al., 2012; Kasinski et al., 2015; Panebianco et al., 2019; Pramanik et al., 2011; Trang et al., 2011; Wiggins et al., 2010).

Intriguingly, in such studies, the amount of delivered miRNAs is rarely mentioned as a potential hazard. On the contrary, it is explicitly stated that miR-34a mimics should act like a naturally occurring miRNA by affecting all mRNAs inherently regulated by endogenous miR-34a in normal cells, and that such uptake of miRNA mimics should not affect healthy cells because pathways regulated by the miRNA mimic are already activated by the endogenous miRNA in these cells, which incidentally make dispensable local delivery (Bader, 2012; Trang et al., 2011; Wiggins et al., 2010). These assumptions reflect a misconception of miRNA targeting and neglect the notions of regulator:target availability and affinity. As detailed in our review, *cf.* section 3.1.1, an uncontrolled over-expression of miRNAs can lead to various confounding effects. To summarize, supra-physiological amounts of small RNAs can (i) titrate components of the endogenous miRNA machinery, thus relieving repression of other miRNA targets; (ii) induce the activation of the interferon pathway; (iii) increase endogenous target occupancy, hence achieving higher target repression; or else (iv) affect novel, low-affinity targets that are not repressed with endogenous miRNA concentrations. That is why, ectopic delivery of miR-34a must be verified to distinguish between endogenous anti-proliferative or excessive cytotoxic miR-34a effects. *A fortiori*, it was early stated that the definition of a tumor-suppressor gene excludes genes that are cytostatic or cytotoxic when inappropriately overexpressed (Weinberg, 1991). Thus, in the case where miR-34a is cytotoxic because anti-proliferative only in supra-physiological amounts, it would be wrong to define it as a tumor-suppressor miRNA.

Furthermore, if the miR-34a-based treatment is cytotoxic, it would contradict the initial motive to use a natural tumor-suppressor miRNA to modulate cellular pathways involved in tumor-suppressive functions. The tumor-suppressive function of miR-34a has also been justified by the identification of direct miR-34a targets known to act in tumor-suppressive pathways, such as the antiapoptotic proteins BCL-2 and SIRT-1 (Bommer et al., 2007; Li et al., 2013; Yamakuchi et al., 2008), the programmed cell death protein 1 ligand PDL1 (Wang et al., 2015b), or factors required for G1/S transition including c-Myc (Yamamura et al., 2012). However, all of these experiments have been performed with ectopic expression of miR-34a and therefore are not fully convincing.

Actually, MRX34, a liposome-delivered synthetic miR-34a mimic is the first miRNA-based treatment to have reached the phase I clinical trial in oncology, *i.e.* has been evaluated for its toxicity and its maximum tolerated dose in humans. Unfortunately, this trial has been closed early due to serious immune-mediated adverse events that resulted in four patient deaths and the clinical development of MRX34 was also halted since it has been concluded that the risk of severe adverse events was not outweighed by the 4% overall partial positive response rate (Beg et al., 2017; Hong et al., 2020).



### **Aim of the project**

Following our contradictory results about the effect of miR-34a on cell proliferation and our study of the literature about miR-34a in cancer, we speculated that the endogenous function of miR-34a as an anti-proliferative miRNA, and subsequently as a general tumor-suppressor miRNA, has been over-estimated due to misconceived experiments and results. Therefore, we decided to investigate further the pieces of evidence that proved the tumor suppressive function of miR-34a from its general downregulation in cancer and its antiproliferative activity in a cell line in which it has been extensively studied (HCT-116).

The main inputs of our work was to use, on one hand, the most up-to-date cancer genetics data from The Cancer Genome Atlas Program , and on another hand, an absolute ddPCR-based miRNA quantification method instead of RT-qPCR. The Cancer Genome Atlas Program is a cancer genomic program begun in 2006 and which generated data from over 20,000 primary cancer and matched normal samples spanning 33 cancer types. These data are standardized and available on the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>). Such dataset allows us to provide a large-scale analysis of miRNA expression and gene integrity in primary tumors and in normal adjacent tissues. We also used RT-ddPCR for accurate and, most importantly, absolute miRNA quantification (Hindson et al., 2013). This method is based on water-oil emulsion droplet technology and consists in fractionating a sample into 20,000 droplets in which PCR amplifications occur independently. Following, individual droplets are analyzed to determine the fraction of PCR-positive droplets. Because this system is based on a large number (20,000 independent PCRs) of rare events (presence of the template in the droplet), the target DNA template concentration in the original sample can be determined by Poisson statistics. Consequently, this technology requires a very small amount of input sample and provides an absolute count of target DNA copies per input sample.

The results of this work have been submitted for publication and are currently under revision. My participation covered the conception and design of the study, the generation of mutant cell lines, the absolute miRNA quantification by RT-ddPCR, the proliferation assays with the help of Elizabeth Houbron, the double-check of small RNA sequencing analysis, data analysis and interpretation, and paper writing. The majority of the bioinformatics analyses, including data mining from the GDC portal, were performed by Hervé Seitz.

### **3.3.2 Paper: A rationalized definition of tumor suppressor microRNAs excludes miR-34a**

# A rationalized definition of tumor suppressor microRNAs excludes miR-34a

Sophie Mockly<sup>1</sup>, Élisabeth Houbbron<sup>1</sup> and Hervé Seitz<sup>1,2</sup>

<sup>1</sup> IGH (CNRS and University of Montpellier), Montpellier, France

<sup>2</sup> Corresponding author; telephone: (+33)434359936; email: herve.seitz@igh.cnrs.fr

## Abstract

While several microRNAs (miRNAs) have been proposed to act as tumor suppressors, a consensual definition of tumor suppressing miRNAs is still missing. Similarly to coding genes, we propose that tumor suppressor miRNAs must show evidence of genetic or epigenetic inactivation in cancers, and exhibit an anti-proliferative activity under endogenous expression levels. Here we observe that this definition excludes the most extensively studied tumor suppressor candidate miRNA, miR-34a. In analyzable cancer types, miR-34a does not appear to be down-regulated in primary tumors relatively to normal adjacent tissues. Deletion of *miR-34a* is occasionally found in human cancers, but it does not seem to be driven by an anti-tumorigenic activity of the miRNA, since it is not observed upon smaller, *miR-34a*-specific alterations. Its anti-proliferative action was observed upon large, supra-physiological transfection of synthetic miR-34a in cultured cells, and our data indicates that endogenous miR-34a levels do not have such an effect. Our results therefore argue against a tumor suppressive function for miR-34a, providing an explanation to the lack of efficiency of synthetic miR-34a administration against solid tumors.

**Keywords:** microRNA / miR-34 / proliferation / tumor suppressor

## Introduction

Tumor suppressors are genes whose activity antagonizes tumorigenesis. Consequently, they are frequently silenced, either by germline-inherited or somatic mutation, or otherwise inactivated, in cancers [1]. Mechanistically, tumor suppressors mediate cellular environment-induced inhibition of cell proliferation, therefore exhibiting anti-proliferative activity under their natural expression levels: a gene displaying cytotoxic or cytostatic activity only when inappropriately overexpressed is therefore excluded from that definition [2].

miRNAs are small regulatory RNAs, guiding their effector proteins to specific target RNAs, which are repressed by various mechanisms (target RNA degradation and translational inhibition) [3]. Targets are recognized by sequence complementarity, with most targets bearing a perfect match to the miRNA “seed” (nt 2–7) [4]. Such a short binding motif makes miRNA/target binding poorly specific, and more than 60% of human genes are predicted to be targeted by at least one miRNA [5]. Because such gene regulators can act in signal transduction cascades, they may participate in tumor-suppressive pathways. A consensual definition for “tumor suppressor miRNAs” is still lacking, with some tentative definitions being based on miRNA down-regulation in cancer cells [6], on the targets’ annotation [7], or both [8]. We rather propose to follow the initial definition of tumor suppressors [2], considering that there is no reason to particularize miRNAs among other types of tumor suppressors. We thus advocate for the following definition of tumor suppressor miRNAs: **(i)** there is evidence for their frequent inactivation in cancer (either by genetic or epigenetic alteration; potentially only

in specific cancer types); and **(ii)** they repress cell proliferation under their endogenous expression level, rather than upon unrealistic overexpression.

We applied this definition to interrogate the status of the most highly-studied tumor suppressor candidate miRNA, miR-34a. It is a member of the miR-34 family, comprising six members in human and in mouse: miR-34a, miR-34b, miR-34c, miR-449a, miR-449b and miR-449c (Supplementary Figure S1). The three miR-34a/b/c subfamily members are transcriptionally controlled by the p53 tumor suppressor, which suggested that these miRNAs could participate in the tumor suppressive activity of the p53 network [9, 10, 11, 12, 13, 14, 15]. Indeed, the miR-34a member is down-regulated or lost in various cancer models (tumor samples or transformed cell lines) relatively to normal samples [9, 11, 16, 10, 14, 17, 18, 19, 20]. This observation suggested that the inactivation of *miR-34a* is involved in tumorigenesis, and that other family members (*miR-34b* and *c*, *miR-449a*, *b* and *c*) could not compensate for this loss. Yet *miR-34a*<sup>-/-</sup>, *miR-34b*<sup>-/-</sup>, *miR-34c*<sup>-/-</sup> triple knock-out mice do not exhibit obvious defects in p53-dependent proliferation control or in tumor suppression [21]. And, while pre-clinical studies in mice gave encouraging results (reviewed in [22, 23]), administration of a synthetic miR-34a to human patients with solid tumors failed to repress tumor growth reproducibly [24]. An alternate administration regimen (allowing increased drug exposure) did not clearly improve clinical outcomes, while triggering poorly-understood, severe adverse effects [23].

## Materials and Methods

### Analysis of miR-34a expression and integrity in human cancers

miRNA expression data was downloaded from the GDC portal on April 29, 2021. Cancer types where at least 10 cases were available (with Small RNA-Seq data from normal solid tissue and primary tumor for each case) were selected, and depth-normalized read counts were compared between normal tissue and tumor for each case. The heatmap shown on Figure 1A shows the median log-ratio between tumor and normal tissue, with non-significant changes (calculated with the Wilcoxon test, FDR-adjusted for multiple hypothesis testing) being colored in white.

miRNA gene ploidy data was downloaded from the GDC portal on March 4, 2021. Erroneous miRNA gene coordinates were corrected using information from miRBase. For the heatmap shown on Figure 1B, the percentage of cases with miRNA gene loss (either homo- or heterozygous) was evaluated for each miRNA, selecting cancer types where ploidy was determined in at least 100 cases.

miRNA sequence variation data was downloaded from the GDC portal on February 24, 2021. SNP location was intersected with miRNA hairpin and mature miRNA coordinates from miRBase (as well as with miRNA seed coordinates, defined as nt 2–7 of the mature miRNA). For the heatmaps shown on Supplementary Figure S2, the percentage of cases with sequence variations in miRNA genes (hairpin, mature or seed sequences) is displayed, selecting cancer types with at least 100 analyzed cases.

For each of these heatmaps, miRNAs and cancer types were clustered with the heatmap.2 command with the R software.

### CRISPR/Cas9-mediated mutagenesis

Four sgRNAs were designed using CRISPOR (<http://crispor.tefor.net/> [25]) to target each side of the human pre-mir-34a sequence, and cloned into an expression plasmid for *S. pyogenes* Cas9 (pSpCas9(BB)-2A-GFP plasmid (PX458), a gift from Feng Zhang [26]; Addgene plasmid #48138; <http://n2t.net/addgene:48138>; RRID:Addgene\_48138). Targeting efficiency of each plasmid was estimated by Sanger sequencing of the targeted locus in transfected HCT-116 cells, and analyzed with the Synthego ICE Analysis online tool (<https://ice.synthego.com/#/>). Mutagenesis was performed using the most efficient sgRNA sequence on each side of the targeted locus

(AAGCTCTTCTGCGCCACGGT**GGG** and GCCGGTCCACGGCATCCGGAG**GGG**; PAM sequences in bold; also see Supplementary Figure S6).

HCT-116 (ATCC® cat. #CCL247) and HAP1 (Horizon Discovery cat. #C631) cells were grown till 80% confluency and transfected with the two plasmids (15 µg each) following the protocol for Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific). After 24 hours, Cas9-GFP-expressing single cells were isolated in 96-well plates by flow cytometry on a BD FACSMelody (Becton Dickinson), then grown for 10 days. Homozygous wild-type and mutant clones were first tested by discriminative PCRs (with primer pairs ACTTCTAGGGCAGTATACTTGCT and GCTGTGAGTGTCTTTTGGC; and TCCTCCCCACATTTCTTCT and GCAAACCTTCTCCCAGCCAAA), and eventually validated by Sanger sequencing of their *miR-34a* locus. For the HAP1 cell line, mutagenesis efficiency was so high that we were unable to isolate wild-type clones after cotransfection of sgRNA-carrying PX458 plasmids. Wild-type clones were therefore generated by transfection of HAP1 cells with a plasmid expressing SpCas9-HF1 variant but no sgRNA (the VP12 plasmid, a gift from Keith Joung [27]; Addgene plasmid #72247 ; <http://n2t.net/addgene:72247>; RRID:Addgene\_72247), and went through the same isolation and selection process as mutant clones.

## RNA extraction

Cells plated in 10 cm Petri dishes were lysed and scrapped in 6 mL ice-cold TRIzol™ Reagent (Invitrogen) added directly to the culture dish after removal of the growth medium, and mixed with 1.2 mL of water-saturated chloroform. Samples were homogenized by vigorous shaking for 1 min and centrifuged for 5 min at 12,000 g and 4°C to allow phase separation. The aqueous phase was transferred in a new tube and mixed with 3 mL isopropanol for precipitation. After a 10 min incubation at room temperature, samples were centrifuged for 10 min at 12,000 g and 4°C and the supernatant was removed. The RNA pellet was washed with 6 mL of 70% ethanol and samples were centrifuged for 5 min at 12,000 g and 4°C. After complete removal of ethanol, the RNA pellet was resuspended in 20 µL RNase-free water and the quantity of total RNA was determined by spectrophotometry on a NanoDrop ND-1000.

## Small RNA-Seq

Total RNA of each cell line was extracted 48 h after seeding and quality was assessed on electrophoretic spectra from a Fragment Analyzer (Agilent), analyzed with the PROSize software (v. 3.0.1.6). Libraries were prepared using NEXTflex™ Small RNA-Seq Kit v3 (Bioo Scientific) following the manufacturer's instructions. Libraries were verified by DNA quantification using Fragment Analyzer (kit High Sensitivity NGS), and by qPCR (ROCHE Light Cycler 480). Libraries were sequenced on Illumina NovaSeq 6000 using NovaSeq Reagent Kit (100 cycles). RNA quality assessment, library preparation, validation and sequencing were performed by the MGX sequencing facility.

Adapters ended with 4 randomized nucleotides in order to reduce ligation biases. Because of the sequencing design, the adapter sequence (5' GTTCAGAGTTCTACAGTCCGACGATCNNNN 3') appears at the beginning of the read sequence, and the final 4 nucleotides of the read are the initial randomized nucleotides of the other adapter, whose other nucleotides are not read. Hence small RNA reads can be extracted from the fastq files with the following command:

```
cutadapt -g GTTCAGAGTTCTACAGTCCGACGATCNNNN --discard-untrimmed -m 18 -M 30 \  
$input_file.fastq | cutadapt -u -4 -
```

## Cell transfection

Cells were transfected 24 hours after seeding either with a control duplex, siRNA against eGFP: 5'-GGCAAGCUGACCCUGAAGUdTdT-3' / 5'-ACUUCAGGGUCAGCUUGCCdTdT-3'

or with a hsa-miR-34a mimic duplex:

5'-P-UGGCAGUGUCUUAGCUGGUUGUU-3' / 5'-P-CAAUCAGCAAGUAUACUGCCCUA-3'

according to the protocol for Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific).

### **Proliferation assays**

Because the mere procedure of isolating and selecting mutated clones may artifactually select clones with exceptionally high proliferation rates, we applied the same isolation and selection procedure to wild-type clones, and we measured proliferation rates on several independent wild-type and mutant clones. Each cell line was seeded in 96-well plates (Figure 3C: in 4 replicates at  $3 \times 10^3$  cells/well per time point; Figures 4A and B: in 6 replicates at  $6 \times 10^3$  cells/well). From 24 hours after cell seeding or transfection, to 3 days later, the number of living cells was determined twice a day by CellTiter-Glo Luminescent Cell Viability Assay (Promega) according to the manufacturer's protocol and recorded with a TriStar LB 941 (Berthold Technologies). Linear regression of log-transformed cell counts relative to time and genotype (in R syntax:  $\log\text{-transformed cell counts} \sim \text{time} * \text{genotype}$ ) or transfected duplex identity ( $\log\text{-transformed cell counts} \sim \text{time} * \text{duplex identity}$ ) was used to measure doubling time and to estimate the significance of the effect of genotype or transfected duplex.

For Figure 3D and E, doxorubicin (Sigma-Aldrich) was diluted in molecular biology-grade water and 5-fluorouracil (5-FU) (Sigma-Aldrich) diluted in dimethyl sulfoxide (SigmaAldrich). In a preliminary experiment, half-maximal inhibitory concentration (IC50) was estimated after 72 h drug exposure:  $7 \times 10^{-8}$  M and  $8 \times 10^{-6}$  M for doxorubicin and 5-FU respectively. Cell lines were seeded in 3 replicates per drug concentration at  $2.5 \times 10^3$  cells/well in 96-well plates. After 24 hours, culture medium was replaced with drug-containing medium (concentration range centered on the IC50 with  $2.5 \times$  increments), or solvent-containing medium for untreated controls, and the number of living cells was determined 72 h later by CellTiterGlo Luminescent Cell Viability Assay (Promega). Cell counts were normalized to the mean cell number in untreated controls. Normalized cell number was fitted to an asymptotic model for each clone to assess the significance of the effect of genotype (using an analysis of variance to compare a model not informed by clone genotype, to a genotype-informed model).

### **miRNA quantification by RT-ddPCR**

Reverse transcription of a specific miRNA in HCT-116 cells was performed on 10 ng total RNA using the TaqMan microRNA Reverse Transcription Kit (Thermo Fisher Scientific) in a total volume of 15  $\mu$ L, according to the manufacturer's protocol, with miRNA-specific RT primers from the TaqMan MicroRNA Assay Kit (assay IDs for hsa-miR-34a-5p and miR-21b-5p are respectively 000426 and 000397). ddPCR amplification of the cDNA was performed on 1.33  $\mu$ L of each cDNA combined with 1  $\mu$ L of miRNA-specific 20X TaqMan MicroRNA Reagent containing probes and primers for amplification from the TaqMan MicroRNA Assay Kit (Thermo Fisher Scientific), 10  $\mu$ L of 2X ddPCR Supermix for probes (no dUTP) (Bio-Rad), and 7.67  $\mu$ L of molecular biology-grade water. Droplets were generated, thermal cycled and detected by the QX200 Droplet Digital PCR System (Bio-Rad) according to the ddPCR Supermix protocol and manufacturer's instructions. Data were extracted using QuantaSoft Pro Software (Bio-Rad).

### **Data and script availability**

Deep-sequencing data has been deposited at SRA and linked to BioProject number PRJNA695193. Scripts, raw, intermediate and final data files are available at [https://github.com/HKKeyHKKey/Mockly\\_et\\_al\\_2021](https://github.com/HKKeyHKKey/Mockly_et_al_2021) and at <https://www.igh.cnrs.fr/en/research/departments/genetics-development/systemic-impact-of-small-regulatory-rnas#programmes-informatiques/>.

# Results

## No evidence for miR-34a loss or inactivation in cancers

It is now possible to compare miRNA levels between tumors and normal adjacent tissues on a large collection of human cases [28], allowing a rigorous assessment of miR-34a expression in tumorigenesis. Selecting every cancer type where miRNA expression is available for primary tumor and normal adjacent tissue, in at least 10 studied cases (n=20 cancer types), we did not find any cancer type where miR-34a was significantly down-regulated (Figure 1A). Hence in this collection of cancer types, human primary tumors do not tend to under-express miR-34a, contradicting the notion that genetic or epigenetic silencing of *miR-34a* could participate in tumorigenesis.

Accordingly, genetic alterations affecting *miR-34a* are very rare in cancer: focusing on every cancer type for which gene-level copy number was measured in at least 100 cases (n=29 cancer types), we did not observe any tendency for the loss of *miR-34a* relatively to other miRNA genes (see Figure 1B). Similarly, we did not find any evidence for the selective mutation of the pre-miR-34a hairpin precursor sequence, mature miR-34a or the miR-34a seed in cancers (n=30 analyzed cancer types; Supplementary Figure S2). In contrast to *miR-34a*, 105 miRNA loci tend to be frequently lost in 19 cancer types (red area at the top left corner of the heatmap in Figure 1B; listed in Supplementary Table S1): these miRNAs are more convincing tumor suppressor candidates than *miR-34a* in this respect.

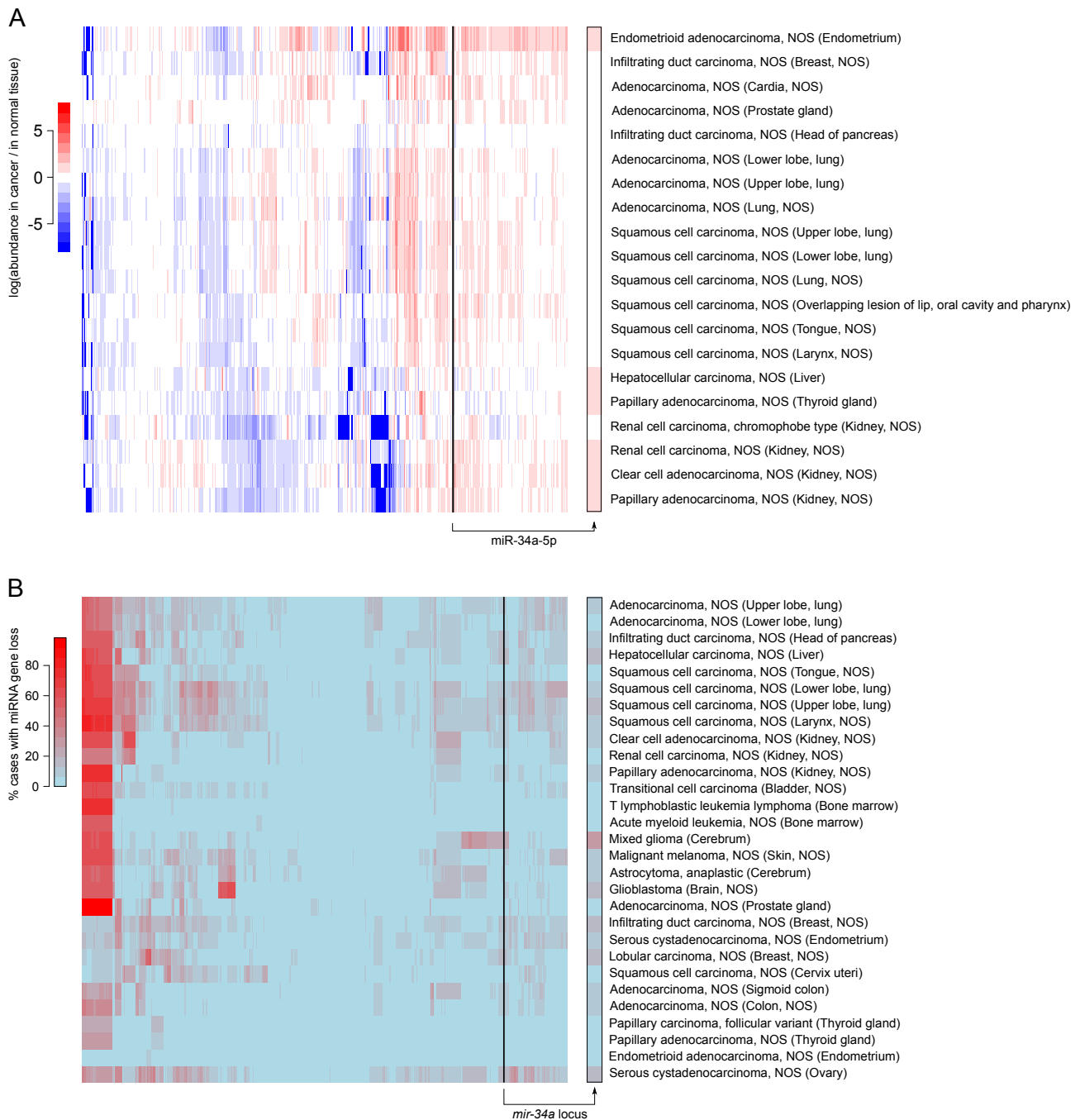
It could be argued that *miR-34a* inactivation by itself is insufficient to contribute to tumorigenesis, while it may play a role in a sensitized context, where additional, cooperative mutations may reveal the oncogenicity of miR-34a down-regulation. In that case, *miR-34a* inactivation could be enriched in just a subset of highly mutated cancers, and it would not be visible in the analyses shown in Figure 1 and Supplementary Figure S2. Yet, stratifying cases by cancer grade, we did not observe any tendency for the most aggressive tumors to inactivate *miR-34a* (Supplementary Figure S3), indicating that even the most sensitized tumors do not show any evidence of *miR-34a* inactivation.

Similarly, it is conceivable that miR-34a plays a tumor suppressive role only in the presence of functional p53, and the frequent mutation of p53 in the samples analyzed in Figure 1 may have obscured its behaviour in p53<sup>+/+</sup> tumors. But the selective analysis of cancer cases without any mutation in *p53* gives a very similar result, without miR-34a being down-regulated in any analyzed cancer type (see Figure 2).

Hence the loss or mutation of *miR-34a* does not appear to be enriched in cancer. We note that *miR-34a* is located on cytogenetic band 1p36, which is often altered in a wide variety of cancers. But our analyses suggest that the inactivation of *miR-34a* is not the actual driver for deletion selection – and because a convincing tumor suppressor is already known at 1p36 (the *CHD5* gene [29]), we propose that the occasional deletion of *miR-34a* in cancer is rather a consequence of its genomic proximity with such a real tumor suppressor. Accordingly, whenever a limited region of consistent deletion could be mapped in 1p36, that region excludes *miR-34a* (with the only exception of myelodysplastic syndromes, but with low experimental support): see Supplementary Figure S4.

## The reported anti-proliferative action of miR-34a is artifactual

miR-34a has also been considered a tumor suppressor candidate on the basis of the apparent anti-proliferative activity of miR-34 family miRNAs. Numerous studies in cultured cell lines indeed showed that miR-34 transfection inhibits cell proliferation, either by slowing down cell division or by increasing cell death [16, 9, 11, 12, 13, 14, 15]. But miRNA over-expression generates false positives, raising the possibility that this reported anti-proliferative role is artifactual [30]. We thus deleted the *miR-34a* gene in HCT-116 cells, where it has been proposed to be anti-proliferative by several independent studies [9, 11, 14] (mutagenesis strategy in Supplementary Figure S5). Deletion of the *miR-34a* locus eliminated 94% of the expression of the whole miR-34 family (Figure 3A and B). Our results do not show any significant difference in the growth rate of *miR-34a*<sup>-/-</sup> and wild-type clones



**Figure 1: miR-34a is not generally down-regulated or lost in cancers.** **A.** miRNA abundance (normalized by the number of mapped miRNA reads) was compared between primary tumors and normal adjacent tissues. Only cancer types for which at least 10 cases were analyzed have been considered ( $n=20$  cancer types; rows), and miRNAs with a null variance across cancer types were excluded (remaining:  $n=545$  miRNAs; columns). For each miRNA/cancer type pair, the heatmap shows its median fold-change across all cases, with non-significant changes ( $FDR \geq 0.05$ ) being shown in white.  $\log(\text{fold-changes})$  larger than  $+8$  or smaller than  $-8$  were set to  $+8$  or  $-8$  respectively, for graphical clarity. **B.** Only cancer types for which at least 100 cases were analyzed have been considered ( $n=29$  cancer types; rows), and miRNA genes whose ploidy could not be assessed were excluded (remaining:  $n=1,686$  miRNAs; columns). For each miRNA/cancer type pair, the heatmap shows the percentage of cases with monoallelic or biallelic loss of the miRNA gene. **Both panels:** the column showing miR-34a data is magnified on the right margin (framed in black). “NOS”: not otherwise specified.

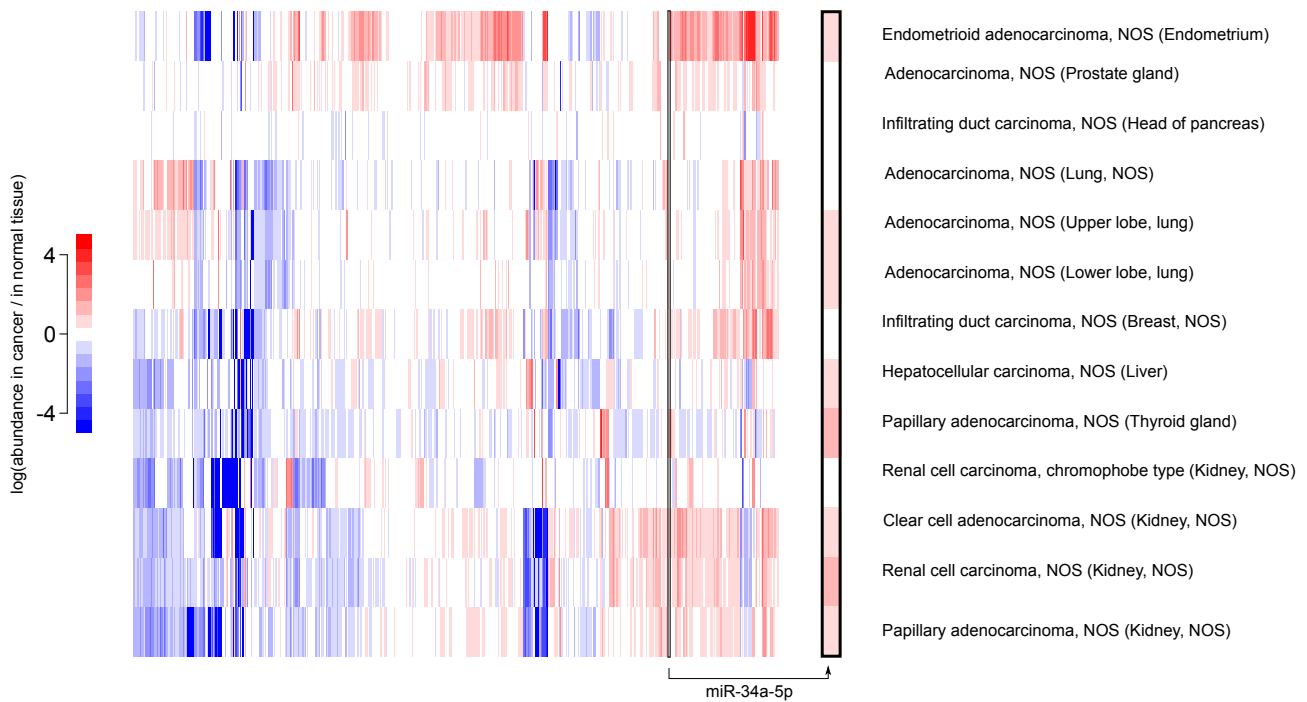


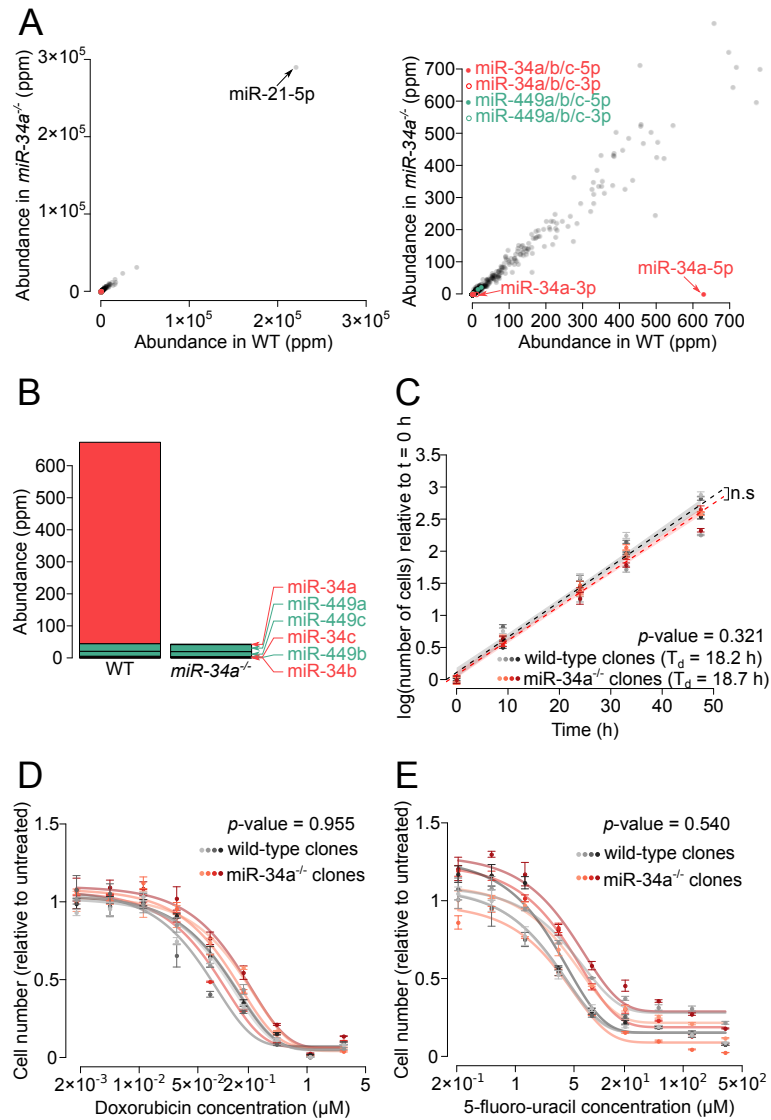
Figure 2: **No evidence for *mir-34a* inactivation in tumors with an intact *p53* gene.** Cancer samples analyzed in Figure 1A were stratified by the mutation status of the *p53* gene, and only cases without any detected mutation in *p53* were selected here (also selecting cancer types with at least 10 cases after this selection). Same conventions than in Figure 1A. miRNA abundance (normalized by the number of mapped miRNA reads) was compared between primary tumors and normal adjacent tissues. The column showing miR-34a data is magnified on the right margin (framed in black). log(fold-changes) larger than +5 or smaller than -5 were set to +5 or -5 respectively, for graphical clarity. “NOS”: not otherwise specified.

(Figure 3C). We also prepared *miR-34a*<sup>-</sup> clones from the human haploid HAP1 cell line, where miR-34a is also not anti-proliferative (it is even slightly pro-proliferative; Supplementary Figure S6). It could be argued that *miR-34a* does not inhibit cell proliferation in unstressed conditions, while being anti-proliferative upon genotoxic stress. But we also failed to observe significant differences between wild-type and mutant clones under doxorubicin or 5-fluoro-uracil treatment (Figure 3D and E).

In agreement with published data, we did observe a strong reduction in cell proliferation when we transfected HCT-116 cells with large amounts (10 nM) synthetic miR-34a duplex (Figure 4A), but that effect was lost when transfecting 1 nM duplex (Figure 4B). Absolute miRNA quantification by RT-ddPCR shows that a 10 nM transfection over-expresses miR-34a by >8,000-fold in HCT-116 cells (and a 1 nM transfection over-expresses it by >490-fold), clearly demonstrating that such an experiment results in supra-physiological miRNA concentrations (Figure 4C). For comparison, we measured the increase in miR-34a expression in response to DNA damage: a 72 h treatment with doxorubicin at its IC50 concentration ( $7 \times 10^{-8}$  M in HCT-116 cells; Supplementary Figure S7) over-expresses miR-34a by only 4.7-fold (Figure 4D).

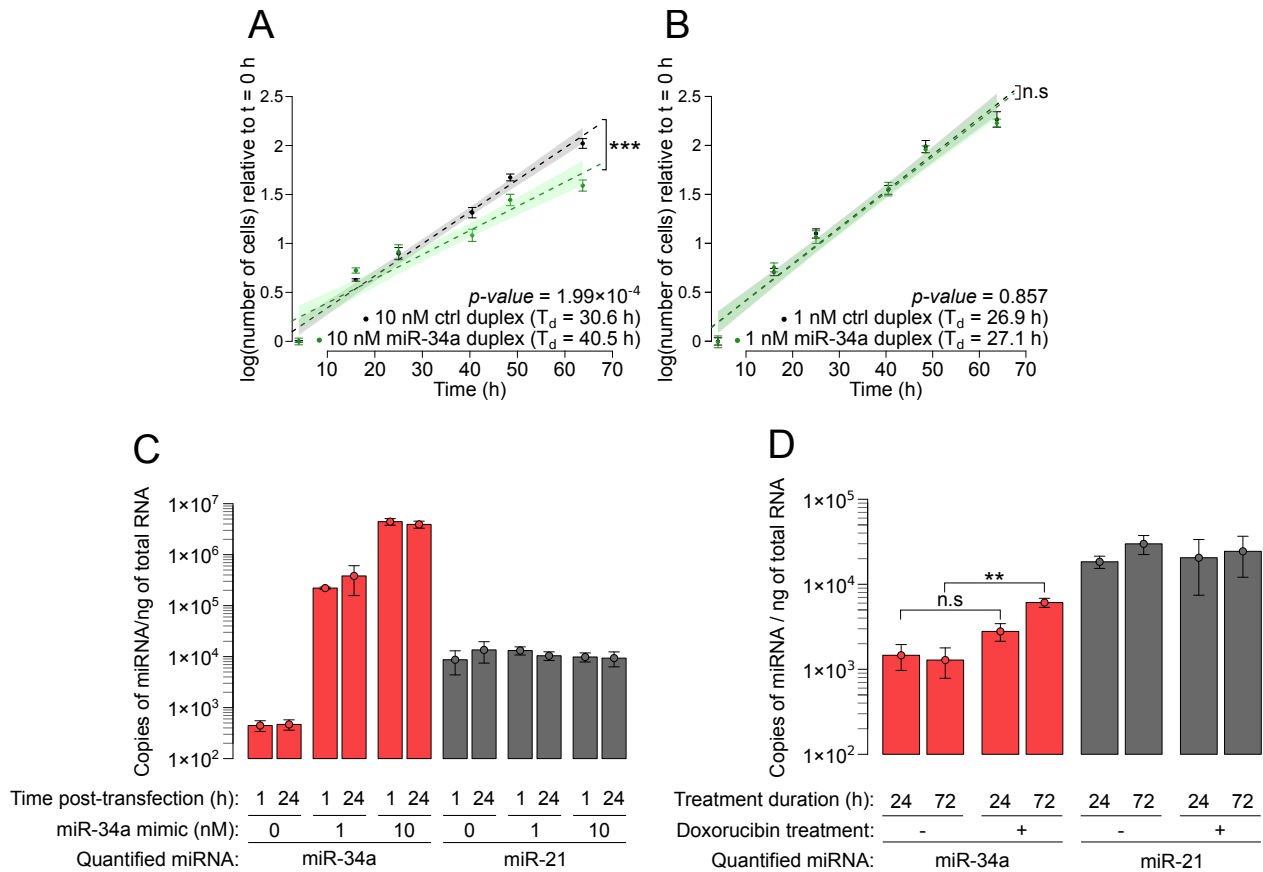
Of note, some authors have previously characterized the proliferative effect of miR-34 using genetic ablation rather than over-expression. In one study, mouse embryonic fibroblasts (MEFs) devoid of miR-34a/b/c appear to grow at the same rate than wild-type MEFs, except, transiently, for one early time-point [21]. In another study, genetic inactivation of the *miR-34a* gene in HCT-116 is reported to accelerate cell proliferation, in stark contrast with our own findings [31]. Such discrepancy would deserve to be investigated, but unfortunately that published mutant cell line has been lost and it is no longer available from the authors (Dr. J. Lieberman, personal communication).





**Figure 3: miR-34 is not a general repressor of cell proliferation.** **A.** miRNA quantification by Small RNA-Seq in a representative wild-type HCT-116 clone (*x* axis) and a representative *miR-34a*<sup>-/-</sup> clone (*y* axis). Right panel: magnification of the left panel. **B.** Cumulated abundance of miR-34 family members in the two clones. miRNAs are sorted vertically according to their abundance in the wild-type clone. **C.** Four wild-type and four *miR-34a* mutant clones were grown in sub-confluent conditions. Means and standard errors of 4 biological replicates are represented by dots and error bars. Linear modeling of log-transformed cell counts relative to time was used to measure doubling time ( $T_d$ ), and to estimate the significance of the effect of genotype ( $p$ -value is given in the inset). Shaded areas represent the 95% confidence interval for theoretical future measurements. **D, E.** Cell number after 3 days of culture in presence of varying doses of **(D)** doxorubicin or **(E)** 5-fluoro-uracil (4 clones of each genotype were analyzed; 3 biological replicates for each drug concentration; mean  $\pm$  st. error is shown). Cell number was normalized to cell number count in untreated replicates. Normalized cell number was fitted to an asymptotic model for each clone (fitted models are represented by curves). In order to assess the significance of the effect of genotype, a naïve (non-informed by clone genotype) and a genotype-informed model were compared by an analysis of variance ( $p$ -value is indicated in the inset).

While the miR-34 family is believed to exert a tumor suppressive action in a diversity of cancers [32], we observed that it is hardly expressed in cultured cell lines, primary tissues and body fluids (Supplementary Figure S8–S10). It could be argued that a low level of miR-34 expression is expected in normal tissues, where p53 is mostly inactive. But p53 is clearly not the only regulator for miR-34,



**Figure 4: Supra-physiological transfection of miR-34a inhibits cell proliferation.** Wild-type HCT-116 cells were transfected with 10 nM (panel **A**) or 1 nM (panel **B**) duplex (either a control siRNA duplex, or miR-34a/miR-34a\* duplex) and grown in sub-confluent conditions. Means and standard errors of 6 biological replicates are represented by dots and error bars. Linear modeling of log-transformed cell counts relative to time was used to measure doubling time ( $T_d$ ), and to estimate the significance of the effect of duplex identity ( $p$ -values are given in the inset; asterisks denote  $p$ -value  $< 0.05$ , “n.s.” indicates larger  $p$ -values). Shaded areas represent the 95% confidence interval for theoretical future measurements. **C.** Cellular abundance of miR-34a (red bars) or a control miRNA (miR-21; gray bars) 1 or 24 h after transfection of HCT-116 cells with 0, 1 or 10 nM miR-34a/miR-34a\* duplex. **D.** HCT-116 cells were treated for 24 or 72 h with  $7 \times 10^{-8}$  M doxorubicin, and their intracellular miR-34a and miR-21 were quantified by RT-ddPCR. Two-way ANOVA analysis shows that doxorubicin treatment has an effect on miR-34a levels ( $p=0.0013$ ), and post-hoc pairwise t-tests find the effect significant only after 72 h exposure to the drug ( $p=0.0521$  for 24 h exposure,  $p=0.00138$  for 72 h exposure, indicated by “n.s.” and “\*\*\*” respectively). A similar two-way ANOVA analysis does not detect a significant effect of doxorubicin treatment on miR-21 levels ( $p=0.768$ ). **Panels C and D:** Means and standard errors of 3 biological replicates are represented by dots and error bars, respectively.

and the expression of miR-34 does not mirror p53 activity [21]. Current RNA detection technologies can be extremely sensitive, and they can detect miRNAs which are too poorly abundant to induce any clear change in target expression [33]. Hence we anticipate that in all the cell lines for which we analyzed miRNA abundance, and in most cells in the analyzed tissues, miR-34 family miRNAs are actually non-functional.

Yet we do not question the overall functionality of miR-34 miRNAs *in vivo*. Because that family is deeply conserved in evolution (shared between, *e.g.*, vertebrates and insects), it certainly plays important biological functions, perhaps only in a small number of cells, or at very specific developmental stages, where its abundance would be high enough. In mouse, the miR-34 family is particularly ex-

pressed in lungs and testes [21, 34]. Mutation of all 6 members of the miR-34 family causes severe ciliogenesis defects, leading to respiratory distress and impaired gametogenesis – translating into sterility and premature mortality [34]. Unsurprisingly then, the most obvious biological functions of that miRNA family seem to take place in the tissues where miR-34 miRNAs are highly expressed, in contrast with the widely-accepted notion of their broad anti-tumorigenic activity.

## Discussion

While the original definition for tumor suppressors had been formulated with coding genes in mind, we consider that there is no objective reason for adopting a different definition for tumor suppressor miRNAs. In this view, the most heavily studied candidate tumor suppressor miRNA, miR-34a, does not appear to be a tumor suppressor. It remains formally possible that miR-34a inactivation is frequent in specific cancer types, distinct from those we could analyze in Figures 1 and 2 and Supplementary Figure S2–3. In that case, miR-34a may be a tumor suppressor in these particular cancers, but rigorous investigation – while avoiding the pitfalls described above – of its impact on cell proliferation and tumorigenesis would be necessary to conclude so.

We confirmed that a large artificial over-expression (10 nM) of miR-34a indeed represses cell proliferation. It could be argued that this cytotoxic effect could provide the ground for an efficient anti-cancer treatment, no matter how un-natural it is. But the whole purpose of using natural tumor suppressors (e.g., miRNAs) is that they are expected to be well tolerated, because they already exist endogenously. Administering large amounts of cytotoxic agents to patients may indeed kill cancer cells – but it will also likely trigger unwanted adverse effects. In this view, synthetic miR-34a behaves similarly to existing anti-cancer drugs, which are based on exogenous molecules. It is therefore not surprising to observe a variety of adverse secondary effects when the MRX34 miR-34a mimic is administered to patients [24, 23]. More innocuous miRNA-based treatments may be possible, but they would have to rely on rigorously established tumor-suppressive activity of the endogenous miRNA.

## Acknowledgments

We thank G. Canal, C. Theillet, P.D. Zamore and members of the Seitz lab for critical reading of the manuscript, and A. Pélisson and K. Mochizuki for useful discussions. This research was supported by Cancéropôle GSO “Émergence” grant and Projet Fondation ARC #PJA 20191209613. We thank the MGX facility (Biocampus Montpellier, CNRS, INSERM, Univ. Montpellier, Montpellier, France) for sequencing the Small RNA-Seq libraries.

## Author contributions

S.M. and É.H. performed experiments; S.M. and H.S. performed computational analyses; S.M. and H.S. wrote the manuscript and prepared figures.

## Conflict of interest

The authors do not declare any conflict of interest.

## References

- [1] Green, A. R. (1988) Recessive mechanisms of malignancy. *Br J Cancer*, **58**(2), 115–121.

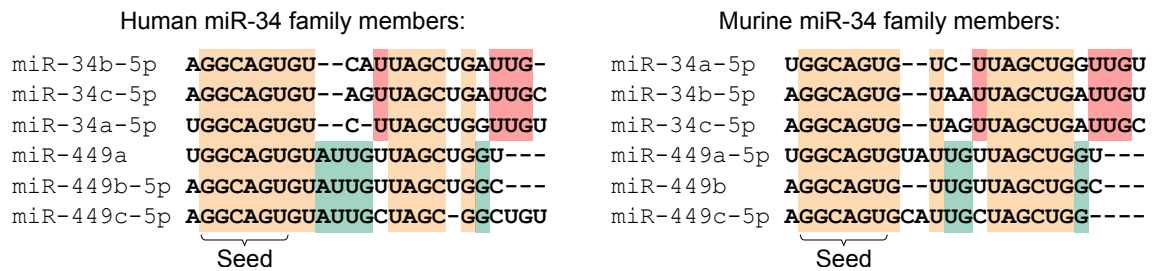
- [2] Weinberg, R. A. (1991) Tumor suppressor genes. *Science*, **254**(5035), 1138–1146.
- [3] Iwakawa, H. O. and Tomari, Y. (2015) The functions of MicroRNAs: mRNA decay and translational repression. *Trends Cell Biol*, **25**(11), 651–665.
- [4] Bartel, D. P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**(2), 215–233.
- [5] Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, **19**(1), 92–105.
- [6] Zhang, B., Pan, X., Cobb, G. P., and Anderson, T. A. (2007) microRNAs as oncogenes and tumor suppressors. *Dev Biol*, **302**(1), 1–12.
- [7] Wong, K. Y., Yu, L., and Chim, C. S. (2011) DNA methylation of tumor suppressor miRNA genes: a lesson from the miR-34 family. *Epigenomics*, **3**(1), 83–92.
- [8] Adams, B. D., Parsons, C., and Slack, F. J. (2016) The tumor-suppressive and potential therapeutic functions of miR-34a in epithelial carcinomas. *Expert Opin Ther Targets*, **20**(6), 737–753.
- [9] He, L., He, X., Lim, L. P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., Jackson, A. L., Linsley, P. S., Chen, C., Lowe, S. W., Cleary, M. A., and Hannon, G. J. (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**(7148), 1130–1134.
- [10] Bommer, G. T., Gerin, I., Feng, Y., Kaczorowski, A. J., Kuick, R., Love, R. E., Zhai, Y., Giordano, T. J., Qin, Z. S., Moore, B. B., MacDougald, O. A., Cho, K. R., and Fearon, E. R. (2007) p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr Biol*, **17**(15), 1298–1307.
- [11] Chang, T.-C., Wentzel, E. A., Kent, O. A., Ramachandran, K., Mullendore, M., Lee, K. H., Feldmann, G., Yamakuchi, M., Ferlito, M., Lowenstein, C. J., Arking, D. E., Beer, M. A., Maitra, A., and Mendell, J. T. (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell*, **26**(5), 745–752.
- [12] Corney, D. C., Flesken-Nikitin, A., Godwin, A. K., Wang, W., and Nikitin, A. Y. (2007) MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res*, **67**(18), 8433–8438.
- [13] Tarasov, V., Jung, P., Verdoodt, B., Lodygin, D., Epanchintsev, A., Menssen, A., Meister, G., and Hermeking, H. (2007) Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle*, **6**(13), 1586–1593.
- [14] Tazawa, H., Tsuchiya, N., Izumiya, M., and Nakagama, H. (2007) Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci USA*, **104**(39), 15472–15477.
- [15] Raver-Shapira, N., Marciano, E., Meiri, E., Spector, Y., Rosenfeld, N., Moskovits, N., Bentwich, Z., and Oren, M. (2007) Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell*, **26**(5), 731–743.
- [16] Welch, C., Chen, Y., and Stallings, R. L. (2007) MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene*, **26**(34), 5017–5022.
- [17] Lodygin, D., Tarasov, V., Epanchintsev, A., Berking, C., Knyazeva, T., Körner, H., Knyazev, P., Diebold, J., and Hermeking, H. (2008) Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle*, **7**(16), 2591–2600.
- [18] Gallardo, E., Navarro, A., Viñolas, N., Marrades, R. M., Diaz, T., Gel, B., Quera, A., Bandres, E., Garcia-Foncillas, J., Ramirez, J., and Monzo, M. (2009) miR-34a as a prognostic marker of relapse in surgically resected non-small-cell lung cancer. *Carcinogenesis*, **30**(11), 1903–1909.

- [19] Wiggins, J. F., Ruffino, L., Kelnar, K., Omotola, M., Patrawala, L., Brown, D., and Bader, A. G. (2010) Development of a lung cancer therapeutic based on the tumor suppressor microRNA-34. *Cancer Res*, **70**(14), 5923–5930.
- [20] Corney, D. C., Hwang, C.-I., Matoso, A., Vogt, M., Flesken-Nikitin, A., Godwin, A. K., Kamat, A. A., Sood, A. K., Ellenson, L. H., Hermeking, H., and Nikitin, A. Y. (2010) Frequent downregulation of miR-34 family in human ovarian cancers. *Clin Cancer Res*, **16**(4), 1119–1128.
- [21] Concepcion, C. P., Han, Y.-C., Mu, P., Bonetti, C., Yao, E., D'Andrea, A., Vidigal, J. A., Maughan, W. P., Ogrodowski, P., and Ventura, A. (2012) Intact p53-dependent responses in miR-34-deficient mice. *PLoS Genet*, **8**(7), e1002797.
- [22] Bader, A. G. (2012) miR-34 - a microRNA replacement therapy is headed to the clinic. *Front Genet*, **3**, 120.
- [23] Hong, D. S., Kang, Y.-K., Borad, M., Sachdev, J., Ejadi, S., Lim, H. Y., Brenner, A. J., Park, K., Lee, J.-L., Kim, T.-Y., Shin, S., Becerra, C. R., Falchook, G., Stoudemire, J., Martin, D., Kelnar, K., Peltier, H., Bonato, V., Bader, A. G., Smith, S., Kim, S., O'Neill, V., and Beg, M. S. (2020) Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours. *Br J Cancer*, **122**(11), 1630–1637.
- [24] Beg, M. S., Brenner, A. J., Sachdev, J., Borad, M., Kang, Y.-K., Stoudemire, J., Smith, S., Bader, A. G., Kim, S., and Hong, D. S. (2017) Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Invest New Drugs*, **35**(2), 180–188.
- [25] Concordet, J.-P. and Haeussler, M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*, **46**(W1), W242–W245.
- [26] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, **8**(11), 2281–2308.
- [27] Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., and Joung, J. K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**(7587), 490–495.
- [28] Zhang, Z., Hernandez, K., Savage, J., Li, S., Miller, D., Agrawal, S., Ortuno, F., Staudt, L. M., Heath, A., and Grossman, R. L. (2021) Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat Commun*, **12**(1), 1226.
- [29] Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H., and Mills, A. A. (2007) *CHD5* is a tumor suppressor at human *1p36*. *Cell*, **128**(3), 459–475.
- [30] Mockly, S. and Seitz, H. (2019) Inconsistencies and Limitations of Current MicroRNA Target Identification Methods. *Methods Mol Biol*, **1970**, 291–314.
- [31] Navarro, F. and Lieberman, J. (2015) miR-34 and p53: New Insights into a Complex Functional Relationship. *PLoS One*, **10**(7), e0132767.
- [32] Slack, F. J. and Chinnaiyan, A. M. (2019) The Role of Non-coding RNAs in Oncology. *Cell*, **179**(5), 1033–1055.
- [33] Mullokandov, G., Baccarini, A., Ruzo, A., Jayaprakash, A. D., Tung, N., Israelow, B., Evans, M. J., Sachidanandam, R., and Brown, B. D. (2012) High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods*, **9**(8), 840–846.
- [34] Song, R., Walentek, P., Sponer, N., Klimke, A., Lee, J. S., Dixon, G., Harland, R., Wan, Y., Lishko, P., Lize, M., Kessel, M., and He, L. (2014) miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. *Nature*, **510**(7503), 115–120.

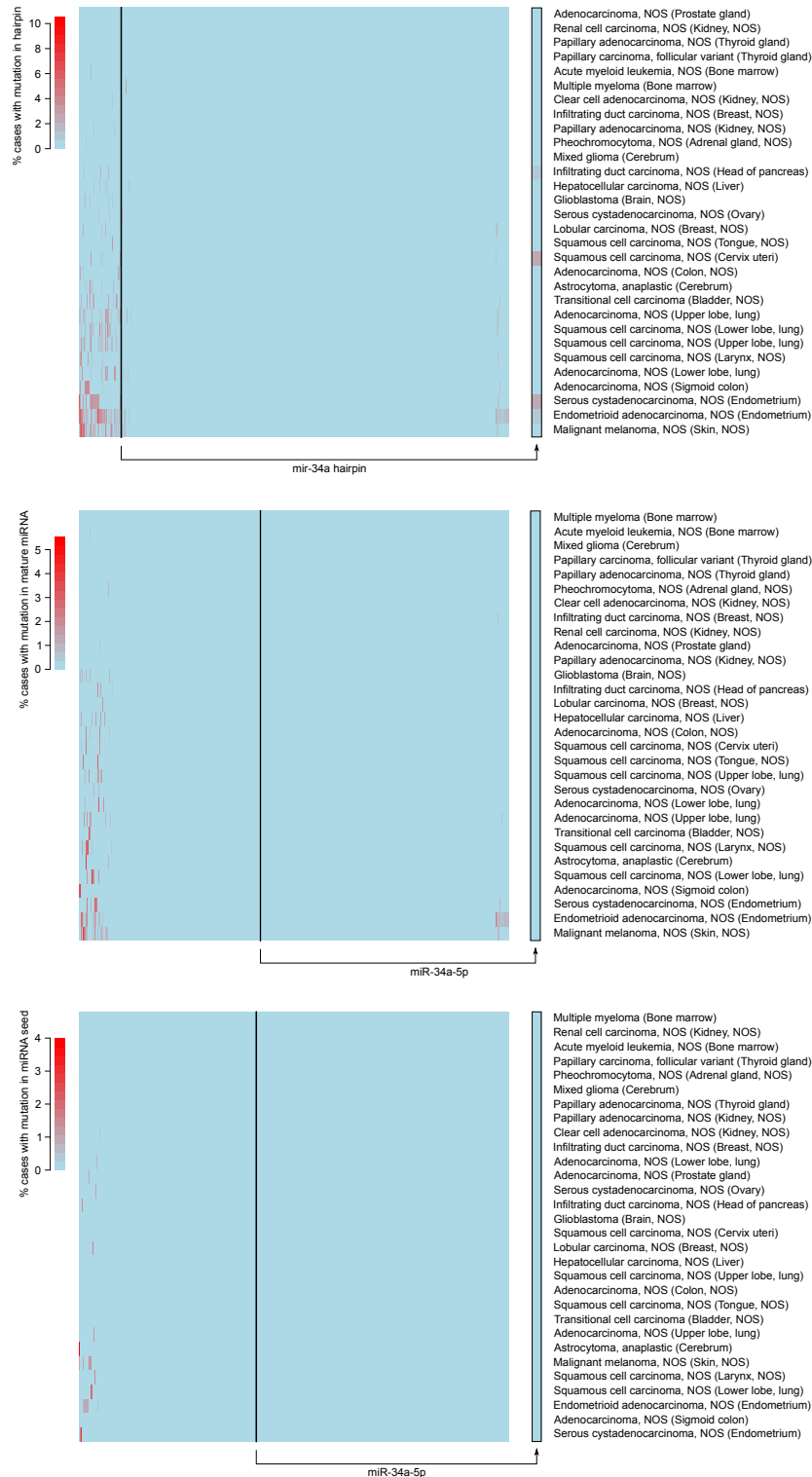
# Supplementary Data

A rationalized definition of tumor suppressor microRNAs excludes miR-34a

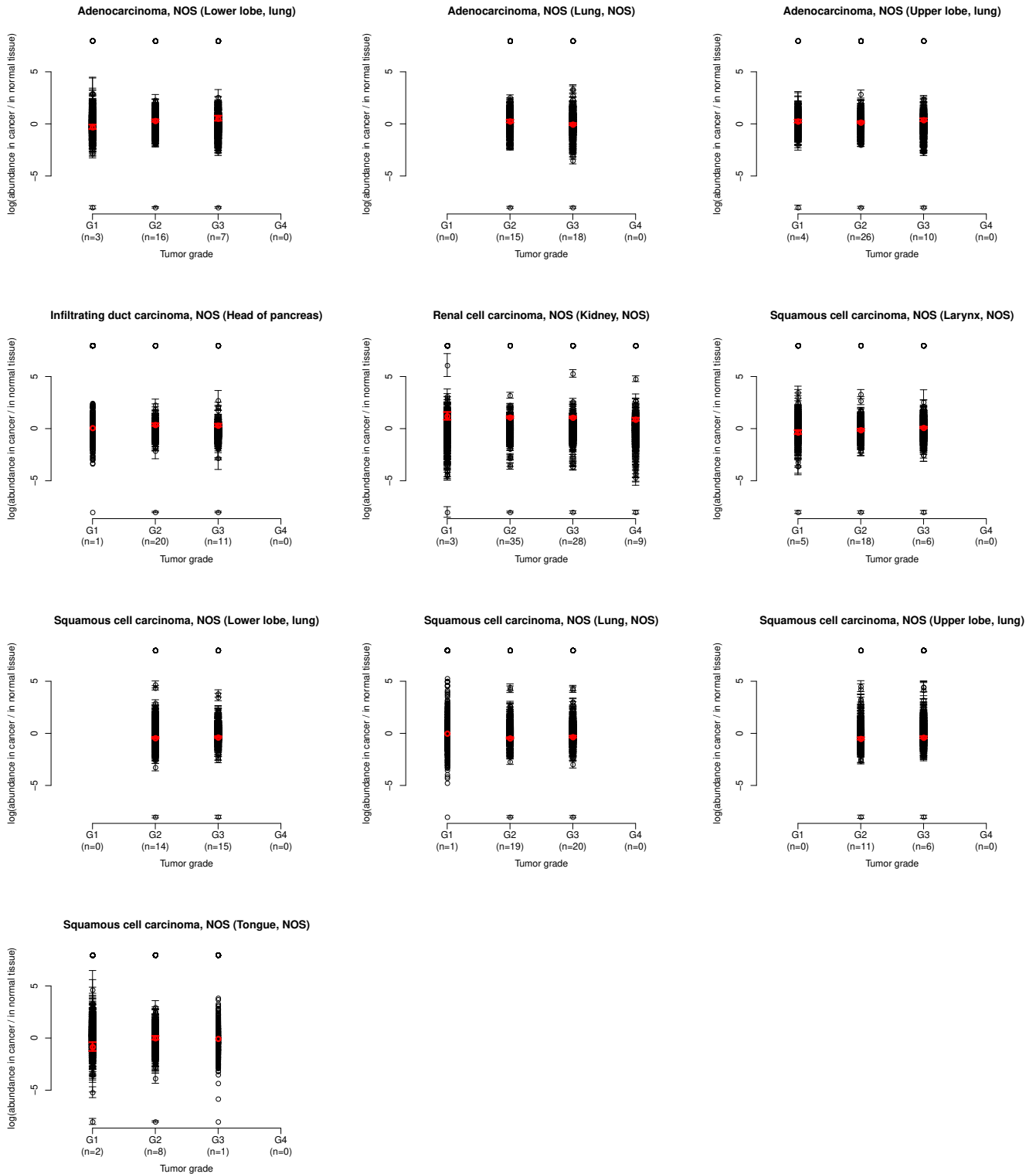
Sophie Mockly, Élisabeth Houbron and Hervé Seitz



Supplementary Figure S1: **Human and murine miR-34 family members.** Sequence alignment of the human (left) and murine (right) members of the miR-34 family. Nucleotides conserved between every family member are shown in orange. Nucleotides specific for the miR-34a/b/c subfamily are in red, those specific for the miR-449a/b/c subfamily are in green.

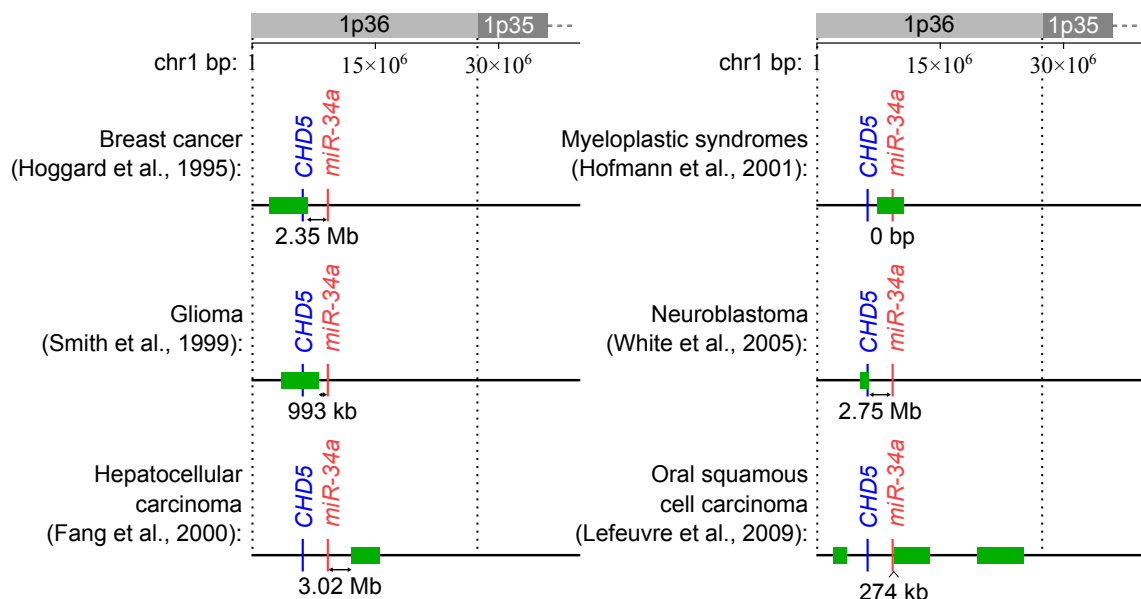


Supplementary Figure S2: **mir-34a is not generally mutated in cancers.** Only cancer types for which at least 100 cases were analyzed have been considered (n=30 cancer types; rows), and miRNA genes are shown in columns (n=1,750 hairpin loci in the top panel; 2,588 mature miRNA loci in the middle panel; 2,588 miRNA seed loci in the bottom panel). For each miRNA/cancer type pair, the heatmap shows the percentage of cases with recorded sequence variations either in the miRNA hairpin precursor (top panel), in the mature miRNA sequence (middle panel) or in the miRNA seed (bottom panel). miRNA loci were defined as in miRBase v.21. The column showing *mir-34a* data is magnified on the right margin (framed in black). “NOS”: not otherwise specified.

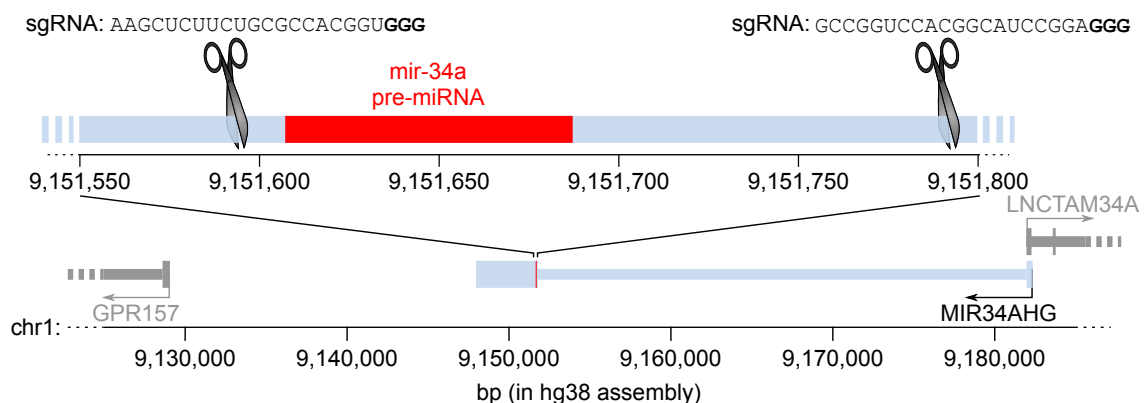


Supplementary Figure S3: **No evidence for miR-34a inactivation in the most aggressive tumors.** Cancer samples analyzed in Figure 1A were stratified by cancer grade (excluding cases whose grade was not determined, and selecting cancer types with at least 10 cases after this selection). miRNA abundance (normalized by the number of mapped miRNA reads) was compared between primary tumors and normal adjacent tissues. Each miRNA is represented by a dot (mean and standard error across independent cases are shown as a circle and error bars, respectively), with miR-34a being shown in red.  $\log(\text{fold-changes})$  larger than +8 or smaller than -8 were set to +8 or -8 respectively, for graphical clarity. “NOS”: not otherwise specified. The number of analyzed cases is indicated under the  $x$ -axis for each tumor grade.

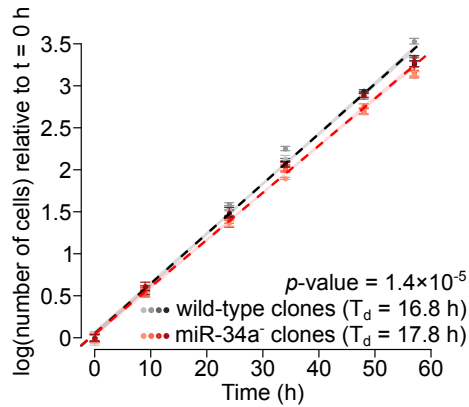




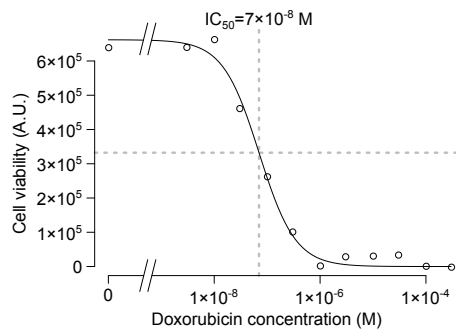
Supplementary Figure S4: **Smallest regions of consistent deletion identified in the 1p36 locus.** The 1p36 locus contains 885 annotated coding genes and 36 miRNA genes (Ensembl v.101); for clarity, only *CHD5* (in blue) and *miR-34a* (in red) are shown. Smallest identified region of consistent deletion are in green, and their distance to *miR-34a* is indicated under the map. Genomic coordinates relate to the hg38 assembly.



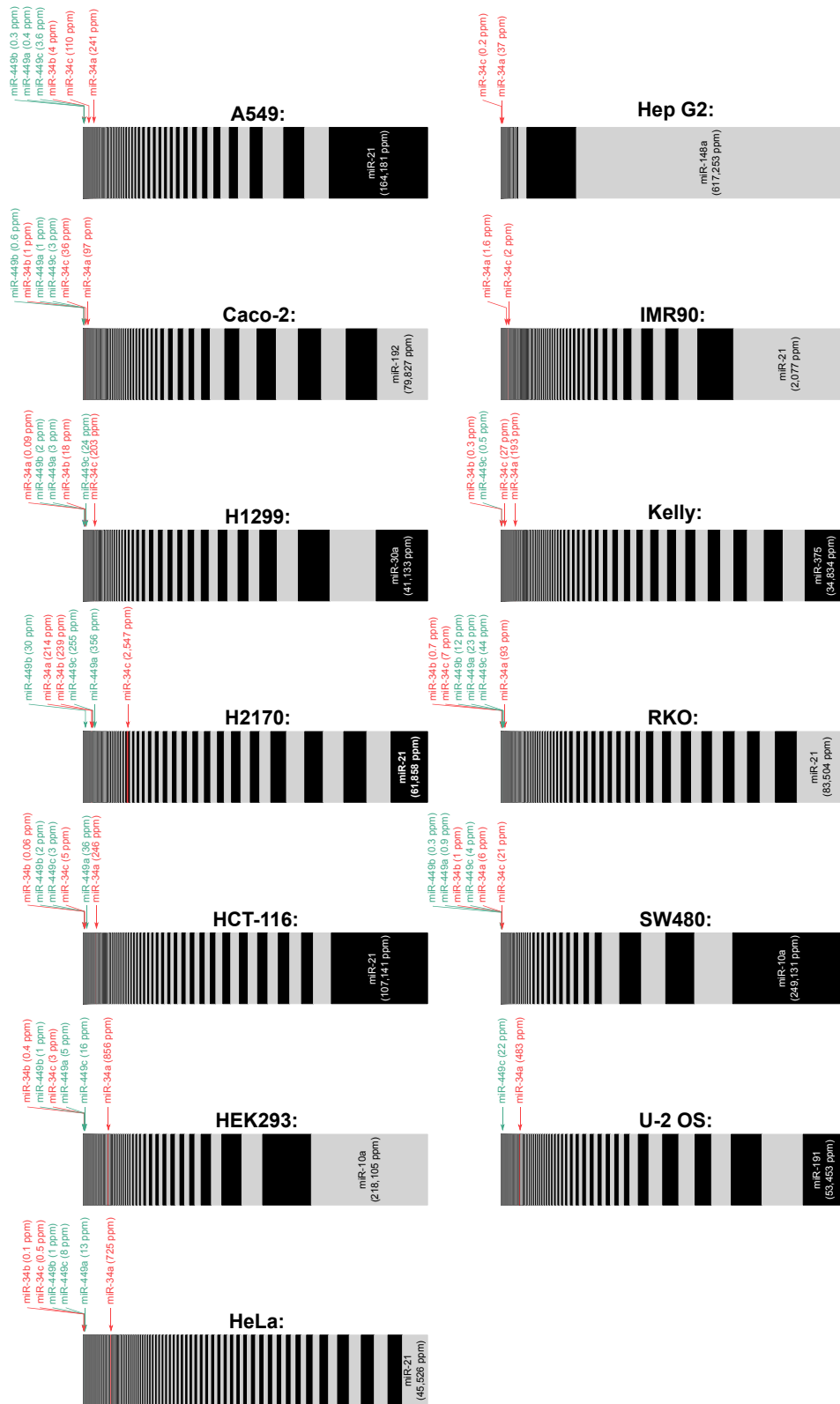
Supplementary Figure S5: **Strategy for CRISPR-mediated deletion of the miR-34a locus.** Thick lines: exons; thin lines: introns. Scissors represent Cas9-mediated cleavage sites, and the sequences of the cognate sgRNAs are written above each (PAM sequences in bold).



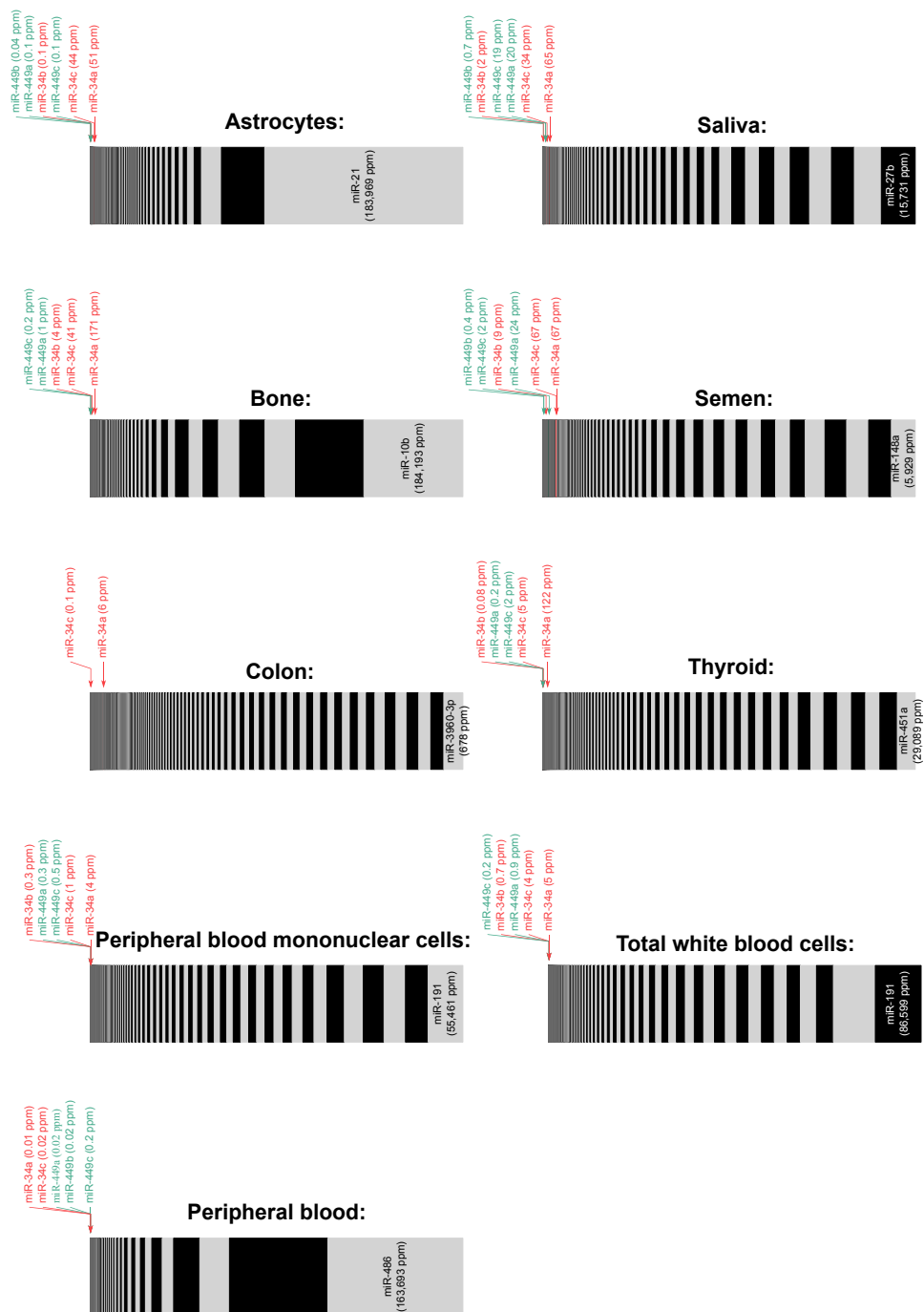
Supplementary Figure S6: **Proliferative effect of miR-34a in HAP1 cells.** Four wild-type and four *miR-34a* mutant clones were grown in sub-confluent conditions. Means and standard errors of 4 biological replicates are represented by dots and error bars. Linear modeling of log-transformed cell counts relative to time was used to measure doubling time ( $T_d$ ), and to estimate the significance of the effect of genotype ( $p$ -value is given in the inset). Shaded areas represent the 95% confidence interval for theoretical future measurements.



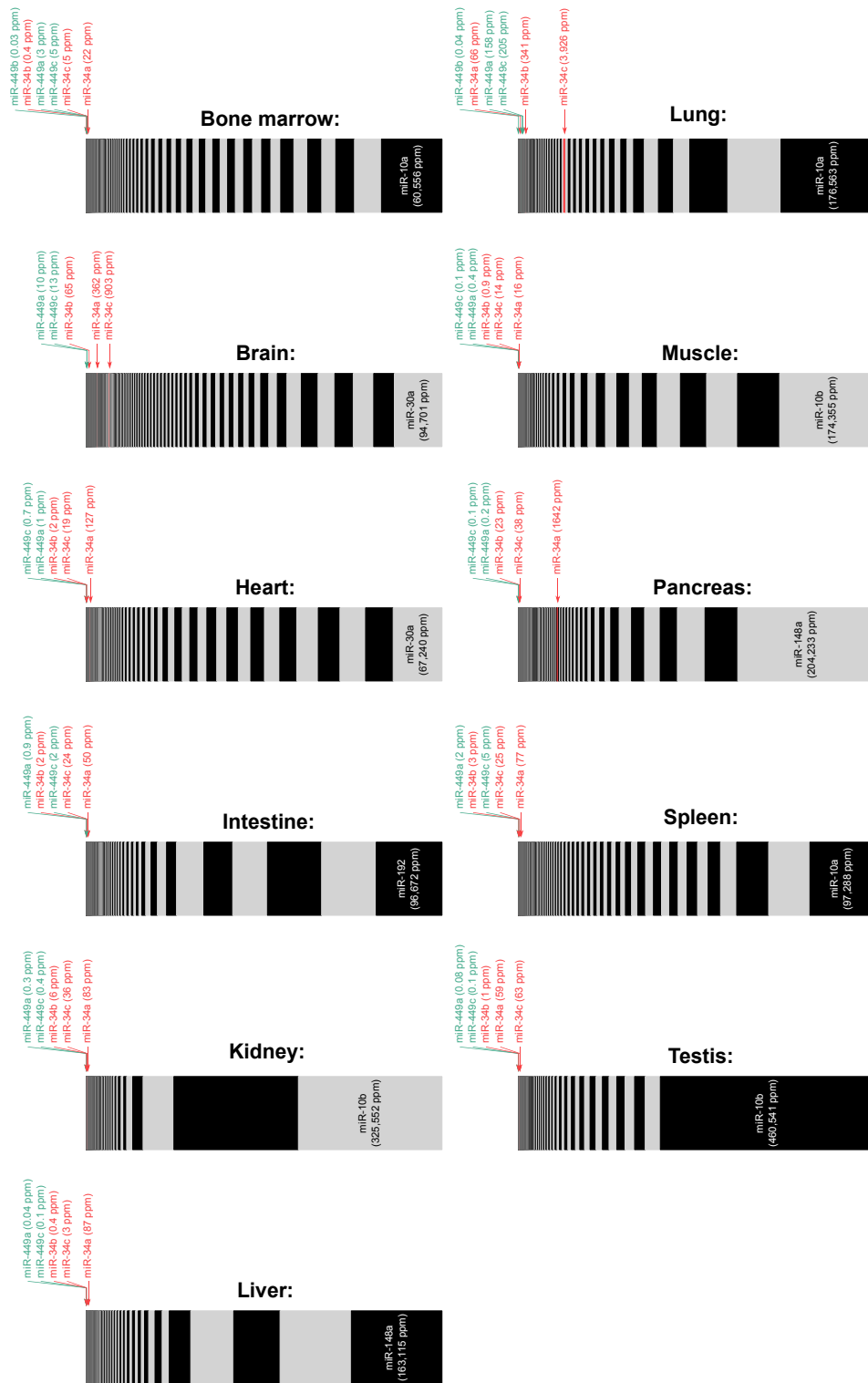
Supplementary Figure S7: **Determination of the half-maximal inhibitory concentration ( $IC_{50}$ ) of doxorubicin in HCT-116 cells.** Wild-type HCT-116 cells were grown in various concentrations of doxorubicin (shown on the  $x$ -axis), and cell viability was measured after 72 h using by ATP quantification (results shown on the  $y$ -axis).



Supplementary Figure S8: **miRNA abundance in human cell lines.** miRNAs are ranked by increasing abundance from left to right, and the width of each rectangle is proportional to miRNA abundance. Members of the miR-34 family, and their expression level, are shown in red and green (red for the miR-34a/b/c subfamily, green for the miR-449a/b/c subfamily). miRNA abundance is normalized to the total number of genome-matching reads, and expressed as *parts per million* (ppm). Small RNA-Seq datasets used in this figure are listed in Supplementary Table 2.



Supplementary Figure S9: **miRNA abundance in human tissues and body fluids.** miRNAs are ranked by increasing abundance from left to right, and the width of each rectangle is proportional to miRNA abundance. Members of the miR-34 family, and their expression level, are shown in red and green (red for the miR-34a/b/c subfamily, green for the miR-449a/b/c subfamily). miRNA abundance is normalized to the total number of genome-matching reads, and expressed as *parts per million* (ppm). Small RNA-Seq datasets used in this figure are listed in Supplementary Table 3.



Supplementary Figure S10: **miRNA abundance in murine tissues.** miRNAs are ranked by increasing abundance from left to right, and the width of each rectangle is proportional to miRNA abundance. Members of the miR-34 family, and their expression level, are shown in red and green (red for the miR-34a/b/c subfamily, green for the miR-449a/b/c subfamily). miRNA abundance is normalized to the total number of genome-matching reads, and expressed as *parts per million* (ppm). Small RNA-Seq datasets used in this figure are listed in Supplementary Table 4.

miRNA gene	Confidence	miRNA gene	Confidence	miRNA gene	Confidence
mir-4767	lower	mir-325	lower	mir-542	high
mir-4770	lower	mir-545	high	mir-450a-2	high
mir-651	high	mir-374a	high	mir-424	high
mir-548ax	lower	mir-421	lower	mir-450b	high
mir-3915	lower	mir-374b	high	mir-450a-1	high
mir-548f-5	lower	mir-1184-3	lower	mir-503	high
mir-6134	lower	mir-1184-2	lower	mir-4330	lower
mir-4666b	lower	mir-1184-1	lower	mir-452	lower
mir-23c	lower	mir-664b	high	mir-224	lower
mir-548am	lower	mir-6858	lower	mir-767	high
mir-4768	lower	mir-3202-1	lower	mir-105-1	high
mir-548aj-2	lower	mir-718	lower	mir-105-2	high
mir-1587	lower	mir-6087	lower	mir-891a	lower
mir-3937	lower	mir-548m	lower	mir-892c	lower
mir-4769	lower	mir-3672	lower	mir-891b	lower
mir-222	high	mir-766	high	mir-892a	lower
mir-221	high	mir-1277	lower	mir-888	lower
mir-98	high	mir-548an	lower	mir-890	lower
let-7f-2	high	mir-3978	lower	mir-892b	lower
mir-6857	lower	mir-652	high	mir-2114	lower
mir-6895	lower	mir-4329	lower	mir-514a-1	high
mir-6894	lower	mir-1912	lower	mir-514a-3	high
mir-8088	lower	mir-764	lower	mir-514a-2	high
mir-502	lower	mir-448	lower	mir-510	high
mir-660	high	mir-1298	lower	mir-509-3	high
mir-500a	high	mir-1911	lower	mir-509-1	high
mir-532	high	mir-320d-2	lower	mir-513b	lower
mir-501	lower	mir-504	high	mir-513c	high
mir-500b	lower	mir-934	lower	mir-513a-1	lower
mir-362	high	mir-505	high	mir-513a-2	lower
mir-188	high	mir-92a-2	high	mir-514b	high
mir-1468	lower	mir-106a	high	mir-508	high
mir-223	high	mir-363	high	mir-509-2	high
mir-361	high	mir-19b-2	high	mir-507	lower
mir-548i-4	lower	mir-20b	high	mir-506	high

Supplementary Table S1: **Frequently deleted miRNA genes in cancer.** Identity of the 105 miRNA genes frequently deleted in a variety of cancers (red area at the top left corner of Figure 1B). For each miRNA gene, its confidence level (as defined by miRBase v.21; [7]) is indicated.

<b>Cell line:</b>	<b>SRA accession number(s):</b>
A549	SRR6713501, SRR5689166, DRR036695 and SRR1304309
Caco-2	ERR3415707
H1299	DRR036697
H2170	SRR3341761 and SRR3341762
HCT-116	SRR954987, SRR954996, SRR4235725, SRR4235726, ERR3173398, ERR3173397 and ERR3173396
HEK293	SRR1240816 and SRR1240817
HeLa	SRR8311268 and SRR8311269
Hep G2	SRR12054851
IMR90	SRR020286
Kelly	SRR3533075, SRR3533074 and SRR3533073
RKO	ERR3415712
SW480	SRR3923807 and SRR3923808
U-2 OS	SRR10225092

Supplementary Table S2: **Small RNA-Seq datasets used in Supplementary Figure S8.** Several of these cell lines had been used in previous studies to evaluate the proliferative effect of miR-34 by over-expression experiments: A549 [8], H1299 [9], HCT-116 [8, 10, 11], Kelly [12], IMR90 [8], RKO [9], SW480 [13] and U-2 OS [9].

<b>Tissue or body fluid:</b>	<b>SRA accession numbers:</b>
Astrocytes	SRR2915342, SRR2915343 and SRR2915344
Bone	SRR6324194
Colon	SRR6895202, SRR6895203, SRR6895204 and SRR6895205
Peripheral blood mononuclear cells	SRR7412273–SRR7412275, SRR7412278, SRR7412280–SRR7412285, SRR7412296–SRR7412300, SRR7412302–SRR7412311, SRR7412313–SRR7412315 and SRR7412326–SRR7412334
Peripheral blood	SRR9844335–SRR9844346, SRR9844348, SRR9844349, SRR9844351–SRR9844360, SRR9844362, SRR9844364, SRR9844366–SRR9844373, SRR9844375 and SRR9844377–SRR9844386
Saliva	SRR3144036–SRR3144041, SRR3144044–SRR3144053, SRR3144055, SRR3144057 and SRR3144060
Semen	SRR11912557–SRR11912563 and SRR11912574
Thyroid	SRR8393464–SRR8393466
Total white blood cells	SRR7012343

Supplementary Table S3: **Small RNA-Seq datasets used in Supplementary Figure S9.**

<b>Tissue:</b>	<b>SRA accession numbers:</b>
Bone marrow	SRR10695972–SRR10695983 and SRR7807316–SRR7807327
Brain	SRR10695984–SRR10695997 and SRR7807328–SRR7807341
Heart	SRR10695998–SRR10696010 and SRR7807342–SRR7807354
Intestine	SRR10696011–SRR10696022 and SRR7807355–SRR7807366
Kidney	SRR10695896–SRR10695903, SRR10696023–SRR10696028 SRR7807237–SRR7807244 and SRR7807367–SRR7807372
Liver	SRR10695904–SRR10695917 and SRR7807245–SRR7807258
Lung	SRR10695918–SRR10695930 and SRR7807259–SRR7807271
Muscle	SRR10695931–SRR10695944 and SRR7807272–SRR7807285
Pancreas	SRR10695945–SRR10695957 and SRR7807286–SRR7807298
Spleen	SRR10695958–SRR10695971 and SRR7807299–SRR7807312
Testis	SRR10662083–SRR10662085 and SRR7807313–SRR7807315

Supplementary Table S4: **Small RNA-Seq datasets used in Supplementary Figure S10.**



## References

- [1] Hoggard, N., Brintnell, B., Howell, A., Weissenbach, J., and Varley, J. (1995) Allelic imbalance on chromosome 1 in human breast cancer. II. Microsatellite repeat analysis. *Genes Chromosomes Cancer*, **12**(1), 24–31.
- [2] Smith, J. S., Alderete, B., Minn, Y., Borell, T. J., Perry, A., Mohapatra, G., Hosek, S. M., Kimmel, D., O'Fallon, J., Yates, A., Feuerstein, B. G., Burger, P. C., Scheithauer, B. W., and Jenkins, R. B. (1999) Localization of common deletion regions on 1p and 19q in human gliomas and their association with histological subtype. *Oncogene*, **18**(28), 4144–4152.
- [3] Fang, W., Piao, Z., Simon, D., Sheu, J. C., and Huang, S. (2000) Mapping of a minimal deleted region in human hepatocellular carcinoma to 1p36.13-p36.23 and mutational analysis of the *RIZ* (*PRDM2*) gene localized to the region. *Genes Chromosomes Cancer*, **28**(3), 269–275.
- [4] Hofmann, W. K., Takeuchi, S., Xie, D., Miller, C. W., Hoelzer, D., and Koeffler, H. P. (2001) Frequent loss of heterozygosity in the region of D1S450 at 1p36.2 in myelodysplastic syndromes. *Leuk Res*, **25**(10), 855–858.
- [5] White, P. S., Thompson, P. M., Gotoh, T., Okawa, E. R., Igarashi, J., Kok, M., Winter, C., Gregory, S. G., Hogarty, M. D., Maris, J. M., and Brodeur, G. M. (2005) Definition and characterization of a region of 1p36.3 consistently deleted in neuroblastoma. *Oncogene*, **24**(16), 2684–2694.
- [6] Lefeuvre, M., Gunduz, M., Nagatsuka, H., Gunduz, E., Ali, M. A. S., Beder, L., Fukushima, K., Yamanaka, N., Shimizu, K., and Nagai, N. (2009) Fine deletion analysis of 1p36 chromosomal region in oral squamous cell carcinomas. *J Oral Pathol Med*, **38**(1), 94–98.
- [7] Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, **42**(Database issue), D68–D73.
- [8] He, L., He, X., Lim, L. P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., Jackson, A. L., Linsley, P. S., Chen, C., Lowe, S. W., Cleary, M. A., and Hannon, G. J. (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**(7148), 1130–1134.
- [9] Tarasov, V., Jung, P., Verdoodt, B., Lodygin, D., Epanchintsev, A., Menssen, A., Meister, G., and Hermeking, H. (2007) Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle*, **6**(13), 1586–1593.
- [10] Chang, T.-C., Wentzel, E. A., Kent, O. A., Ramachandran, K., Mullendore, M., Lee, K. H., Feldmann, G., Yamakuchi, M., Ferlito, M., Lowenstein, C. J., Arking, D. E., Beer, M. A., Maitra, A., and Mendell, J. T. (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell*, **26**(5), 745–752.
- [11] Tazawa, H., Tsuchiya, N., Izumiya, M., and Nakagama, H. (2007) Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci USA*, **104**(39), 15472–15477.
- [12] Welch, C., Chen, Y., and Stallings, R. L. (2007) MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene*, **26**(34), 5017–5022.
- [13] Bommer, G. T., Gerin, I., Feng, Y., Kaczorowski, A. J., Kuick, R., Love, R. E., Zhai, Y., Giordano, T. J., Qin, Z. S., Moore, B. B., MacDougald, O. A., Cho, K. R., and Fearon, E. R. (2007) p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr Biol*, **17**(15), 1298–1307.

## 3.4 Discussion

The first part of this thesis project was based on the study of the macroscopic impact of miRNAs, and in particular, consisted in participating in the debate concerning the definition of miRNA targets. Indeed, the notion of functional target is not consensual in the small regulatory RNA field, and while one school of thought considers molecular targets as functional, another reduces the potentiality of functional targets to genes for which repression at the molecular level generates a regulatory cascade at the origin of macroscopic responses. Consequently, among the multitude of molecular targets of a given miRNA, only a minority would represent functional targets, also called phenotypic targets to avoid any misunderstanding. In the context of the study of this second hypothesis, we were interested in strategies for the identification of molecular and phenotypic targets of miRNAs by participating, on the one hand, in the writing of a review presenting the methods currently used, as well as their limitations. On the other hand, we proposed a new methodology for analyzing functional targets of miRNAs in the cultured cell system.

Necessarily, the identification of phenotypic targets requires observing a phenotype and identifying miRNA targets that recapitulate the overall effect of the miRNA on this phenotype. Cell lines are a flexible and inexpensive system to study cellular responses in complex organisms; however, the number of observable cellular phenotypes remains limited. Therefore, we focused on cell proliferation, which has the advantage of being simple to quantify and suitable for mathematical modeling. We initiated this project with a miRNA known to be anti-proliferative in multiple cell lines, including in HCT-116 cells which are diploid and thus easily generate homozygous mutants.

However, during the preliminary study involving the quantification of the anti-proliferative effect of this miRNA in HCT-116, our results did not match the previous studies based on over-expression instead on deletion. Obviously, we first assumed that our results were erroneous and tried to resolve this doubt by carefully limiting the variability of behavior between our mutant and wild-type controls. Despite all our precautions, our results on HCT-116 never demonstrated an anti-proliferative effect of miR-34a. We finally contacted the only lab that created miR-34a-KO HCT-116 – by TALEN technology – to compare their proliferation ability (Navarro et al., 2015). Unfortunately, they could not find their cells, so we decided to pay no heed to their results and to investigate further.

For this purpose, we inspected the literature concerning miR-34a and were surprised to observe that the established tumor-suppressor function of this miRNA was essentially based on its anti-proliferative activity which is only demonstrated by miRNA over-expression experiments. Furthermore, a miR-34a-based cancer therapy was developed to replenish the levels of this miRNA in cancerous cells, where it is presumed to be absent or repressed. However, in such experimental or clinical studies, the introduced amount of artificial miRNAs may exceed physiological levels and only then appears to be anti-proliferative. Thus, even if the introduction of miR-34a is efficient as reducing solid tumors, it would not be longer a question of

solving a lack of miRNA – *i.e.* getting back to normal conditions – but of using a cytotoxic agent with unpredictable side effects. Specifically, overwhelming cellular expression of artificial miRNAs is likely to saturate the miRNA machinery and thus to impair the whole miRNA pathway with uncountable defects.

We thus halted the project on the identification of functional miRNA targets involved in cell proliferation to further investigate this so-called master tumor-suppressor miRNA. Indeed, our main concern was to verify whether miR-34a could be a general tumor-suppressor, like p53, since it is often presented as such.

First, we confirmed from the Cancer Genome Atlas Program data that miR-34a is not particularly underexpressed in numerous cancer types compared to adjacent healthy tissue, and its locus is neither particularly deleted nor mutated. This result questions the global character of this miRNA as a tumor suppressor, since a tumor-suppressor gene is partly defined as being mutated or silenced in cancers. However, we must raise the limit that, in the TCGA database, cancers are sorted by histological types of cancer, *i.e.* by organ and if possible by macroscopically differentiable subtype of cancer; yet there is no one type of cancer for each organ but a multitude of subtypes of cancer, even if histologically similar. These subtypes are rather defined by their mutation profiles or even the chronology of their mutations (Levine, 2020). Therefore, we can conceive that our bioinformatics analysis does not have the required granularity to identify cancer types in which miR-34a would be repressed or mutated: this subtype would be diluted in the other cancer subtypes of the same organ. Nevertheless, our results still confirm that miR-34a does not meet the definition of a general tumor suppressor.

We then demonstrated that miR-34a does not affect endogenous proliferation, even under genotoxic stress conditions, in the HCT-116 cell line that has been widely used to establish that miR-34a is anti-proliferative. Finally, to explain the discrepancies between our observation and the literature, we confirmed that ectopic expression of this miRNA is anti-proliferative and results in an intracellular concentration of artificial miRNA hundreds of times higher than a more natural induction (here: genotoxic stress by doxorubicin).

The objective of this project was to produce a pedagogical tool for the study of miRNAs in cancers and simultaneously correct the misassumption that miR-34a is anti-proliferative in HCT-116 cells. Especially, we aimed to demonstrate that, in spite of their overuse, miRNA overexpression studies do not enable to conclude alone about the biological function of a miRNA. Indeed, our results and others (Jin et al., 2015) reveal that overexpression by delivery of synthetic precursors or miRNA duplexes tends to reach supra-physiological intracellular miRNA levels which may lead to the disruption of the global miRNA pathway, among other defects, and result in severe developmental and cellular maintenance defects (*cf.* section 2.1.1). That is why, rigorous quantification of the induced intracellular miRNA concentration should be systematically required for overexpression assay.

Nevertheless, such measurement of miRNA molecules as presented in our work does not allow to distinguish non-functional from functional miRNA molecules *i.e.* miRNAs part of miRISCs and thus able to mediate silencing. To quantify the part of delivered miRNAs indeed

loaded into AGO proteins, we conceived of the measurement of the guide:passenger miRNA ratio from small RNA-seq data – the miRNA passenger strand is degraded after loading of the duplex into AGO – but such estimation would not be informative in case of saturation of the miRNA machinery in which a part of synthetic miRNAs would accumulate.

Interestingly, we observed that whether the introduction of artificial duplexes is cytotoxic in itself in high amounts, the extent of this toxicity is sequence-specific. While it could be argued that the most cytotoxic small RNAs are endogenously involved in cell survival or proliferation, we presume that increasing miRNA levels may affect novel low-affinity targets that are not repressed with lower miRNA concentrations (Bosson et al., 2014), leading, in an unpredictable way, to higher toxicity. Thus, such observation of a cytotoxic phenotype due to sequence-specific effect is in agreement with a saturation of the miRNA targets and consequently, confirms the functionality of transfected synthetic miRNA duplexes.

This work on miR-34a has been focused on its control of cell proliferation, considering apoptosis, senescence, cell cycle arrest, and survival as many mechanisms resulting in the alteration of cell number. However, tumorigenesis is not limited to the proliferative aspect of neoplastic cells. Other cellular features are involved in the establishment of cancer, including invasion and metastasis activation, angiogenesis induction, or else cellular plasticity (Hanahan et al., 2011). Notably, miR-34a has been shown to play a role in this latter cancer hallmark as it restricts the pluripotent cell fate potential of mouse embryonic or induced pluripotent stem cells and acts as a barrier to somatic reprogramming (Choi et al., 2011; Choi et al., 2017). Thus, just as we do not question miR-34a function in the ciliogenesis of lungs and reproductive organs, we do not refute miR-34a role in cell fate. Solely, we demonstrate that miR-34a is not endogenously anti-proliferative in HCT-116 and HAP1 cells, nor a potential general tumor-suppressor as it is not generally mutated or absent in cancers.

Eventually, in the meantime, we attempted our strategy of identifying functional miRNA targets involved in cell proliferation with a pro-proliferative miRNA, miR-21, without success. Indeed, the miR-21 locus deletion could not be achieved even with a robust mutagenesis protocol. Therefore, this miRNA is likely essential for the survival of HAP1 cells. Since examples of essential miRNAs are rare (see section 2.1.2), we could consider from this *ex vivo* observation that miR-21 would also be essential in vertebrate model organisms. However, homozygous miR-21 deficient mice are viable, fertile, and born in expected Mendelian ratios without any gross phenotypic differences. But interestingly, the deletion of miR-21 suppresses *K-ras*-induced tumor development *in vivo* (Hatley et al., 2010; Ma et al., 2011). Therefore, it is of matter to consider the limits of cancer cell lines for the study of endogenous miRNA functions and miR-21 is a prominent example. This miRNA is extensively over-expressed in human cancer cell lines (Volinia et al., 2006), as illustrated by its proportion in the total miRNA population from Figure S8 of our paper (*cf.* section 3.3.2). Consequently, the impact of highly-expressed miR-21 in cancer cell lines should not be necessarily expected from endogenously-expressed miR-21 in primary cells. In general terms, this means that endogenous miRNA

functions should be studied as far as possible in primary cells or model organisms, and immortalized and cancer cell lines should be saved for molecular mechanism or cancer studies. As an example, our intended CRISPR screen in HCT-116 and HAP1 cells would not genuinely recapitulate functional targets involved in cell proliferation in humans, but rather functional miRNA targets involved in cancer cells immortality and multiplication.

## Chapter 4

# Identification of Target-Directed microRNA Degradation Inducers

As illustrated in Chapter 2, the regulation of miRNA accumulation has been studied in an unbalanced manner, favoring a focus on the control of transcription and biogenesis. Thus, the understanding of miRNA turnover mechanisms was, until now, restricted to occasional observations of sudden decreases in miRNA levels correlated with miRNA 3'-end modifications in particular contexts. Since the serendipitous discoveries of endogenous Target-Directed miRNA Degradation<sup>1</sup> (TDMD) inducers in animals (Bitetti et al., 2018; Kleaveland et al., 2018), a big step forward has been taken in the understanding of the miRNA decay process as well as the miRNA:target RNA interaction. Until then, the interaction between a miRNA and its target was only observed through the eyes of miRISC-directed silencing or cleavage. It is now necessary to consider this interaction as reciprocal, according to the extent of complementarity. Furthermore, just as canonical miRNA targets have been extensively sought, systematic research of TDMD inducers must be conducted.

To date, the Nicassio lab has reported the only example of a TDMD inducer identified on purpose (Ghini et al., 2018). They took advantage of the fibroblast-serum model – mouse embryonic fibroblasts can be induced to quiescence by serum deprivation and then stimulated to re-enter cell cycle by serum addition – to perform an *in vivo* time course RNA labeling and to measure transcriptional changes associated with the cell cycle. Then, by cross-referencing miRNA target changes in expression during cell cycle re-entry and 3'-complementarity score from TargetScan, their results bring a potential TDMD inducer:miRNA pair out: the Serpine1 transcript and miRNAs miR-30b/c. Eventually, they provided evidence that the Serpine1:miR-30b/c interaction triggers miR-30 degradation through TDMD. Notably, Serpine 1 expression induced by serum stimulation is negatively correlated with miR-30b/c abundance and small mutations in Serpine 1:miR-30b/c pairing site disrupts miR-30b/c decay.

This part of my thesis aims to contribute to this research effort and actively search for TDMD inducers by using a combination of computational analyses and experimental validation *ex vivo* and, in the long term, to understand the rules distinguishing TDMD inducers from canonical miRNA targets.

---

<sup>1</sup>The state-of-the-art considering the mechanism of TDMD is introduced in section 2.2.5.

## 4.1 Development of a TDMD Inducer Prediction Computational Approach

The mechanism underlying TDMD requires extensive pairing between a target and the 3′-end of the miRNA. Indeed, such pairing may be crucial for TDMD by generating a miRISC conformation change favorable for ZSWIM8 interaction, followed by the recruitment of Cullin-RING E3 ubiquitin ligase complex members and the degradation of AGO by the proteasome. Therefore, the distinction between canonical miRNA targets and TDMD-inducing targets – simplified *TDMD inducers* – primarily depends on the pairing geometry. Inspired by computational prediction tools for canonical miRNA targets which search for the presence of conserved ~7mer sites that match the seed region of each miRNA (Lewis et al., 2005), we plan to create a prediction tool for TDMD inducers, searching for TDMD-inducing pairing patterns in regions that have been selectively conserved in evolution. For this purpose, we focused our research on vertebrates since (i) this clade is sufficiently deep to score sequence conservation reliably, (ii) provides an extensive amount of genome sequence data (iii) and of known miRNAs, and (iv) known endogenous TDMD events have been reported in murine and human cells.

In broad strokes, our workflow comprises three phases. It searches TDMD-inducing pairing patterns for each miRNA in the corresponding species genome. Then, it scores conservation of recorded sites. Finally, it provides a selection of promising TDMD inducer candidates in a cell type or tissue of interest according to their expression and their TDMD-triggering site conservation score.

### Preliminary step: Establishment of TDMD-inducing pairing rules

Despite all previously described TDMD inducers, there is no exact consensual definition of the TDMD-inducing pairing. Therefore, we established standard features from all examples – artificial, viral, and endogenous TDMD inducer:miRNA pairs – reported in the literature. Precisely, we evaluated manually 30 examples of active TDMD inducers and 7 examples of inactive sites from Park et al., 2017 and every publication cited in section 2.2.5. As a result, we defined two TDMD-inducing patterns introduced in Figure 4.1.

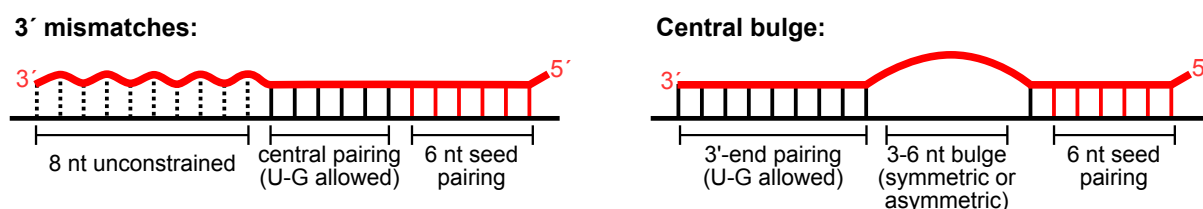


FIGURE 4.1: **Patterns of TDMD-inducing pairing geometry.** The miRNA sequence is represented by the red strand and the target inducer by the black one. 3′ mismatches complementarity: allows mismatches for the last 8 nucleotides at the 3′ end and for the first nucleotide at the 5′ end of the miRNA. Central bulge complementarity: allows a symmetric or asymmetric bulge of 3 to 6 nucleotides just after the seed pairing and followed by a perfect pairing at the 3′ end.

In details, for the 3' mismatches complementarity pattern, the miRNA must be perfectly paired on its nt 2–7 (without GU wobble), then paired (GU wobbles allowed) on any number of nucleotides, then possibly unpaired on the 3' end (no longer than on the 8 last nt). For the central bulge complementarity pattern, two possibilities are accepted: **(i)** the miRNA is perfectly paired on its nt 2-7 (without GU wobble) then paired (GU wobbles included) on nt 8, then unpaired on nt 9 and on the 1–5 next nt, then paired (GU wobbles included) from nt 15 to the end; **(ii)** the miRNA is perfectly paired on its nt 2-7 (without GU wobble) then paired (GU wobbles included) on nt 8-10, then unpaired on nt 11 and on the 1-3 next nt (if only nt 11 and 12 are unpaired, then the inducer has to have 7 or 8 unpaired nt on the other strand; otherwise, as little as 3 unpaired inducer nt are accepted), then paired (GU wobbles allowed) from nt 15 to the end.

In each case, GU wobbles are not allowed in the seed because we hypothesize that TDMD pairing results from the same mechanism of canonical miRNA target recognition by nucleation from the seed region, thus requiring perfect Watson-Crick complementarity.

#### **Step 1: Computational TDMD-inducing pattern search in vertebrates**

In order to localize TDMD-inducing sites, multiple strategies could be applied. The least elaborate would be to search for TDMD-compatible pairings directly in whole genomes for every known mature miRNA sequence in vertebrates. However, since only deeply conserved candidates would be ultimately considered, such unrestricted sequence analysis appears to be a waste of time. For instance, Ghini and colleagues based their TDMD inducer identification strategy on predicted miRNA:target pairs from TargetScan and took advantage of prior TargetScan conservation and 3' pairing contribution (3C-score) evaluations. Similarly, we could conceive to investigate only established conserved genomic sequences – instead of whole genomes – to spare unnecessary calculations. Nevertheless, we could imagine at least two scenarii for which such strategy would not be appropriate. **(i)** TDMD-compatible sites for a given mature miRNA in species X and conserved in species Y are relevant only if an identical mature miRNA does exist in species Y; or **(ii)** mature miRNAs and their TDMD inducer sites could diverge across species in a coevolution manner. Eventually, to consider any eventualities, we decided to favor the conservation of TDMD-inducing pairings before the conservation of TDMD-inducing sites. Thus, we impartially analyzed whole-genome multiple alignments centered on the human genome to identify TDMD-inducing pairings in multiple species independently.

Our workflow starts from two inputs: the Fasta format sequences of all mature miRNA sequences from miRBase (<https://www.mirbase.org/ftp.shtml>) and the multiple alignments of 99 vertebrate genomes with the human genome from UCSC (<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz100way/>). Because miRNAs have not been identified in all vertebrates for which the genome sequence is available, we first created an inner join table to extract the common species names between the two inputs. In the miRBase input, the miRNA species is indicated by its Latin name as well as its three letters prefix, while in the



UCSC input, only the genome assembly name is specified. Thus, we extracted the corresponding Latin name for each genome assembly name from the README of the multiple alignment file and using Latin names as key values, we joined miRBase prefixes and genome assembly names, as illustrated in Figure 4.2.

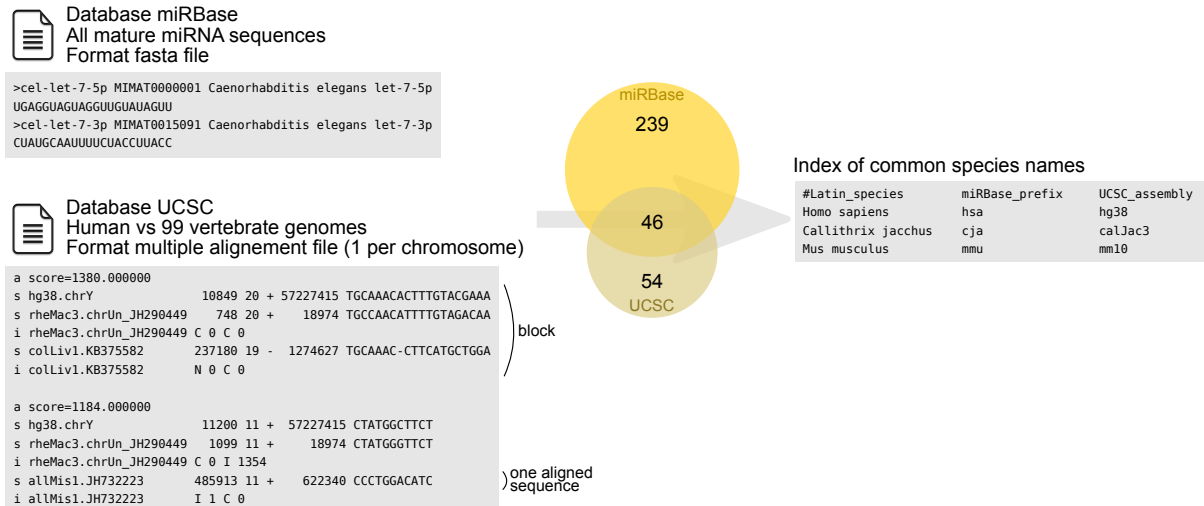


FIGURE 4.2: **Input files pre-processing.** Left: The miRBase fasta file of all mature miRNA sequences gathers miRNAs from 285 species, mainly metazoans, but also viridiplantae, mycetozoans, viruses, alvelolatans and chromoalveolatans (<https://www.mirbase.org/cgi-bin/browse.pl> for more details). A part of the miRBase fasta file is represented. The UCSC multiple alignment file applies to the human genome with 99 vertebrate species. A part of the chromosome Y maf file is represented. Middle: Venn diagram with the number of common species between the two files and unique to each file. Right: A part of the common species name index used in our workflow to link miRNA species with aligned genome species.

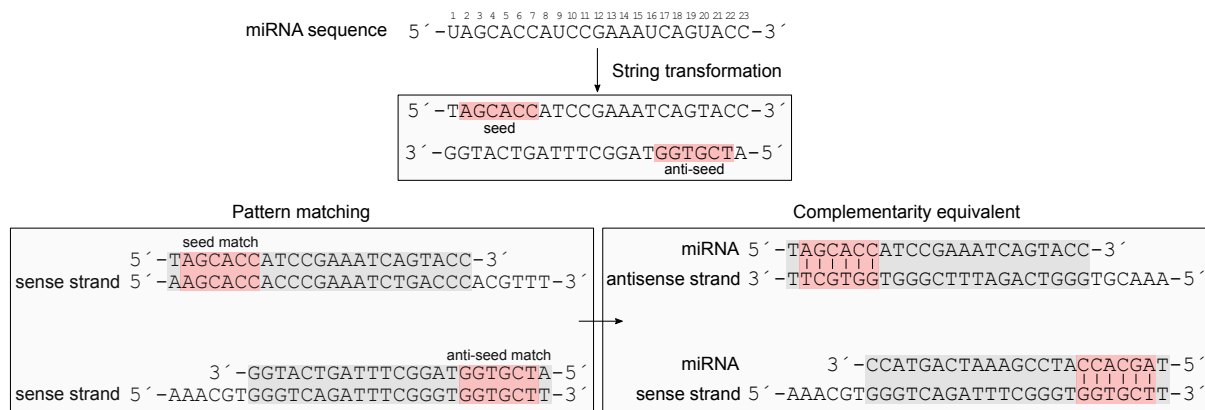
To minimize the calculations of our program, for each miRNA in the 46 species of interest, the following data are stored prior to searching for pattern matching: miRNA ID, miRNA name, species, seed sequence, and reverse complement of seed sequence. All Us are replaced by Ts to allow direct comparison with genome data. Then, unified lists for the seeds and their reverse complement sequences are created since miRNAs exhibit numerous paralogs and orthologs across vertebrates sharing a same seed sequence. In practice, for one seed sequence, we keep a trace of all miRNAs sharing this seed in a hash table: the key is the seed sequence and the corresponding value is a vector of arrays of which each comprises the whole sequence of the miRNA, its name and its species.

Multiple alignment files are divided into 25 files, one per chromosome: 1–22, X, Y, and mitochondrial DNA. The Multiple Alignment Format (maf) stores a series of multiple alignments at the DNA level between entire genomes in a format that is easy to parse. We will refer to individual multiple alignments as *blocks* and each sequence from the block alignment as an *aligned sequence*. Because some blocks contain only the human sequence or consist of aligned sequences shorter than the minimal miRNA length (20 nt), we first clean maf files from such blocks. Our program reads remaining blocks, one by one, and temporarily stores aligned sequences with their corresponding species, chromosome, genomic coordinates, strand, length

#### 4.1. Development of a TDMD Inducer Prediction Computational Approach

sequence, and sequence data in an array. Each sequence is then interrogated in turn, and to facilitate the search of patterns, all sequences are standardized by removing dashes and capitalizing all characters.

For each aligned sequence, our program first tests a match to every seed or reverse complement seed – hereafter, *antiseed* – from the unified lists. To note, a match between a seed and an aligned sequence means that this seed is complementary to the opposite strand of this aligned sequence. In this case, the seed sequence is oriented in 5' to 3' and the substring to compare with TDMD-inducing patterns is : 1 nt before the seed match + the seed match + 14 nt after the seed match. Reciprocally, a match between an antiseed and the aligned sequence means that the seed is complementary to the aligned sequence. In this case, the seed is oriented in 3' to 5' and the substring to test is : 14 nt before the seed match + the seed match + 1 nt after the seed match. The rationale behind such string comparison is illustrated in Figure 4.3. In the text below, only the case of seed match is detailed to ease the method description.



**FIGURE 4.3: Rationale behind miRNA seed match search in sense and antisense strands.** MiRNA sequences are first transformed by replacing Us by Ts. Seed and reverse complement seed subsequences are then extracted to test a match within sequences from the multiple alignment files. A match to the seed sequence indicates that the miRNA is complementary to the antisense strand while a match to the antiseed sequence indicates that the miRNA is complementary to the sense strand.

For each match between a seed and an aligned sequence, all miRNAs sharing this seed and from the same species than the aligned line, are investigated. To find TDMD-inducing patterns, our program tests if the aligned sequence matches a regular expression including the seed sequence and a supplementary subsequence. The miRNA subsequences from nt 8 to the 8th before the miRNA end are named *supplementary MM match* – for 3' mismatches match –, and from nt 15 to the miRNA end are named *supplementary CB match* – for central bulge match. To consider GU wobble pairing, variants are generated by replacing Gs by As and Ts by Cs in all possible combinations. Thus, aligned sequences matching the pattern “seed + supplementary MM match” or “seed + 4-9 nt + supplementary CB match” are recorded as potential 3' mismatches or central bulge hits, respectively.

Each hit is finally tested by RNAduplex (<https://www.tbi.univie.ac.at/RNA/RNAduplex-1.html>) with its cognate miRNA. This function from the ViennaRNA package (<https://www.tbi.univie.ac.at/RNA/index.html>) computes the structure upon hybridization of two RNA strands. By providing the miRNA sequence and the extended TDMD target sequence – in this case, the reverse complement of the recorded matching sequence with 6 extra nt – it is possible to verify that the most stable secondary structure of the duplex fits the 3' mismatches or central bulge complementarity pattern. Especially for central bulges, if the primary structure fits the pattern, it does not assure that the RNA species will spontaneously fold within a central bulge. RNAduplex produces dot-bracket notation of an RNA secondary structure composed by a three-character alphabet: “.” codes for an unpaired base, “(” an open base pair, and “)” a closed base pair. However, it does not distinguish between Watson-Crick pairing and GU wobbles. Therefore, dot-bracket notation of the intermolecular folding between the miRNA and its target is translated in numeric notation of base pairing for each strand: “0” codes for unpairing, “1” a GU wobble, and “2” a Watson-Crick pairing. This workflow is illustrated in Figure 4.4.

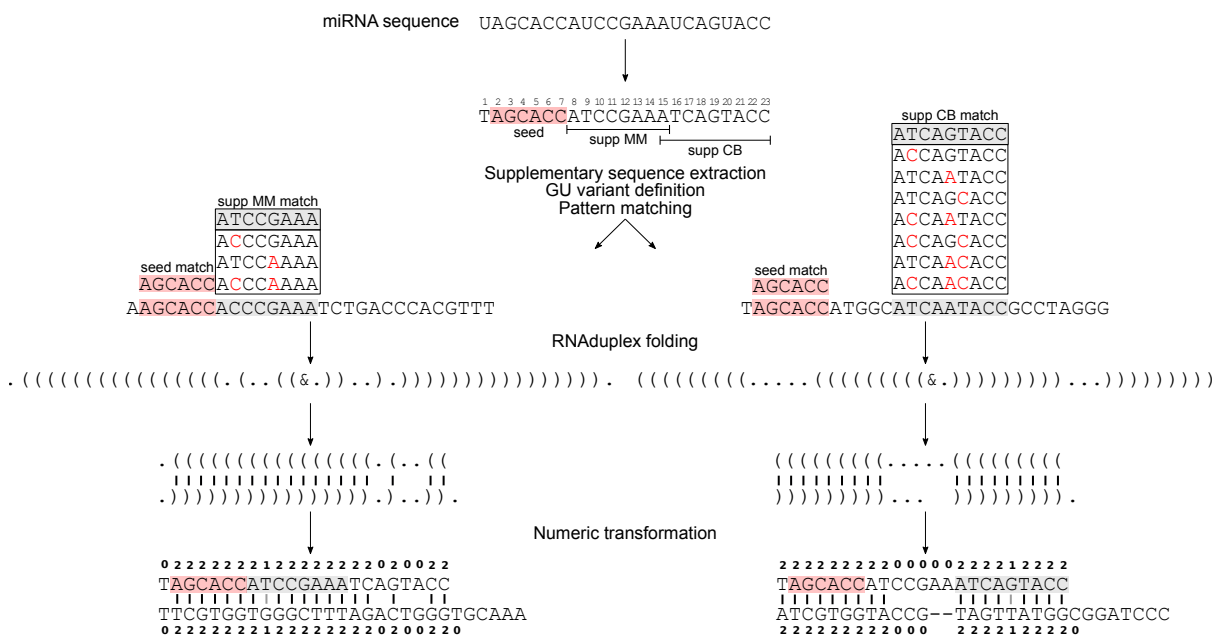


FIGURE 4.4: **Schematic of 3' mismatches and central bulge complementarity identification workflow.** MiRNA sequences are first transformed by replacing Us by Ts. Seed, supplementary MM and supplementary CB subsequences are extracted to test a match within sequences from multiple alignment files. For each match, the duplex structure of the reverse complement of the target and the miRNA sequences is predicted by RNA duplex which generates a dot-bracket format. An indeterminate step is represented to illustrate how this dot-bracket notation is interpreted: bold black lines indicate Watson-Crick base-pairing and bold grey lines GU wobbles. The dot-bracket RNA duplex notation is finally converted in a numeric base-pairing code for both miRNA and target sequences.

At the end of this step, two reports are generated, one for each TDMD-candidate pattern. They comprise all 3' mismatches and central bulge hit sites, including the candidate site genomic localization, its sequence, its pairing code, as well as its cognate miRNA sequence and

pairing code.

##### **Step 2: TDMD-inducing site conservation score in vertebrates**

Our program scores **(i)** the conservation of a human TDMD-compatible sequence in all vertebrates through the identity and number of aligned genomic sequences in the block and **(ii)** the conservation of a TDMD-compatible pairing in all vertebrates through the identity and number of recorded TDMD-compatible sites in other species at the same localization according to multiple alignment data. These two distinct conservation scoring methods were respectively named: the blind and the informed strategies.

Concretely, the blind strategy confirms human TDMD-candidate site conservation within its alignment block regardless of whether these other species possess a miRNA that could bind it. Thus, for each human 3' mismatches or central bulge hit, nt 2-12 or nt 2-7 and 15-20 are extracted, respectively, and tested if the subsequence is found at the same position in another species in each other species aligned line. For the informed strategy, we took advantage that each alignment block starts by the human reference line to test, for each non-human 3' extensive or central bulge hits, if a human TDMD-candidate site had been localized on the human line, and if the nucleotide facing miRNA nt 1 is at the same position.

Such conserved TDMD-candidates are recorded in two other reports, one for each conservation strategy. They comprise all 3' mismatches and central bulge conserved sites for the blind strategy and conserved TDMD-inducing pairing for the informed strategy, at least between humans and another species. The report records conservation by species pairs: the human TDMD-candidate site, its genomic localization, its cognate miRNA, the other species name and localization of the conserved site. At this step, TDMD candidate conservation is recorded but not scored yet.

Ideally, the informed strategy would be preferred over the blind strategy since we explained in subsection 4.1 that we aimed to favor the conservation of TDMD-inducing pairings before the conservation of TDMD-inducing sites. However, the informed strategy relies on identifying TDMD-compatible sites in various vertebrates and thus is limited by the state of discovery about miRNA sequences in these species. Consequently, we inevitably miss TDMD-inducing sites in most non-model vertebrates for which a low amount of miRNAs have been described, thus undeniably increasing the part of false-negative results. On the contrary, because the blind strategy consists of scoring sequence conservation regardless of whether other species possess complementary miRNAs, it is thus prone to false-positive results, for example, in the most conserved regions of the genome. Consequently, the issue would be to find a balance that minimizes both false-positive and false-negative results. For want of anything better, we decided to record outputs from both strategies and preferred the informed strategy stringency in a second phase for experimental validation of the best candidates.

The simplest way to score site conservation would be to count the number of species in which this site has been reported as conserved, according to one of our strategies of conservation evaluation. Nevertheless, from an evolution point of view, a sequence conserved sporadically is not equivalent to a sequence specifically conserved in one clade. To take into consideration the conservation distribution across vertebrates, we created a score of clade-specificity: a site is specific to a phylogenetic clade when present in >50% of species in the clade and absent in every species outside the clade. For this purpose, we manually created a phylogeny file that details the taxonomic classification of each studied species. For example, the chicken *Gallus gallus* belongs to the following taxa: Metazoa, Bilateria, Deuterostoma, Chordata, Vertebrata and Aves. Thus, for each TDMD inducer candidate, it is possible to examine in which species it is conserved and establish its clade specificity.

### **Step 3: TDMD-inducing candidate selection in a tissue of interest**

The previous steps permit the prediction of all potential TDMD-inducing sites in vertebrate genomes and to assign them a conservation score. Our workflow brings out the most conserved and expressed TDMD inducer candidates in a tissue of interest.

First, to limit false-positive candidates, only conserved candidates from the informed strategy were investigated and TDMD candidates localized in genomic region particularly conserved in an obvious miRNA-independent manner – *i.e.* localized in exons of mRNA open reading frames (ORF) – were evicted. Genomic annotation information was obtained from the human NCBI RefSeq gene (coding and non-coding) annotation. This file indicates RefSeq accession numbers of genes – and indirectly the molecule types since mRNAs, for example, begin with the prefix XM\_ if predicted and NM\_ if curated – as well as their coordinates, ORF coordinates, exon number and coordinates.

Secondly, the expression of transcripts hosting TDMD candidate sites is evaluated in a tissue of interest. Our workflow cross-references RNA-seq data with the results of our TDMD inducer site prediction algorithm. To distinguish the most promising candidate sites, it ranks all RNA:miRNA pairing sites by clade specificity and candidate inducer RNA expression.

## 4.2 TDMD Inducer Candidates in Murine Neuronal Cells

In order to verify the prediction efficiency of our algorithm, we focused on a cell type that could be used for experimental validation. It has been shown that the TDMD mechanism is more efficient in primary neurons than in non-neuronal cells or induced neuronal cell lines (de la Mata et al., 2015). We therefore set up a collaboration with the Perroy laboratory (Institute of Functional Genomics, Montpellier) which studies the pathophysiology of synaptic transmission and could provide us in primary neuronal culture. We agreed on an experimental validation with mouse cortical neurons which can be easily collected and in sufficient quantity for our experiments.

Therefore, we cross-referenced the results of our TDMD inducer site prediction algorithm with RNA-seq data of mouse cortical neurons from different laboratories available on NCBI SRA (control samples in SRA study *SRP150077*). To distinguish the most promising candidate sites, TDMD inducer:miRNA pairs are plotted according to their score of pairing conservation and the TDMD inducer RNA host expression in murine cortical neurons (Figure 4.5).

The most promising TDMD inducer candidates are the most conserved and expressed TDMD-compatible sites. First of all, our results recognize the two already reported neuronal TDMD pairs: Nrep:miR-29b and Cyrano:miR-7 (Bitetti et al., 2018; Kleaveland et al., 2018), which are positive controls in our analysis. Then, we colored 3 candidate TDMD inducer:miRNA pairs and one positive control in red for further experimental validation, Nnat:miR-708, Mapre2:miR-1a, Ppp1cb:miR-23b and Nrep:miR-29b. *Nnat* or *Neuronatin* is an imprinted gene exclusively expressed from the paternally inherited allele (Kagitani et al., 1997) and enriched in the developing brain (Joseph et al., 1994). It encodes a proteolipid that may be involved in the regulation of calcium channels during brain development (Dou et al., 1996; Lin et al., 2010; Oyang et al., 2011). *Mapre2*, also known as *EB2* or *RP1*, encodes a microtubule-associated protein (Su et al., 2001) that is necessary for spindle symmetry during mitosis (Brüning-Richardson et al., 2011; Iimori et al., 2016) and for the establishment of cell polarity (Goldspink et al., 2013). *Ppp1cb* encodes for one of the three catalytic subunits of protein phosphatase 1 (Barker et al., 1994) and its mutation may participate in a RASopathy – genetic disorder caused by mutations in multiple genes encoding components of the RAS/MAPK pathway – similar to the Noonan syndrom (Bertola et al., 2017; Gripp et al., 2016; Zambrano et al., 2017). These 3 TDMD inducer candidate sites are localized in 3'-UTRs, and Nnat:miR-708 and Ppp1cb:miR-23b present a 3' mismatches pattern, while Mapre2:miR-1a present a central bulge pattern. In Figure 4.6, the multiple alignments of each site show their sequence conservation in the 100 vertebrates. The Mapre2 TDMD candidate is well conserved in vertebrates, except in some fishes, likewise the Nrep TDMD inducer. The Ppp1cb TDMD candidate is also globally well conserved in amniotes, *i.e.* in all vertebrates in the alignment except xenopus and fishes. Finally, the Nnat TDMD candidate is conserved only in placental mammals and, interestingly, its cognate miRNA miR-708 is also conserved only in placental mammals, suggesting co-selection between the miRNA and its regulator.

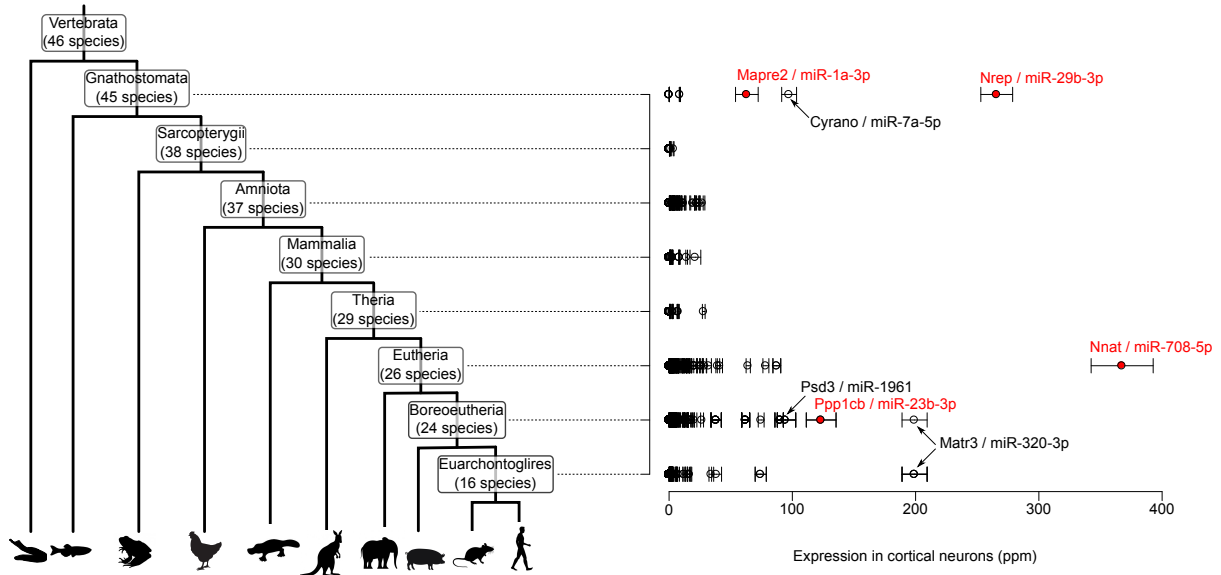


FIGURE 4.5: TDMD inducer candidates in mouse cortex. TDMD site:miRNA pairs are represented according to their conservation and their abundance in murine cortical neurons. In red are shown selected pairs for experimental validation. X-axis: expression of TDMD site host RNA in ppm (mean +/- st. error). Y-axis: ranking of TDMD sites by clade specificity.

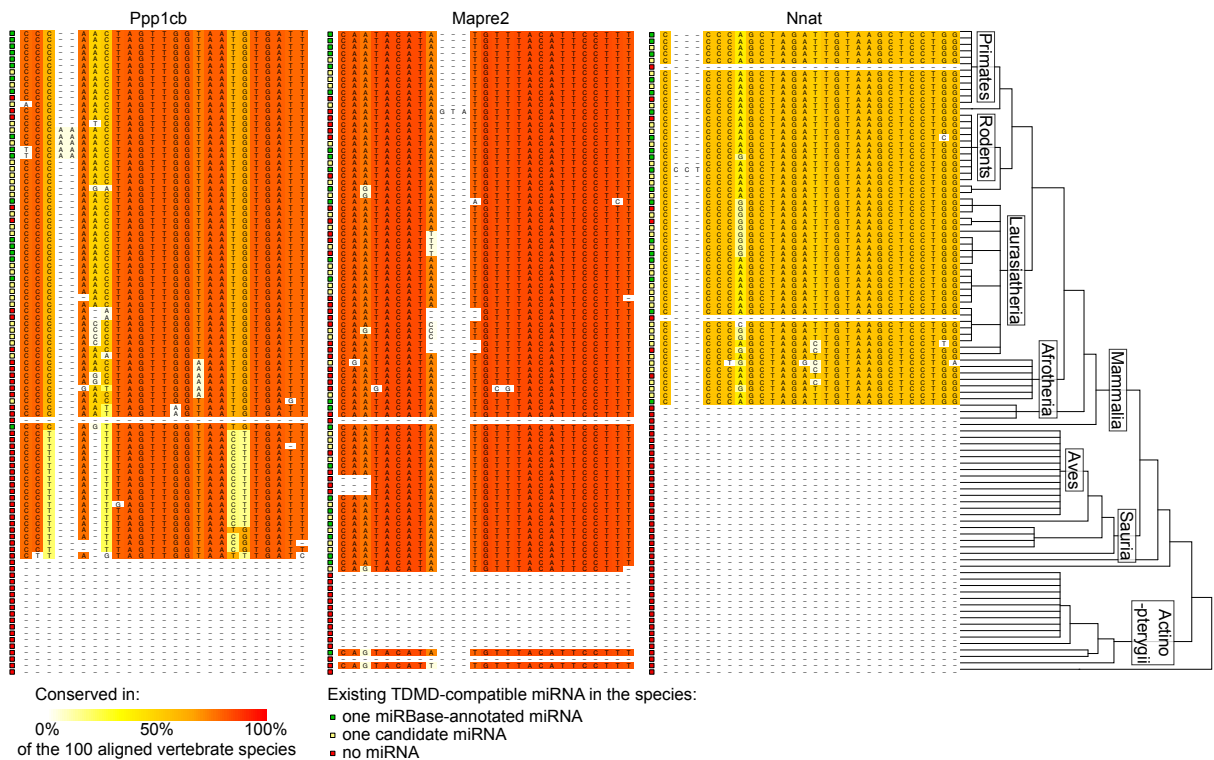


FIGURE 4.6: Multiple alignment of TDMD inducer candidates in vertebrates. From left to right, Ppp1cb, Mapre2 and Nnat TDMD candidate sites from 100 vertebrates are aligned and ranked according to a phylogenetic tree represented on the right. The percentage of identity conservation of a nucleotide at each position in the 100 vertebrates is indicated by a color ranging from yellow (0%) to red (100%). The colored squares on the left of each sequence indicate whether an annotated TDMD-compatible (green), non-annotated (yellow) or no miRNA (red) is identifiable for this sequence in this species.

### 4.3 Experimental Validation of AGO2-bound miR-708 Cleavage Activity

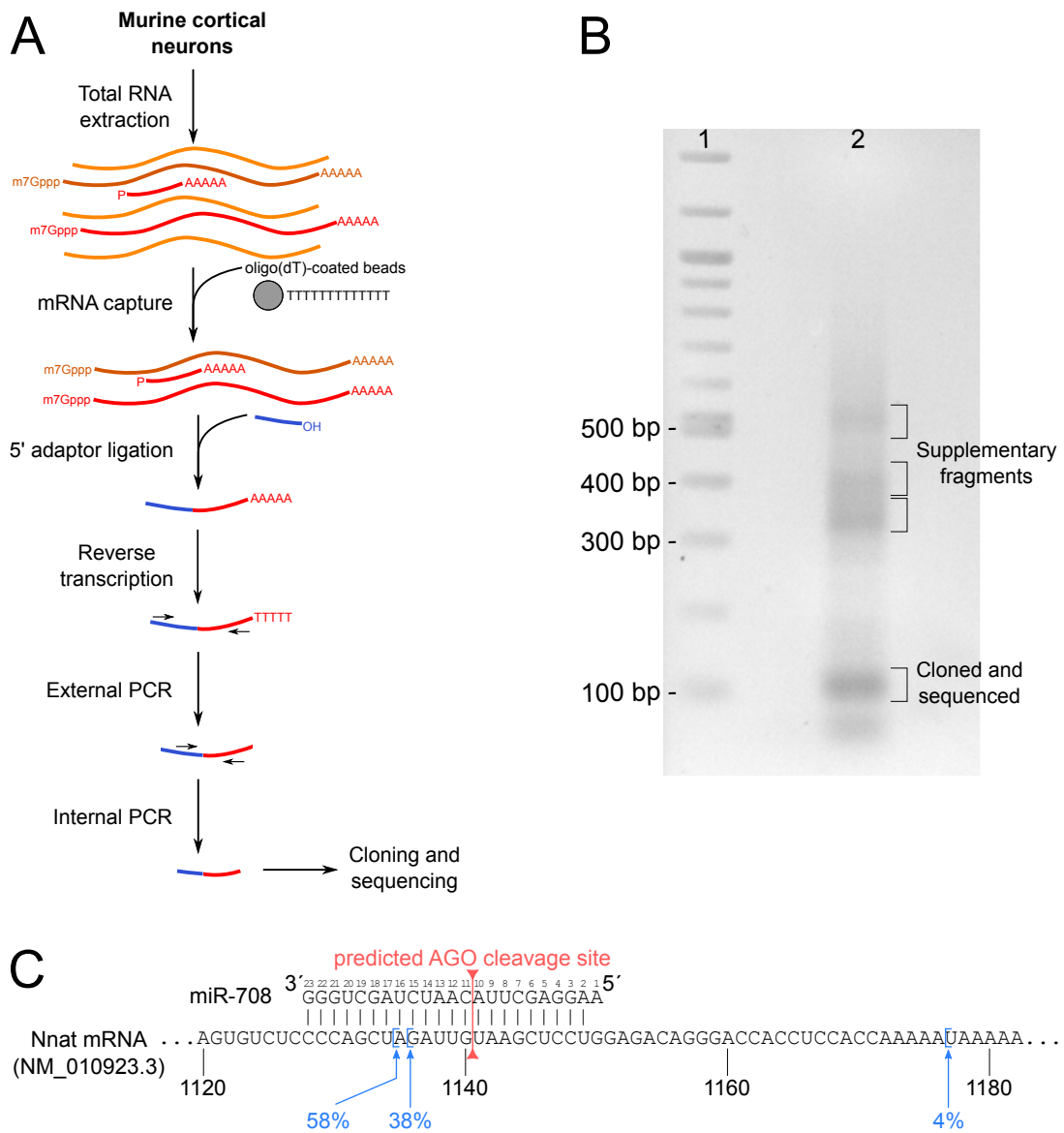
The rigorous examination of the candidate Nnat:miR-708 interaction shows a perfect complementarity, except for the 1st nucleotide at the 5' end of the miRNA. Such complementarity matches the 3' mismatches TDMD pattern and resembles the nearly perfect complementarity required for target miRNA-mediated cleavage. As explained in section 1.3.2, in cases of extensive or perfect complementarity, AGO proteins with active cleavage activity – only AGO2 in vertebrates – can directly cleave the target transcript, which is degraded afterwards. Therefore, it is conceivable that, in vertebrates, Nnat transcript and miR-708 reciprocally control each other in the following manner: miR-708 loaded into AGO1-3-4 may trigger the degradation of miR-708 by TDMD while when loaded into AGO2, the interaction may lead to the cleavage and degradation of Nnat transcript instead.

In order to verify if miR-708-bound AGO2 endogenously triggers Nnat cleavage, we performed an RLM-5'-RACE (RNA Ligase-Mediated Rapid Amplification of cDNA 5' Ends) from murine primary cortical neuron extracts. This modified 5'-RACE method allows to map the 5'-ends of target mRNAs within the expected cleavage site by taking advantage of two characteristics of AGO-mediated slicing: **(i)** it cleaves the target RNA strand between 10th and 11th nucleotides, relative to the 5'-end of the miRNAs, and **(ii)** it leaves fragments ending with 5'-monophosphates, thus competent for T4 RNA ligase-mediated ligation of a 3' hydroxyl-terminated adaptor.

The experimental outline is presented in Figure 4.7.A. External and internal primers were designed to produce a 72 nt-long fragment whether Nnat is indeed cleaved by AGO2. Negative control PCR reactions were performed without the adapter-specific forward or the gene-specific reverse primers to identify amplification products that result from nonspecific binding of the other pair primer. We finally obtained specific Nnat fragments of ~100 bp, 300 bp, 400 bp and 500 bp (Figure 4.7.B). The ~100 bp fragment was extracted, cloned and sequenced. Out of 24 clones, the detected 5'-ends of Nnat map predominantly between nt 15-16 and 16-17 relative to the 5'-end of miR-708, but not between nt 10-11 as expected (Figure 4.7.C). In conclusion, we were not able to detect AGO2-cleaved Nnat transcripts in murine primary cortical neurons.

This result may suggest that in murine primary neurons, miR-708 is not abundant enough to trigger detectable Nnat cleavage, or else that TDMD occurs faster than AGO2-mediated cleavage which may explain the rare observations of endogenous RNAi-like miRNA target slicing.





**FIGURE 4.7: AGO2-cleaved Nnat transcripts are not detected by RLM-5'-RACE in murine primary cortical neurons.** **A:** RLM-5'-RACE outline. Total RNA is extracted from murine primary cortical neurons, then poly(A)-RNAs are captured by oligo(dT)-coated beads and ligated with 5' adaptor. To ensure a higher yield of 5' RACE products and to reduce amplification of nonspecific products, Nnat fragments are amplified by nested PCR. Finally, the 5' RACE products are run on agarose gel and distinctive bands are gel-purified, cloned, and sequenced. **B:** Agarose gel electrophoresis of RLM-5'-RACE PCR products. Lane 1 = 100 bp NEB ladder and lane 2 = nested PCR product. Four gel fragments were extracted, cloned in pGEM-T easy and sequenced. **C:** Illustration of the expected cleavage site of Nnat by AGO2-loaded with miR-708 – between 10th and 11th nucleotides relative to the 5'-end of the miRNA – and detected Nnat 5'-ends from 24 clones of the 100 bp fragment.

## 4.4 Experimental Validation of TDMD Inducer Candidates

Our prediction program underlines three TDMD inducer candidates in the mouse cortex – *Nnat*, *Mapre2* and *Ppp1cb* – that remain to be experimentally validated. For this purpose, we planned to silence these transcripts and a positive control TDMD inducer, *Nrep*, in primary cortical neurons and to measure the subsequent levels of cognate miRNAs.

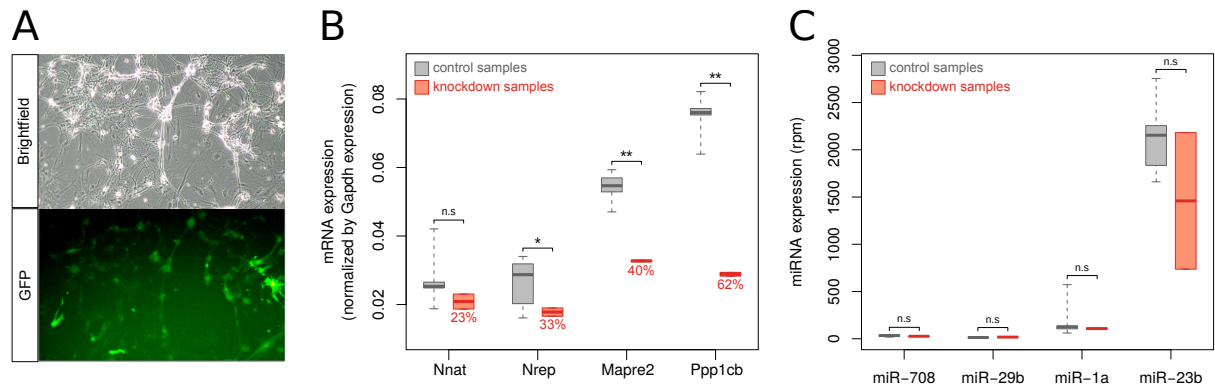
### Strategy 1: RNAi with pre-designed shRNA constructs

To perform a long-term RNAi, we used pre-designed shRNAs from Dharmacon cloned into the pGIPZ lentiviral vector (*cf.* appendix A). The pre-designed shRNA constructs are expressed as human pri-miR-30 transcripts with optimized flanking regions for efficient Microprocessor and Dicer processing (Chang et al., 2013; Silva et al., 2005). The company Dharmacon provided us three to four mouse-specific shRNA lentiviral vectors for each target gene to guarantee above 70% of knockdown.

Packaging of shRNA construct pools – to assure the highest possibility of mRNA targeting, multiple shRNA constructs are used at once to silence each gene – and purification of lentiviral particles have been performed by the Vectorology Platform of Montpellier. Infectious titration determination has been performed by FACS, 72 hours after HeLa cells transduction.

Isolation of primary cortical neurons from newborn mice P0-P3, as well as lentiviral transduction have been performed by our collaborators, the Perroy lab. At day 6 of *in vitro* culture (DIV6), dissociated neurons plated in 10 cm petri dishes were transduced with 25  $\mu$ L of lentiviral particles at  $10^9$  IU/mL. At 10 days post-infection (DIV16), we verified the efficiency of cell transduction and shRNA transcription by GFP imaging (Figure 4.8.A), and extracted total RNA for validation of target mRNA knockdown by RT-ddPCR (Figure 4.8.B) and sequencing of small RNAs (Figure 4.8.C).

GFP imaging validates the transduction of cortical neurons with lentiviral particles. However, as neophytes in neuronal cell culture, we also found glial cells in non-negligible amounts in primary cortical cell cultures. This could participate in the dilution of the derepression of miRNAs generated by the knockdown of TDMD inducers. But more likely, in this experiment the pool of shRNA constructs induced poorly efficient silencing of target genes. Especially, our TDMD positive control, *Nrep*, was likely not sufficiently knockdown (33%) to observe derepression of miR-29b. For the three candidates, we did not observe any miRNA derepression either. For *Ppp1cb*, the most silenced candidate – 62% of knockdown –, this result suggests that either the knockdown is too weak to affect miR-23b abundance, or *Ppp1cb* is not a TDMD inducer. For the two other candidates, we consider that their knockdowns – *Nnat* 23% and *Mapre2* 40%– were too low to induce any consequence.



**FIGURE 4.8: Lentiviral delivery of pGIPZ shRNAs against TDMD inducer candidates resulted in poorly efficient mRNA knock-down.** **A:** GFP imaging of murine primary cortical neurons 10 days after transduction of pGIPZ shRNA constructs against Nrep. **B:** RT-ddPCR measurement of TDMD inducer candidate expression 10 days after transduction of pGIPZ shRNA constructs. In grey are the abundance of the target RNA normalized by Gapdh abundance in control samples, and in red its abundance in knockdown samples. **C:** Small RNA-seq measurement of cognate miRNA levels 10 days after transduction of pGIPZ shRNA constructs. In grey are the abundance of miRNAs in control samples and in red their abundance in knockdown samples. **Panel B and C:** Means and standard errors of 3 biological replicates are represented by dots and error bars, respectively. The significance of the difference in expression between control and knockdown samples was tested by t-tests (“n.s.”: non-significant, “\*\*”: p-value  $\leq 0.05$  and “\*\*\*”: p-value  $\leq 0.01$ ).

We sought explanations for this poor silencing and explored small RNA sequencing data to evaluate siRNA processing. Commercial pGIPZ vectors were sequenced to obtain the pre-designed shRNA sequences based on the miR-30 pri-miRNA backbone as detailed in Chang et al., 2013. From small RNA-seq data, we examined the read coverage profile of shRNA hairpin sequences. Well-processed miR-30-based shRNAs should exhibit higher read coverage on their 3' arm – the siRNA guide strand – with well-defined extremities, resulting from accurate cleavages by DROSHA and DICER. However, if most of the pre-designed shRNA coverage profiles appear to be satisfactory, some shRNAs do not seem properly processed, with reads exhibiting inconsistent extremities and mapping on shRNA arms heterogeneously (Figure 4.9.A).

We also quantified reads mapping on the siRNA guide strand of each shRNA and observed that, for each target gene, only one siRNA looks well-processed ( $>100$  ppm): for Nnat, shRNA #1; for Nrep, shRNA #2; for Mapre2, shRNA #3; and surprisingly, none for Ppp1cb while it is the most effectively depleted RNA in our RNAi assay (Figure 4.9.B).

We then examined the structural accessibility of target mRNA sequences for the most expressed shRNAs. Using various RNA folding strategies (full-length mRNA folding, 7–15 nt windows by 2 nt increments, 17–33 nt windows by 2 nt increments, and 63–101 nt windows by 2 nt increments), we verified the self-pairing probability of pre-designed siRNA targets. Thus, the most accessible regions are those where the siRNA match has a low pairing probability for each match nucleotide. We found that pre-designed siRNAs target moderately to poorly accessible sequences (Figure 4.10). Altogether, these results suggest that this RNAi assay on potential TDMD inducers failed, partly because of a defect in the processing and targeting of

designed guide siRNAs.

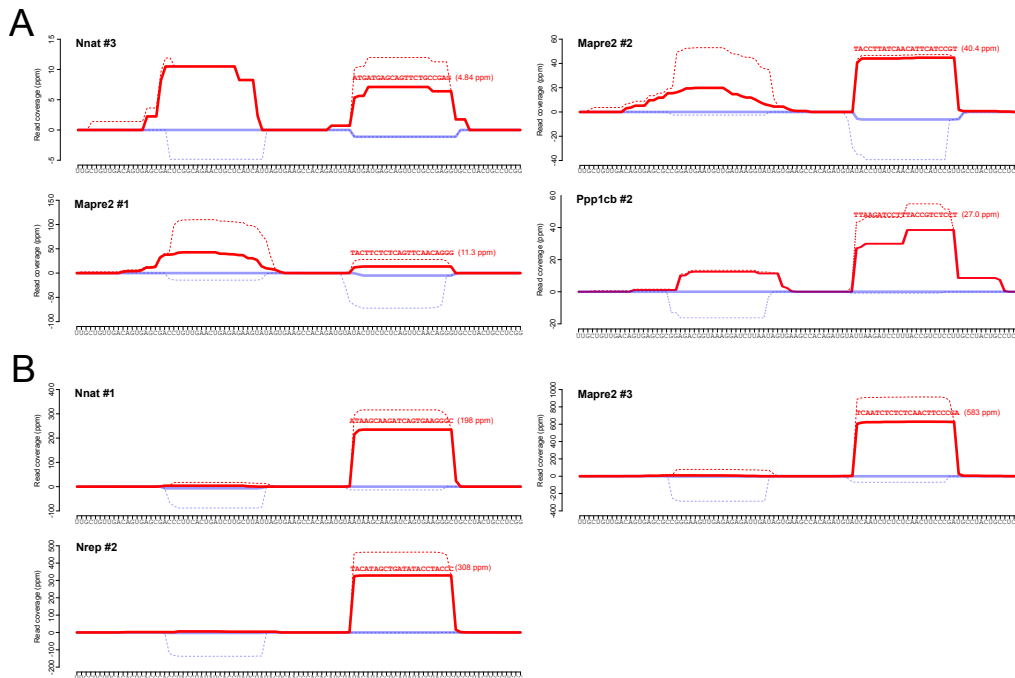


FIGURE 4.9: **Small RNA-Seq coverage profile of pGIPZ shRNAs.** **A:** Poorly processed and **B:** well processed pGIPZ shRNAs. Abundance of reads mapping on the sense and antisense strand of the shRNA are colored in red and blue, respectively. In solid lines are represented reads mapping unambiguously to one place of the loop-stem, and in dotted line, all reads.

### Strategy 2: RNAi with optimized shRNA constructs

Consequently, we designed our own shRNA sequences to clone into the lentiviral pGIPZ vector. The rationale behind our siRNA sequence choice was the accessibility of the gene target site and its compatibility between human and mouse genomes, so we could use them on cells from both species. In detail, we selected 13 nt-long stretches perfectly conserved between between humans and mice, then selected those with high structural accessibility. We used the same RNA folding strategies as detailed above. Candidates were ranked according to their accessibility by progressively increasing the pairing probability cutoff, and candidates captured by the same cutoff in both species were selected.<sup>2</sup>

Finally, to ensure better processing of shRNAs and to favor the guide strand selection, siRNA sense and antisense sequences were not inserted in the same pri-miR-30 backbone but instead in an artificial pri-miRNA backbone yielding large amounts of accurately processed miRNAs in Fang et al., 2015 (Figure 4.11). The antisense strand, which is the guide sequence, is placed on the 3' arm and wobble GU pairs are included at its 5' end to keep stable base pairs near the DROSHA cleavage sites which should favor accurate processing by DROSHA, therefore also by DICER. These shRNA sequences are cloned into the lentiviral pGIPZ vector.

<sup>2</sup>This strategy of siRNA selection was used, as part of a collaboration with the company “Medesis Pharma”, to design siRNAs targeting mRNAs of the SARS-CoV-2. The related invention patent is attached in appendix B.



#### 4.4. Experimental Validation of TDMD Inducer Candidates

quantified by RT-ddPCR. Again, none of the newly designed shRNA constructs induced efficient knockdown of the four target genes (Figure 4.12).

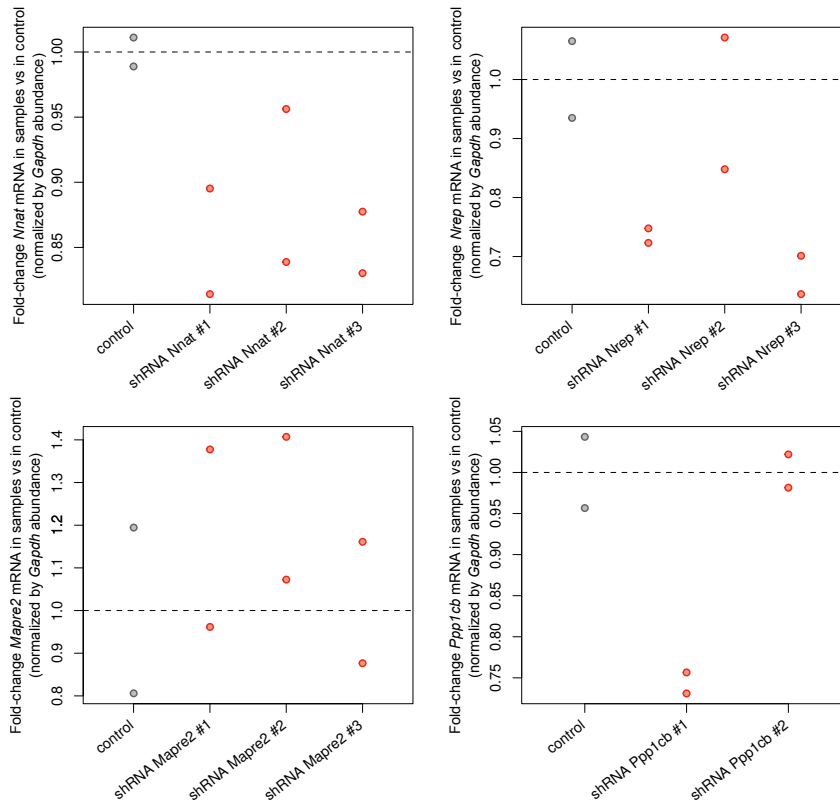


FIGURE 4.12: **Quantification of TDMD inducer candidates knockdown with optimized shRNA constructs.** RT-ddPCR measurement of TDMD inducer candidate expression 2 days after transfection of pGIPZ shRNA constructs. In grey are fold-change of abundance of the target RNA relative to the control and normalized by Gapdh abundance in control samples and in red in knockdown samples (two biological replicates are displayed). Despite the low number of replicates, two-way ANOVA analysis was performed to test whether siRNA transfection has an effect on target mRNA expression (Nnat p-value=0.141; Nrep p-value=0.0585; Mapre2 p-value=0.749; Ppp1cb p-value=0.0121), and post-hoc pairwise t-tests was performed for Ppp1cb to test whether sh11 or sh12 affect significantly Nrep mRNA levels (p-value=0.08561 and p-value=0.9755 respectively).

#### Strategy 3: dCas-KRAB-mediated gene silencing

Simultaneously, we attempted an RNAi-independent strategy based on the use of catalytically inactive Cas9 protein lacking endonuclease activity, also known as dead Cas9 or dCas9. Dead Cas9 was found to maintain guide RNA-directed binding of specific DNA sequences and, in this way, to sterically block transcription of coding regions in *Escherichia coli* (Qi et al., 2013). This system was called CRISPR interference (CRISPRi) and was recapitulated in mammalian cells but at a lower extent (Qi et al., 2013). However, the fusion of transcriptional repressor domains to dCas9, for example, of the Krüppel-associated box (KRAB) domains which promote heterochromatin formation (Gröner et al., 2010), does allow efficient and specific transcription repression in human cells (Gilbert et al., 2013). Since, this CRISPR-mediated gene silencing system is widely used for genetic screens in mammalian cells (Kampmann, 2018) and efficiently

replaces RNAi strategies.

Recently, the ZIM3 KRAB domain was shown to be more effective in target gene silencing and less sensitive to guide RNA selection than currently existing systems (Alerasool et al., 2020). Therefore, we decided to use the same vector expressing dCas9-KRAB/ZIM3 (Addgene 154472, cf. appendix A), in addition with a lentiviral dual CRISPR gRNA expression vector from mouse U6 and human U6 promoters (Addgene 72667, cf. appendix A). Guide RNAs against *Nnat* were designed to target sequences at 100 nt distance upstream or downstream of *Nnat* transcriptional start site (TSS) and following CRISPOR recommendations (<http://crispor.tefor.net/>) (Concordet et al., 2018). To allow a 5 days-long puromycin selection of transfected cells, co-transfection was performed twice: on day 1 and day 4, while puromycin treatment started on day 3. At day 7, total RNA was extracted and mRNA *Nnat* abundance was quantified by RT-ddPCR. Unfortunately, none of the tested gRNA resulted in *Nnat* knockdown (Figure 4.13).

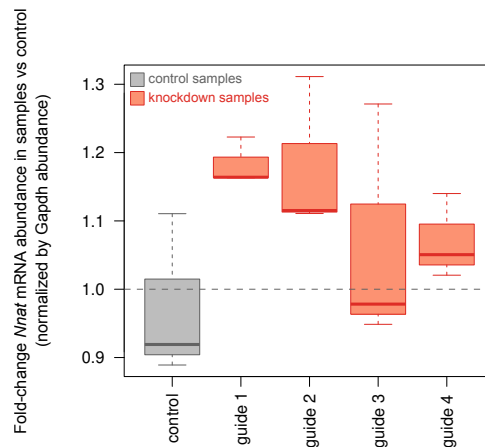


FIGURE 4.13: **Quantification of TDMD inducer candidates knockdown with dCas9-KRAB/ZIM3.** RT-ddPCR measurement of *Nnat* expression 7 days after co-transfection of dCas9-KRAB/ZIM3 and gRNA expression vectors and puromycin selection. In grey are fold-change of abundance of the target RNA relative to the control and normalized by *Gapdh* abundance in control samples and in red in knockdown samples (three biological replicates, mean  $\pm$  st.error). Two-way ANOVA analysis shows that none of the dCas9/KRAB treatment affect significantly the *Nnat* mRNA expression ( $p$ -value=0.202).

## 4.5 Analysis of Degradation Intermediates from small RNA-seq Data

Following the early characterization of the TDMD mechanism and the discovery of endogenous TDMD inducers in metazoan neuronal cells, miRNA trimming and A or U-tailing were found to be correlated with the TDMD process. Thus, it was initially assumed that the 3′A or U-tailed and 3′ trimmed miRNA isoforms were intermediate degradation species, and consequently their distribution was used to be explored to demonstrate a TDMD event. Later, the publication of Han et al., 2020 and Shi et al., 2020 papers have demonstrated that TDMD and miRNA 3′ end enzymatic modification, or TDTT, are more likely independent processes that may influence each other but are mechanistically uncoupled.

Analysis of small RNA-seq data typically involves counting reads aligned to miRNA loci whose ends match the native form of the miRNA from the miRBase database. Thus miRNA variants are either included in this count without distinction, or completely excluded depending on the stringency setting. We, therefore, developed a program to distinguish native miRNAs from canonical and non-canonical 3′ isomiRs in small RNA sequencing datasets, and count them apart. We differentiate canonical variants which are generated by an alternative maturation cleavage, from non-canonical derived from 3′ end addition or trimming of nucleotides. The initial scope of this tool was to provide insights into miRNAs ongoing TDMD, but it also could help to identify miRNAs subjected to 3′-end enzymatic modifications in a given biological context and provide an exact quantification of miRNAs and their isomiRs. Such computational pipeline has been developed like isomiRage from the Nicassio lab, a web application for the analysis of mouse and human dataset (Muller et al., 2014). In our case, this workflow has been developed within the scope of a collaboration with Dr. Sebastian Canzler and Dr. Jörg Hackermüller (Helmholtz Centre for Environmental Research - UFZ, Germany), in part of the project “XomeTox”, an oral rat toxicity multi-omics-based study focusing on direct and indirect thyroid toxicity.

### **XomeTox samples preparation**

Preparation of samples and total RNA extraction were covered by our collaborators. In detail, samples were grouped in preparation sets to contain 15 liver and 15 thyroid samples from each of the 15 treatment groups. The subsequent processing of those samples in each set was randomized meaning that the samples were not prepared in their originally assigned order. Additionally in each preparation set a negative control (water) and a positive control (human liver total RNA) was included. Total RNA quality was assessed on electrophoretic spectra from a Fragment Analyzer (Agilent), analyzed with the PROSize software (v. 3.0.1.6). Libraries were prepared using NEXTflex Small RNA-Seq Kit v3 (Bio Scientific) following the manufacturer’s instructions. Libraries were verified by DNA quantification using Fragment Analyzer (kit High Sensitivity NGS), and by qPCR (ROCHE Light Cycler 480). Libraries were sequenced on Illumina NovaSeq 6000 using NovaSeq Reagent Kit (100 cycles). RNA quality assessment, library preparation, validation and sequencing were performed by the MGX sequencing facility. Adapters ended with 4 randomized nucleotides in order to reduce ligation biases. Because



of the sequencing design, the adapter sequence (5' GTTCAGAGTTCTACAGTCCGACGATC-NNNN 3') appears at the beginning of the read sequence, and the final 4 nucleotides of the read are the initial randomized nucleotides of the other adapter, whose other nucleotides are not read.

### Development of a miRNA isoform quantification approach

The adapter-trimmed reads were mapped with HISAT2 (Kim et al., 2015) on rat pre-miRNA hairpin sequences from miRBase (version October 2018) extended by 10 nt on each flank according to the rat genome sequence. In order to account for potential non-canonical trimming and tailing at the miRNAs 3' end, a multi-pass mapping procedure was employed. First, full-length reads are mapped with default settings, and unmapped reads are recorded in a dedicated file. By turns, unmapped reads are mapped with the "-3" option, which trim a set number of bases from the 3' end of each read before alignment, and unmapped reads are still recorded in a dedicated file. This operation is repeated from one trimmed until five trimmed nucleotides. At the end of the procedure, miRNA read counts are expressed as "rpm" (reads per million) after normalization per sequencing depth, *i.e.*; the number of adapter-trimmed reads matching the rat genome according to HISAT2 on full-length reads.

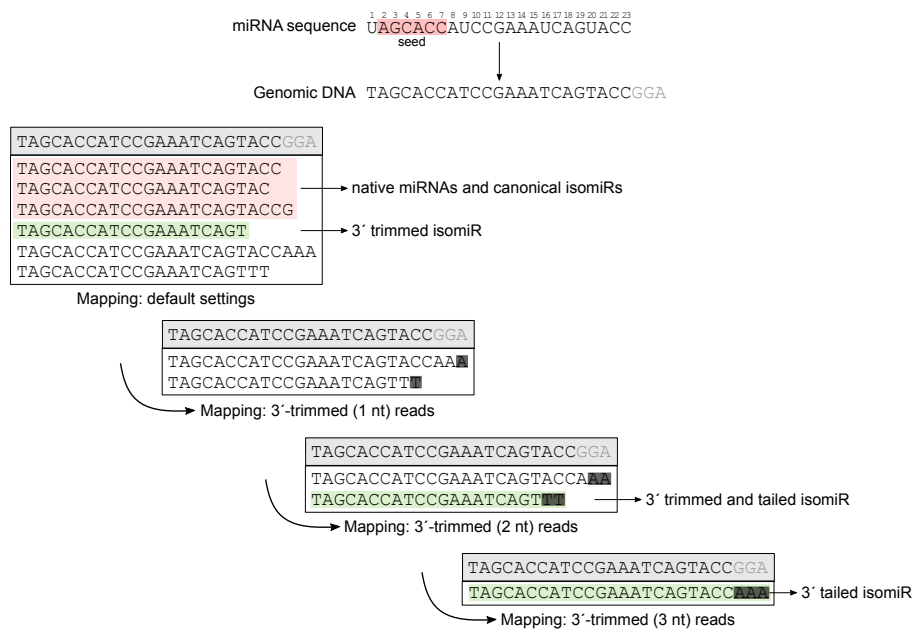


FIGURE 4.14: Example of the mapping strategy to record canonical and non-canonical isomiRs. The initial mapping is performed with default setting. Full-size reads that align to the miRNA locus (highlighted in grey) are recorded as miRNAs (read length = miRNA length) and canonical isomiRs (read length = miRNA length -1 or +1 to +3 of genomically templated nt) or as 3' trimmed isomiRs (read length = miRNA length -2 to -3). Non-mapped reads go through successive 3'-trimmings and mappings. 3'-trimmed reads that align to the miRNA locus are recorded as 3' tailed isomiRs (3'-trimmed read length = miRNA length -1 or +1 to +3 of genomically templated nt) or 3' trimmed and tailed isomiRs (3'-trimmed read length = miRNA length -2 to -3). Canonical isomiRs are highlighted in red and non-canonical isomiRs in green.

Native miRNA sequences and canonical isomiRs comprise reads with a same 5' end than in the miRBase-annotated sequence, and a 3' end trimmed by no more than 1 nt, and extended

by no more than 3 genomically templated nucleotides. By contrast, non-canonical isomiRs correspond to reads with a same 5' end but a 3' end trimmed by 2 to 3 nt, extended by 3 to 5 non-templated nt, or else trimmed by 1 to 3 nt then tailed by 3 to 5 non-templated nt.

From the XomeTox dataset, we found that native miRNAs and canonical isomiRs account for most sequences in the libraries: around 70-75% of genome-matching reads, minimum 67.4% and maximum 82% (Figure 4.15).

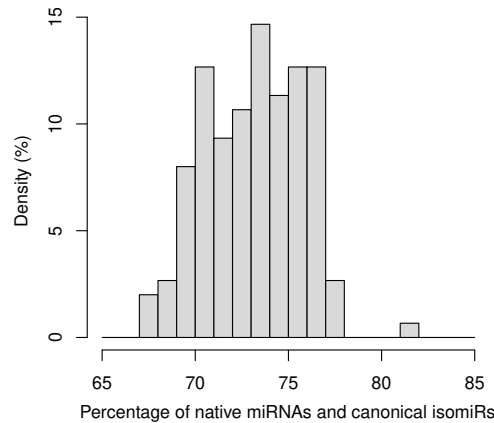


FIGURE 4.15: **Distribution of native miRNAs sequence percentage in Xometox libraries.** X-axis: the percentage of canonical miRNAs and isomiRs in a library. Y-axis: percentage of libraries.

#### Inconsistencies in rat miRNA annotation

During the development of this workflow, we noticed several non-negligible misannotations in the rat miRNA dataset from miRBase. Among them:

- Some pre-miRNAs do not match the rat genome.
- Some pre-miRNAs recorded in miRBase and the miRNA hairpin Fasta file are missing from the miRBase GFF3 file.
- Some miRNAs generated from multiple loci are assigned to a single locus in miRBase (e.g., rno-mir-26a is attributed to a single locus on chr8, while it also has another hairpin on chr7)
- For 47 miRNA hairpins, their antisense has also been annotated as a bona fide hairpin (e.g., rno-mir-3556b-1 is the antisense of rno-mir-29c-1).

Consequently, the classical strategy of mapping reads on the genome and then intersecting with the miRBase GFF3 annotation file would have provide misestimations of miRNA counts. To resolve this issue, we prepared a manually curated Fasta file with hairpin sequences extended by 10 nt on both ends – in order to account for reads with imprecise maturation – except for the 47 antisense-fake hairpins. From that cleaned-up Fasta file, we generated a new HISAT2 index and mapped reads on that index.

## 4.6 Discussion and Perspectives

The second part of this thesis involves the reciprocal regulation of miRNAs by their targets and the recent discovery of an active and highly specific mechanism of miRNA decay induced by miRNA targets, the Target-Directed miRNA Degradation pathway. Mostly, our work consisted in providing a computational tool for the systemic research of TDMD inducers, which will complement the current collection of prediction strategies for miRNAs or miRNA targets.

During this project, we designed an adaptable workflow for the identification of conserved TDMD-compatible pairing sites in any tissue or cell line of interest, and attempted to validate the best *in silico* candidates experimentally. For this purpose, we had to modulate the TDMD inducer expression and measure the cognate miRNA's subsequent derepression. TDMD examples have been reported or suggested by ZSWIM8 depletion in various cell types: human cancer (HeLa, MCF7, A549) and inducible neuronal cells (SH-SY5Y); rodent embryonic cells (MEFs, mESC, embryonic hypothalamic neurons) and cell lines (NIH3T3, K562); drosophila cells (S2); and in *C.elegans* (de la Mata et al., 2015; Ghini et al., 2018; Han et al., 2020; Shi et al., 2020). Despite the possibility to study more effortless cells, we chose primary neuronal cells for our experimental validation as the miRNA derepression extent following the TDMD pathway disruption would be greater in this system as illustrated in de la Mata et al., 2015 with a 7 to 12-fold derepression after mutation of the TDMD inducer against 1.5 to 2-fold in non-primary neuronal cells.

The implementation of our TDMD inducer prediction program has resulted in the selection of three candidates to be experimentally validated: Nnat for the regulation of miR-708-5p, Mapre2 for miR-1a-3p and Ppp1cb for miR-23b-3p. Unfortunately, the TDMD candidate silencing strategies we opted for were either insufficient or required more optimization. At this time, we chose to turn to CRISPR-Cas9 technologies in place of RNAi through shRNA expression. We have already initiated the design of dCas9-KRAB-based silencing strategy as well as the genetic invalidation of the TDMD inducer candidates by Cas9-mediated knockout. Furthermore, to resolve the potential issue that primary neurons have to be co-cultured with glial cells to survive – which can dilute the studied effect – we plan to purify neurons from glia by growing them in a device allowing simple mechanical separation of the two cell types (Viviani, 2006).

In the course of selecting TDMD inducer candidates, the target:miRNA pair Nnat:miR-708 aroused our curiosity because of its complementarity geometry: these two RNAs are perfectly complementary, except on miRNA nt 1. This observation raises the obviousness that the TDMD-inducing 3' extensive complementarity is compatible with the perfect – or nearly perfect (Schwarz et al., 2006) – complementarity required for AGO-induced target slicing. It can therefore be anticipated that some miRNA:target pairs will simultaneously trigger the degradation of each RNA. In *Drosophila* for example, this phenomenon is not questioned since AGO1, the miRNA-loading AGO, does not exhibit catalytically efficient endonuclease activity (Förstemann et al., 2007). However, in vertebrates, AGO2 has conserved its mRNA slicing

ability (Liu et al., 2004; Meister et al., 2004). Thus, it would be possible that the miRNA:target pairing function would depend on the nature of the AGO protein loading the miRNA.

We attempted to identify products of Nnat cleavage by miR-708-guided AGO2 in murine primary cortical neurons using RLM-5'-RACE but did not obtain any fragments mapping the expected AGO2 cleavage site. In light of the low levels of miR-708 in comparison with the high expression of Nnat transcript in this cellular context (Figure 4.8), miR-708 may not be abundant enough to induce any detectable AGO2-induced slicing. Indeed, it was precisely demonstrated that low abundant miRNAs induced virtually no mRNA degradation, whereas highly abundant miRNAs trigger efficient degradation of the target mRNAs, suggesting that variations in miRNA abundance can balance between TDMD and mRNA decay processes (de la Mata et al., 2015; Mullokandov et al., 2012). To confirm this hypothesis, we could overexpress miR-708 in primary cortical neurons by using the pGIPZ lentiviral vector cloned with the pri-miR-708 instead of the shRNA-adapted pri-miR-30, and then reiterate a RLM-5'-RACE. It would be also possible to study this mutual interaction in another cellular system easier to mutate by CRISPR/Cas9, such as HCT-116 cells (especially suitable for genome editing) or the human neuroblastoma SK-N-BE(2) cell line (more similar to the primary neuron model), once it would have been verified that one of this line is TDMD-competent with a positive TDMD control: Nrep:miR-29b or Cyrano:miR-7 interaction. Exactly, we could measure both components of the reciprocal regulation: comparing miR-708 levels in wild-type and Zswim8-deficient cells would quantify the amplitude of Nnat-guided TDMD, while comparing Nnat mRNA levels in wild-type and Ago2-deficient cells would evaluate the extent of miR-708-guided Nnat cleavage.

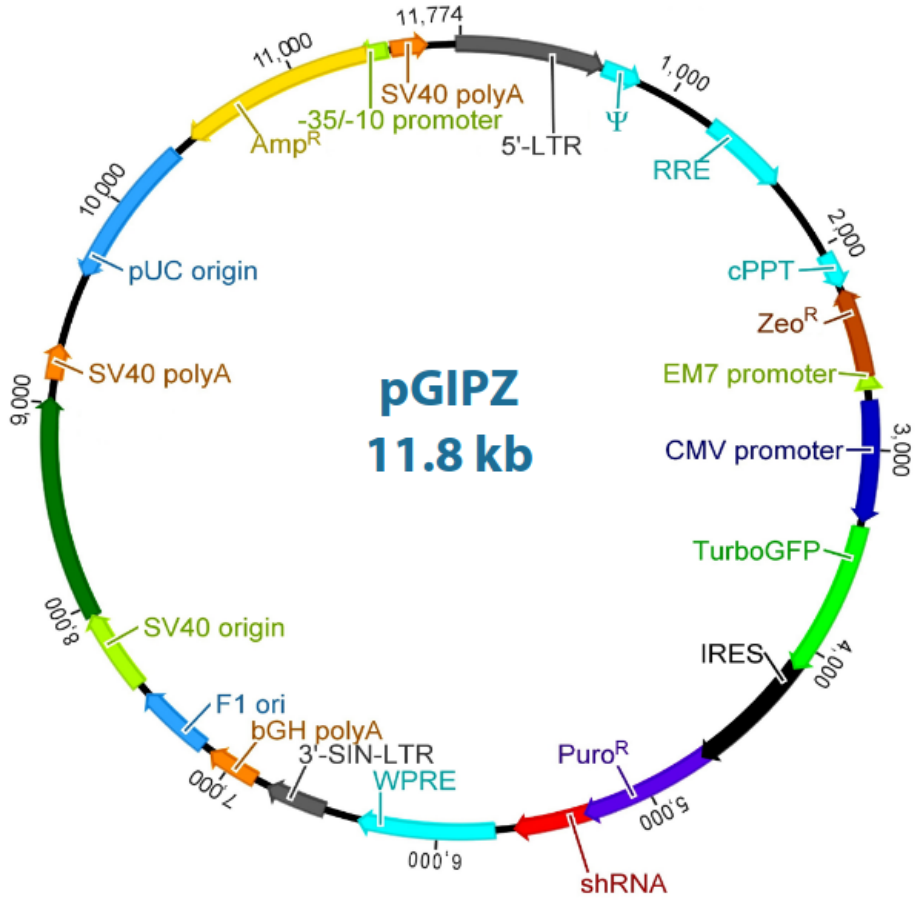
Furthermore, because mammalian miRNAs appear to associate to every available Ago protein without any clear preference (Meister et al., 2004), it is expected that miR-708 loads equally on AGO1, AGO2, AGO3 and AGO4, in proportion to their relative amounts. But because only AGO2 is able to efficiently cleave the Nnat mRNA, we anticipate that TDMD of miR-708 will be less active on the AGO2-loaded pool of miRNA than on the AGO1-, AGO3- and AGO4-loaded pools. It is possible to verify that prediction by capturing miR-708-5p/AGO complexes by a biotinylated oligo (Hutvágner et al., 2004), then measuring each AGO's abundance by quantitative mass-spectrometry to detect potential enrichments for AGO2 in the miRISC-miR-708 complexes.

Finally, the differential potency of TDMD between neuronal and non-neuronal cells was not explained for now. Because Nnat exhibits a neuron-specific expression, we could conceive that the presumed reciprocal regulation between Nnat and miR-708 may explain Nnat distribution, either because neuronal cells display a favorable context for TDMD (higher AGO1-3-4/AGO2 ratio, higher levels of ZSIW8 or other Cullin-RING E3 ubiquitin ligase complex member...), or instead for Nnat-specific TDMD (lower levels of miR-708). More precisely, the TDMD-compatible site located in Nnat 3'-UTR could mediate canonical miRNA-mediated silencing or cleavage of the mRNA in non-neuronal cells, or turn to an inducer of miRNA decay through TDMD in neuronal cells. Such assumption would require to investigate both the miR-708 distribution in various cells and the correlation with Nnat expression, and the presumed tissue-specificity of TDMD.

Using the UCSC whole genome multiple alignments from 124 insects, we could also apply the same strategy in order to identify TDMD inducer candidates in *D. melanogaster*. Because many *Drosophila* mutant lines are already available, mutants will probably already exist for many candidates and would greatly facilitate the experimental validation of our computational predictions. Whenever a candidate mutant would be available and viable, we would use RT-ddPCR to evaluate miRNA abundance in these mutants and compare with cognate heterozygotes. Candidates with an apparent TDMD activity and candidates that could not be tested because the mutant was lethal and whose phylogenetic conservation would be particularly deep will then be properly validated. Using genome editing to mutate only their TDMD-inducing MRE, we could also characterize their *in vivo* phenotypes, including development, motricity, and learning abilities. However, available RNA-Seq from NCBI for the wild-type *Drosophila* brain are limited. Therefore, we must first generate RNA-Seq data from whole brain at various stages using tissue-specific GFP markers to facilitate dissection in the earliest stages.

# Appendix A

## Vector Maps



Drawing was created using Geneious (<http://www.geneious.com>).

FIGURE A.1: **Vector map of pGIPZ lentiviral vector from Dharmacon.** The Dharmacon GIPZ Lentiviral shRNA Library was developed in collaboration with Dr. Greg Hannon of Cold Spring Harbor Laboratory [CSHL] and Dr. Steve Elledge of Harvard Medical School. From the online technical manual provided by Dharmacon.

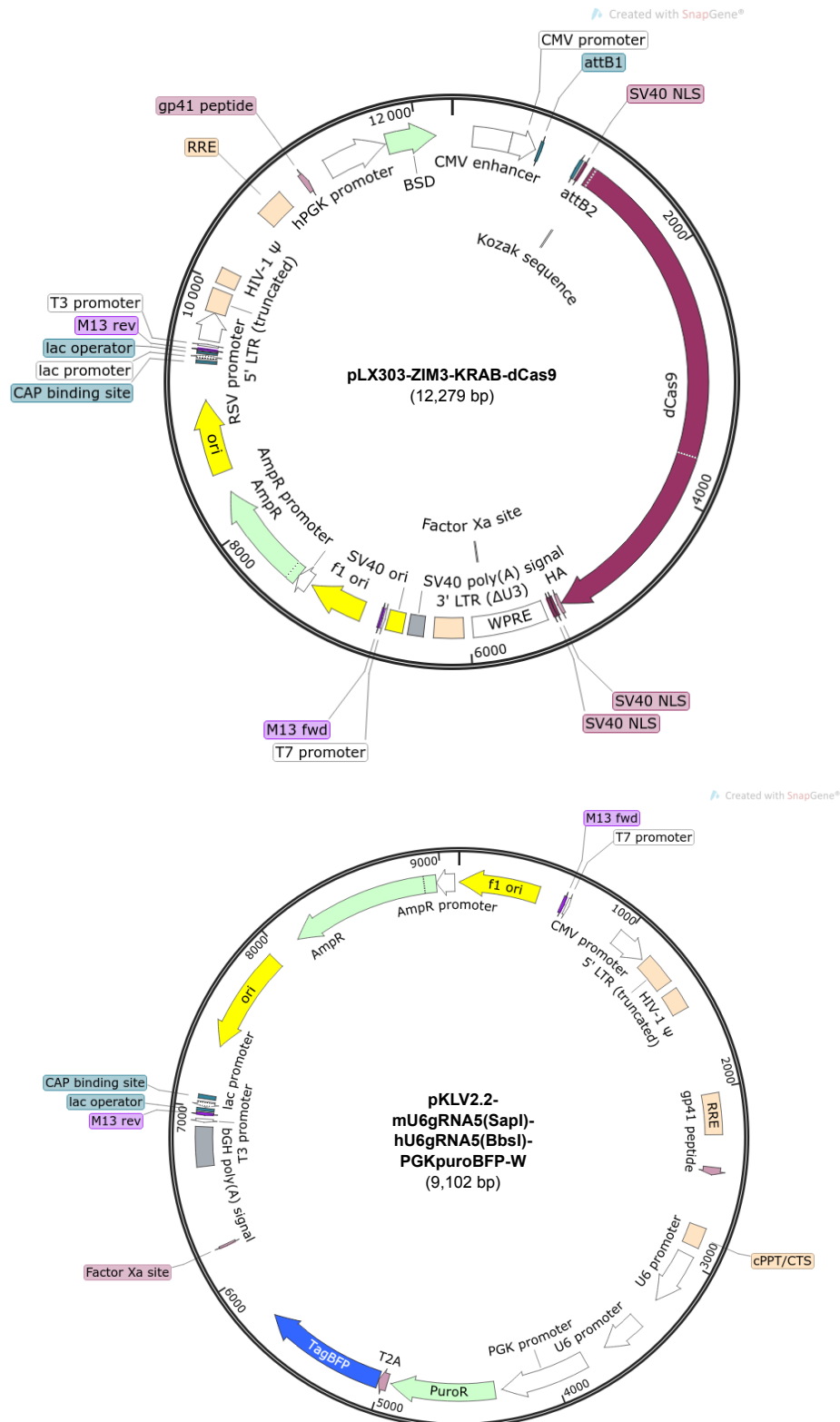


FIGURE A.2: Vector maps of plasmids for dCas9-KRAB/ZIM3-mediated gene silencing. Left: The pKLV2.2 plasmid is a lentiviral dual CRISPR gRNA expression vector from mouse U6 and human U6 promoters (Addgene 72667). Right: pLX303 is a lentiviral vector expressing dCas9-KRAB/ZIM3 (Addgene 154472)

## Appendix B

# Patent WO2021185766

The strategy of siRNA selection detailed in section 4.4 have been initially developed as part of a collaboration with the company Medesis Pharma to design siRNAs targeting mRNAs of the SARS-CoV-2.

The first draft of the siRNA selection workflow have been developed in February 2020 and my participation was the double-check of this siRNA selection from SARS-Cov-2 genomic data available in March 2020 (at that time, the number of SARS-Cov-2 genome sequences posted online was increasing day by day). An invention patent was filed in March 2020 and is currently being made public: "Treatment of Covid-19 with reverse micelle system comprising unmodified oligonucleotides" (<https://data.inpi.fr/brevets/WO2021185766>).

Below is the content of this patent, except for pages 22 to 31 which consist of the Sars-CoV-2 genome sequence.



administration routes: ocular, skin, oral, intramuscular. Those attempts have not been totally satisfactory so far. For instance, some of these attempts, more specifically assays with nucleic acids in liposome carriers have stimulated immune response.

5 The object of the present invention is to overcome disadvantages of the prior art. There is an obvious need for a safe and efficient nucleic acids therapeutic strategy for the treatment of diseases related to SARS-CoV-2 virus (or for the treatment of COVID-19), and in particular for new tools that are able to achieve efficient gene expression modulation-based therapy in order to treat diseases related to SARS-CoV-2 virus. More particularly, it is an object of the invention to provide a drug delivery system comprising an unmodified oligonucleotide targeting one or more genes of SARS-CoV-2, which can be for instance administered via buccal mucosa, giving rise to a satisfactory drug bioavailability in an active form.

#### SUMMARY OF THE INVENTION

10 The present invention relates to a delivery system for the *in vivo*, or *ex vivo* release of unmodified oligonucleotides targeting SARS-CoV-2 RNAs, by administration to the buccal or rectal mucosa of said delivery system, as well as the compositions and methods for preparing the delivery system.

More specifically, the delivery system is a reverse micelle system comprising at least one sterol, acylglycerol, phospholipid, an alcohol, and at least one unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 virus.

20 Herein described are reverse micelle systems designed to reach this goal in a safe and controlled manner. The reverse micelle systems are able to be absorbed through mucosa and to vectorize unmodified oligonucleotides under a protected form to all cells of any tissue of the organism.

The invention also relates to a pharmaceutical composition comprising a reverse micelle system as defined herein and a pharmacologically acceptable support or carrier.

25 The present invention relates to the use of unmodified oligonucleotides targeting CoV\_2019 RNAs, in particular siRNAs whose sequences have been designed to inhibit one or more genes expression of this virus, in the preparation of reverse micelle systems or pharmaceutical compositions comprising the same in the treatment of diseases related to SARS-CoV-2 (or in the treatment of COVID-19).

30 The aim of the present invention is to provide unmodified oligonucleotides targeting SARS-CoV-2 RNAs in a delivery system that allows to vectorize said oligonucleotides as to down regulate or knock

#### TREATMENT OF COVID-19 WITH REVERSE MICELLE SYSTEM COMPRISING UNMODIFIED OLIGONUCLEOTIDES

##### FIELD OF THE INVENTION

5 The present invention relates to specific reverse micelle system of the invention which allows the administration and intracellular delivery of unmodified oligonucleotide, such as siRNA, targeting one or more genes of the SARS-CoV-2 virus. The reverse micelle system of the invention is thus particularly useful for the treatment of the viral pathology linked to the SARS-CoV-2 virus.

##### BACKGROUND OF THE INVENTION

10 Coronavirus disease 2019 (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease has spread globally since 2019, resulting in the 2019–21 coronavirus pandemic. Common symptoms include fever, cough and shortness of breath. Muscle pain, sputum production and sore throat are less common symptoms. While the majority of cases result in mild symptoms, some progress to pneumonia and multi-organ failure. The deaths per number of diagnosed cases is estimated at between 1% and 5% but varies by age and other health conditions.

15 There is consequently a major public health emergency to treat CoVID\_19 which is spreading worldwide very quickly.

RNAi (RNA interference) and antisense (AS) strategies consist in silencing the expression of a target gene by the use of nucleic acids which allow the degradation or the translational arrest of mRNA target. New antisense applications (exon skipping, alternative splicing correction), by masking the mutation responsible for an alternative splicing default, have permitted the synthesis of a functional protein. Aptamers are nucleic acids capable of interacting with a target protein and down regulating its synthesis. The discovery of all these nucleic acids, and more recently siRNA and miRNA, has opened wide perspectives in therapeutics for the treatment of diseases like genetic diseases, cancers, neurodegenerative diseases, infectious and inflammatory diseases or to block cell proliferation and diseases caused thereby.

25 However, these molecules are unstable in biological fluids, *in vitro* and *in vivo*, they display a poor intracellular penetration and low bioavailability. These critical drawbacks have limited their use in therapeutics. As a result, clinical applications of said nucleic acids have required chemical modifications with the aim of retaining their capacity to knockdown protein expression while increasing stability and cellular penetration. Research groups have also applied the nanotechnology approach to improve their delivery, to overcome most barriers that hampered the development of nucleic acids delivery-based therapies. To improve bioavailability, many researchers have also attempted to use alternative

transparent formulation, and duration of the stirring is such that the stirring may be stopped a few minutes after obtaining the visually transparent formulation.

The present invention further relates to a pharmaceutical composition comprising reverse micelles of the invention and a pharmaceutically acceptable carrier, excipient or support.

## 5 DETAILED DESCRIPTION OF THE INVENTION

The following description is of preferred embodiments by way of examples only and without limitation to the combination of features necessary for implementing the invention.

### Reverse micelles

10 The reverse micelle system according to the invention is characterized as a micro-emulsion comprising a dispersion of water-nanodroplets in oil. The dispersion is stabilised by two surfactants (acylglycerol, more preferably a diacylglycerol of fatty acids and a phospholipid, more preferably phosphatidylcholine) and a co-surfactant (alcohol) that are most likely at the water/oil interface. The reverse micelle phase can be defined as a system wherein water forms the internal phase and the hydrophobic tails of the lipids form the continuous phase. Reverse micelles containing oil(s), surfactant(s), co-surfactant(s), and an aqueous phase are also characterized as water-in-oil micro-emulsions.

Generally, the size of micelles according to the invention is very small, more particularly, it is less than 10 nm; more specifically it is less than 8 nm and more preferably less than 5 nm. The size may vary with the quantity of added water and phospholipid. The present invention relates more particularly to 20 reverse micelles with an aqueous core of 3 to 5 nm, preferably from 3.5 to 5 nm, in particular from 3.7 to 4.5 nm.

The reverse micelles and the size of their aqueous core can be characterized by various methods, including:

- 25 - Small Angle X-Ray Scattering (SAXS)
- Neutrons Scattering
- Transmission Electron Microscopy (TEM)
- Dynamic Light Scattering (DLS)

The ratios of the lipidic constituents (including sterol, acylglycerol and phospholipid) in the reverse-micelle system according to the invention can vary. For instance, the weight ratio sterol/acylglycerol can range from 0.015 to 0.05, more particularly from 0.03 to 0.04. The weight ratio phospholipid/acylglycerol is from 0.06 to 0.25. For the calculation of these ratios, the weight of phospholipid corresponds to the total weight of the mixture of phospholipids, for instance the weight of lecithin, used in the formulation.

down the expression of a target nucleic acid of SARS-CoV-2 virus, with high efficiency and limited off-target-mediated secondary effects.

Drug delivery technology according to the invention allows intracellular delivery to all tissues and organs using HDL lipoprotein (High Density Lipoprotein) or vHDL lipoprotein (Very-High-Density Lipoprotein) receptors.

5 More particularly, the present invention provides a reverse-micelle transport system for delivering unmodified oligonucleotides capable of modulation of gene expression or duplication of genes of SARS-CoV-2 virus. More specifically, reverse micelles according to the invention allow the incorporation thereof in HDL and vHDL lipoprotein in the buccal or rectal mucosa. Reverse micelles according to the invention are thus carried in a protected lymphatic transport form, then in the general blood circulation which finally allows an intracellular delivery of said oligonucleotides by the membrane receptors of HDL lipoproteins, of the SRB-1 type (Scavenger receptor class B type 1).

10 Advantageously, at no time can the subject's immune system detect the presence of the oligonucleotide, with absence of immune reaction when unmodified oligonucleotides are administered in the technology according to the invention.

The reverse micelles can be prepared according to a method described below using at least a sterol, an acylglycerol, a phospholipid, an alcohol, water, and at least one unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 virus.

20 Said micelles are more particularly obtainable by the following method:

- (a) Contacting (i) sterol, preferably sitosterol or cholesterol, (ii) acylglycerol, preferably diacylglycerol of fatty acids, (iii) phospholipid, preferably phosphatidylcholine, (iv) alcohol, (v) water, preferably purified water, and (vi) at least an unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 virus,
- 25 (b) Stirring mixture obtained in step (a), at 40 °C or less, and for a time sufficient to obtain formation of reverse micelles, said stirring being preferably carried out mechanically or by sonication.

The parameters of the mechanical stirring, for instance duration and speed, can be readily determined by anyone skilled in the art and depend on experimental conditions. In practice, these parameters are such that a micro-emulsion is obtained; the speed is determined so as to enable formation of a visually 30

The oligonucleotides described in the present invention are necessarily unmodified in order to be water-soluble.

Accordingly, the invention ensures absorption of the compounds to be delivered across mucosa, preferably across mouth, nasal and/or rectal mucosa, more preferably across mouth mucosa. Also, reverse micelles of the present invention provide an important bioavailability with low variability of absorption.

#### Method for preparing reverse micelles

In a particular embodiment, the invention relates to a method for preparing reverse micelles as defined above (involving more specifically at least one unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 virus, a sterol, an acylglycerol, a phospholipid, an alcohol, and water), wherein said method comprises the following steps :

- (a) Contacting (i) sterol, (ii) acylglycerol, preferably diacylglycerol of fatty acids, (iii) phospholipid, preferably phosphatidylcholine, (iv) alcohol, (v) water, preferably purified water, and (vi) at least one unmodified oligonucleotide capable of targeting one or more genes of SARS-CoV-2 virus,

- (b) Stirring mixture obtained in step (a), at 40 °C or less, and for a time sufficient to obtain formation of reverse micelles, said stirring being carried out mechanically or by sonication.

The obtained and recovered reverse micelles are then particularly useful as a delivery system for unmodified oligonucleotides. Step (b) of the process is of particular importance since it allows reverse micelles to be obtained, said reverse micelles being then useful as a transport system to deliver unmodified oligonucleotides directly into the cytoplasm of all cells in all tissues and organs, through the cell membrane lipoprotein receptors.

In a particular embodiment, the unmodified oligonucleotide is first solubilised in water (preferably purified water) to form an aqueous phase. Said aqueous phase is then introduced into the oily phase (according to step(a)). The oily phase preferably comprises at least a sterol, an acylglycerol, a phospholipid and an alcohol.

The compounds involved in step (a) will be described in more details below.

Stirring of the mixture obtained by step (a) is carried out at a temperature less than or equal to 40°C, preferably ranging from 30 °C to 38 °C, more preferably from 30 °C to 35 °C, for a time sufficient to form of reverse micelles. The time sufficient can vary in particular upon the used stirring techniques,

The compounds of the reverse-micelle system can be analysed by appropriate means. More specifically, sterols can be identified by gas chromatographic analysis and acylglycerol by high-performance liquid chromatography (HPLC), in particular with a light scattering detector, on a silica column (kromasil C18), in the presence of an eluent, e.g. isocratic acetonitrile. Gas chromatography can also be used to analyse diacylglycerols. Phospholipids can be analysed by high-performance liquid chromatography (HPLC), with a diol column with a light scattering detector.

Reverse micelles are dynamic systems. Brownian motion causes perpetual collisions of micelles, which lead to coalescence of micelles and exchange of the aqueous cores. Separation and regeneration of micelles occur and allow chemical reactions between different solutions. The exchange rate between micelles increases in particular with temperature, the length of hydrocarbon chains of the surfactant, and the water/surfactant ratio. Within the context of the invention and contrary to what is expected in nanotechnology, aqueous cores of micelles must have a specific size allowing one or more molecules of unmodified oligonucleotide, in particular nucleic acid capable of mediating RNA interference, to be stabilised in the prepared micelles. As mentioned above, the size of the aqueous core is around 4 nm, preferably from 3 to 5 nm, more preferably from 3.5 to 5 nm, in particular from 3.7 to 4.5 nm.

Without being bound to any theory, it seems that inclusion of a phospholipid in the reverse micelle system allows formation of micelles with greater diameter and volume, thus allowing vectorization of greater quantities of oligonucleotide.

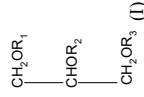
In addition, it seems that, when applied to mucosa tissue, the reverse micelle system triggers formation of lipoproteins which after a lymphatic transport then in the blood circulation cross the cellular membrane and allow delivery of the oligonucleotide, in particular the nucleic acid capable of modulating gene expression of SARS-CoV-2 virus into the cells.

After the deposition of the micro emulsion on the buccal mucosa (or rectal mucosa) in the subject, the Brownian dynamics of the reverse micelles promotes intramucosal penetration into the intercellular spaces, and in contact with the apoproteins present physiologically in the mucosa, there takes place a structure in lipoproteins vHDL and HDL.

Oligonucleotides must be soluble in water, so as not to interfere with the water/oil interface of the reverse micelles according to the invention.

An amphiphilic molecule modifies the water solubility in the nano micelles, interferes with the interface and removes the fluidity of the permanent Brownian-like motions of the micelles which is necessary for their passage in the mucosa and their absorption through the structuration in lipoproteins.

Acylglycerols can be mono- and/or diacylglycerols. In a particular embodiment, mono- or diacylglycerols preferentially used in the present invention present the following formula (I):



in which:

5 - R<sub>1</sub> is an acyl residue of a linear or branched, saturated or unsaturated fatty acid having between 14 and 24 carbon atoms, a hydrogen atom, or a mono-, di- or tri-galactose or glucose;

- R<sub>2</sub> is an acyl residue of a linear or branched, saturated or unsaturated fatty acid having between 2 and 18 carbon atoms;

10 - R<sub>3</sub> is an acyl residue of a linear or branched, saturated or unsaturated fatty acid having between 14 and 24 carbon atoms, or a hydrogen atom.

According to a particular embodiment, R<sub>1</sub> or R<sub>2</sub>, preferably only one of R<sub>1</sub> and R<sub>2</sub>, in particular only R<sub>1</sub>, represents an acyl residue of oleic acid (C18: 1[cis]-9), including in particular glycerol monooleate.

15 According to a particular aspect, R<sub>2</sub> has one unsaturated bond (e.g. ethylenic bond) and has advantageously 18 carbon atoms, preferably R<sub>2</sub> is an oleic acid residue (oleoyl group), one of its positional isomers with respect to the double bond (cis-6,7,9,11 and 13) or one of its iso-branched isomers.

According to another particular aspect, R<sub>1</sub> represents an oleoyl group.

According to another particular aspect, R<sub>2</sub> represents an acetyl group.

According to another particular aspect, R<sub>3</sub> is a hydrogen atom.

20 As a general rule, oil containing a high concentration of oleic acid will be chosen as a useful source of acylglycerols according to the invention. Such oil usually contains a high proportion of acylglycerols useful according to the invention.

According to a particular aspect of the invention, the preferred diglycerols of fatty acids are selected in the group consisting of 1,2-diolein and 1-oleoyl-2-acetyl glycerol.

25 A certain number of them, and more particularly those which are found to be the most active in the applications sought after, are also available commercially. This is the case particularly for 1-oleoyl-2-

i.e., mechanical stirring or sonication. The time of mechanical stirring or sonication is more specifically the time needed to convert the initial mixture into a visually transparent reverse micelle solution.

5 One skilled in the art knows how to select excipients or components that may be used along with the composition according to the present invention in order to keep their beneficial properties. In particular, the presence of glycerol can, when introduced in large amount, prevent the formation of reverse micelles or break the reverse micelle system. More specifically, no more than 2.5%, and preferably no glycerol (percent expressed by weight of glycerol / weight of acylglycerol) is used for the preparation of the reverse micelles of the present invention.

10 Other compounds can be introduced in step (a). One can cite for instance colouring agents and/or flavouring substances.

In an advantageous manner, the compounds cited above or the commercially available mixtures containing them are the only ingredients introduced to prepare the micelle system and consequently the only ones present in the micelle system of the invention.

15 Physical parameters, in particular time- for instance comprised between 3 and 5 minutes, in one or several times-, are dependent on the used material, volumes of the mixture and viscosity thereof. One skilled in the art can readily define such parameters. Temperature of the mixture is less than 40°C. Such a temperature avoids degradation of the reactants. Temperature is preferably ranging from 30 °C to 38 °C, more preferably from 30 °C to 35 °C.

20 The usual materials use propellers whose fast movements generate turbulences and swirls allowing interpenetration of particles and formation of reverse micelles within the mixture.

Stirring speed is preferably ranging from 200 to 2 000 r/minute, more preferably from 300 to 700 r/minute. The implemented volumes, device, and stirring speed depend on and should be adapted with the reactants and amounts thereof.

25 As described above, temperature of the mixture must not exceed 40°C. Temperature is preferably ranging from 30 °C to 38 °C, more preferably from 30 °C to 35 °C.

#### REVERSE MICELLES COMPOUNDS

##### ACYLGLYCEROL

30 Acylglycerols, more particularly acylglycerols of fatty acids, useful for the preparation of the reverse-micelle system according to the invention can be isolated from the majority of animals and more preferably plants.

Phosphatidylcholine is formed from a choline, a phosphate group, a glycerol and two fatty acids. It is actually a group of molecules, wherein the fatty acid compositions vary from one molecule to another. Phosphatidylcholine may be obtained from commercial lecithin that contains phosphatidylcholine in concentrations of 20 to 98%. The lecithin preferably used for the preparation of the reverse micelles according to the invention is Epikuron 200® and contains phosphatidylcholine at a concentration of more than 90%.

The weight ratio phospholipid/acylglycerol in compositions or reverse-micelle systems according to the invention is from 0.06 to 0.30.

#### ALCOHOLS

The alcohols useful for the preparation of the reverse-micelle system according to the invention are preferably linear or branched mono-alcohols from C2 to C6. Examples of alcohols are ethanol, 1-butanol, 2-butanol, 3-methyl-1-butanol, 2-methyl-1-propanol, 1-pentanol, 1-propanol, 2-propanol and any mixture thereof. In a particular embodiment of the invention, alcohol is ethanol.

The alcohol is preferably incorporated or comprised in the composition or reverse-micelle system in an amount by weight ranging from 5 g to 17 g with respect to 100 g of the total weight of the composition or reverse-micelle system according to the invention.

#### OLIGONUCLEOTIDES

The unmodified oligonucleotides targeting SARS-CoV-2 RNAs or targeting one or more genes of SARS-CoV-2 virus can be any nucleic acid molecule capable of modulating gene expression by down regulating or knocking down the expression of a target nucleic acid sequence of SARS-CoV-2 virus.

Down regulating or knocking down the expression of a target nucleic acid sequence can be commonly accomplished via RNA interference (RNAi). RNAi generally designates a phenomenon by which dsRNA specifically reduces expression of a target gene at post-translational level. In normal conditions, RNA interference is initiated by double-stranded RNA molecules (dsRNA) of various length, for example ranging from 15 to 30 base pair length. *In vivo*, dsRNA introduced into a cell is cleaved into a mixture of short dsRNA molecules.

Nucleic acid molecules capable of modulating gene expression by down regulating or knocking down the expression of a target nucleic acid sequence SARS-CoV-2 virus can thus include "antisense oligonucleotides", "short interfering nucleic acid" (siNA), "short interfering RNA" (siRNA), "short interfering nucleic acid molecule", "short interfering oligonucleotide molecule", "miRNA", "micro RNA", guide RNA (gRNA), short guide RNA (sgRNA) of a CRISPR system, "short hairpin RNA" (shRNA) or a mixture thereof.

acylglycerol and 1,2-dioleoylglycerol, which exist as commercial products with a high purity content. In particular, glycerol monooleate containing about 44 % of dioleic glycerol, from which about 14 % is 1,2-diolein. Such a compound is pharmaceutically accepted (*European Pharmacopoeia* (4<sup>th</sup> Edition), *USP 25/NE20*, and *Japanese Standard of food Additives*). Such product is for instance commercially available by Gattefossé Company under the name PECEOL .

The acylglycerols are preferably incorporated or comprised in the composition or reverse-micelle system in an amount by weight ranging from 55 g to 90 g with respect to 100 g of the total weight of the composition or reverse-micelle system according to the invention.

#### STEROLS

The sterols useful for the preparation of the reverse-micelle system according to the invention are preferably natural sterols, such as cholesterol or phytosterols (vegetable sterols). Sitosterol or cholesterol are the preferred sterols useful for the reverse-micelle system according to the invention.

Sitosterol and cholesterol are commercially available. More particularly, commercial sitosterol which is extracted from soya can be used. In such a product, the sitosterol generally represents from 50 to 70% by weight of the product and is generally found in a mixture with campesterol and sitostanol in respective proportions in the order of 15% each. Commercial sitosterol which is extracted from a variety of pine called tall oil can also be used. In general, it will be possible to use sitosterol in mixture with sitostanol. Preferably, said mixture comprises at least 50% sitosterol by weight of the mixture.

As mentioned above, the ratios of the lipidic constituents (sterols, acylglycerol and phospholipids) in the reverse-micelle system according to the invention can vary in a wide range, for instance the weight ratio sterols/acylglycerol can range from 0.015 to 0.05, more particularly from 0.03 to 0.04.

#### PHOSPHOLIPIDS

Phospholipids are formed of a glycerol linked to 2 fatty acids and to a phosphate group. The variability of phospholipids relies on the fatty acids that are attached to the glycerol and on the chemical groups that are susceptible to link to the phosphate group. Phospholipids are the major lipidic constituents of biological membranes.

Among phospholipids useful in the present invention may be cited phosphatidylethanolamine, phosphatidylserine, phosphatidylglycerol, diphosphatidylglycerol, phosphatidylinositol, and phosphatidylcholine.

In a particular embodiment, the phospholipid is phosphatidylcholine. Phosphatidylcholine is also known as 1,2-diacyl-glycerol-3-phosphocholine or PtdCho.

Whether RNAi is a functional antiviral pathway in mammals is still contentious, since production of siRNA molecules from long dsRNAs cannot be explicitly demonstrated in mammalian cells due to the fact that dsRNA longer than 30 bp triggers activation of interferon (IFN) response which shuts down the natural RNAi. However, mammalian cells do possess all the components of evolutionary conserved RNAi machinery that can be harnessed to inhibit the expression of cognate mRNA by exogenous siRNA molecules. The antiviral potential of siRNAs was first demonstrated against respiratory syncytial virus and thereafter numerous studies describing antiviral activity of siRNAs against viruses with DNA and RNA genomes *in vitro* and *in vivo* have been published. RNAi-based drugs thus appear to be a viable option to treat severe viral infections, against which effective vaccines or specific cure is not available yet, such as Ebola virus or emerging viruses, in particular SARS-CoV-2 virus.

The first step in production of antiviral siRNAs is *in silico* selection of highly conservative sequences in the targeted virus genome in order to achieve strong antiviral activity and avoid off-target effects.

After internalization of siRNA duplexes in treated cells, the duplexes are loaded on proteins of the "Ago" family, forming a molecular complex named "RISC" (RNA-induced silencing complex). There are 4 Ago proteins in mouse and in human, but only one (called "Ago2") is able to degrade target RNAs by endonucleolytic cleavage. That reaction generally occurs when the guide strand and the target are highly complementary (a perfect match to the "seed" [preferably nucleotides 2–7 of the guide strand] is important for target binding; and a perfect match to the central part [preferably nucleotides 8–14] of the guide strand is important for target cleavage).

According to a preferred embodiment, oligonucleotides of the present invention are designed to have complementarity to a target nucleic acid sequence of SARS-CoV-2 virus genome (such as SEQ ID NO 7). This complementarity involves at least 13 bases, typically between 13 and 25 bases, preferably at least 14 bases, even more preferably at least 18 bases of the oligonucleotides of the present invention.

The term "complementary" or "complementarity" refers herein to the ability of oligonucleotides to form base pairs with another nucleotide molecule. Base pairs are typically formed by hydrogen bonds between nucleotide units in antiparallel polynucleotide strands. Complementary polynucleotide strands can base pair in the Watson-Crick manner (e.g., A to T, A to U, C to G), or in any other manner that allows for the formation of duplexes. As persons skilled in the art are aware, when using RNA as opposed to DNA, uracil rather than thymine is the base that is considered to be complementary to adenosine. However, when a U is denoted in the context of the present invention, the ability to substitute a T is implied, unless otherwise stated. Perfect complementarity or 100 percent complementarity refers to the situation in which each nucleotide unit of the oligonucleotide strand of the invention can bind to a nucleotide unit of a second oligonucleotide strand. Less than perfect complementarity refers to the situation in which some, but not all, nucleotide units of two strands can bind with each other. For example, for two 20-

Unmodified oligonucleotides as defined above, such as unmodified siRNAs, are prone to rapid degradation by ubiquitous endo- and exonucleases and they are generally undetectable in the blood already 10 min after administration.

The oligonucleotides used in the present invention are necessarily chemically unmodified in order to be perfectly water-soluble. More specifically, unmodified oligonucleotides refer to oligonucleotides without any structural modifications at the ribose level (e.g. 2'-fluoro, 2'-methyl, and/or 2'-methoxy), at the base level and at the backbone level (e.g. phosphodiester; phosphorothioate).

According to a preferred embodiment, oligonucleotides of the present invention are at least 10, 15, 20 or 25 nucleotides (nt) long, more preferably in the range of 19 to 25 nucleotides long, or typically 19, 20, 21, 22, 23, 24 or 25 nt long.

According to a preferred embodiment, oligonucleotides of the present invention are designed to have complementarity to the target sequence. In the context of the present invention, they are more specifically designed to have complementarity to a target nucleic acid sequence of the SARS-CoV-2 virus genome. Said viral genome is for instance as described by SEQ ID NO 7.

The term RNA interference (RNAi) is used to describe gene silencing or knocking down at the mRNA level guided by small complementary non-coding RNA species. There are several classes of RNAi mediators, one of which, namely small interfering RNAs (siRNAs). The source of siRNAs during infection is viral double-stranded RNA (dsRNA), which is cleaved by cytoplasmic RNase III family enzyme Dicer into 19–27 base pair (bp) long molecules with a perfectly complementary middle region and 2-nt overhangs on both 3' ends. These siRNAs are incorporated into a multiprotein RNA-induced silencing complex (RISC). Following the strand separation, the antisense strand (i.e. guide strand) guides the RISC to recognize and cut target RNA transcripts (the other strand is called passenger strand).

Unmodified oligonucleotides as defined above, such as unmodified siRNAs, are aimed at inhibiting or reducing contagiousness of SARS-CoV-2 virus, more specifically, by protecting host from viral infection, inhibiting the expression of viral antigen or accessory genes, controlling the transcription, retro transcription or replication of viral genome, hindering the assembly of viral particles, or displaying influences in virus-host interactions.

Unmodified oligonucleotides as defined above, such as unmodified siRNAs, can thus be used to protect host from viral infection, inhibit the expression of viral antigen and accessory genes, control the transcription, retro transcription or replication of viral genome, hinder the assembly of viral particles, or display influences in virus-host interactions.

Natural human miRNAs frequently have a 5' uridine, which may be due to an intrinsically higher affinity of the Ago protein or its loading machinery (at least Ago2 binds preferentially 5' uridines and 5' adenosines).

According to a particular embodiment, the unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 virus, and in particular its guide strand, has a 5' uridine base.

5

In addition to the intended target, introduced siRNAs might bind additional mRNAs ("off-targets"). The main determinant of target recognition is a perfect match between nucleotides 2-7 of the guide strand (the "seed" of the guide strand) and the off-target RNA. If there are many off-targets, the siRNA is likely to be partially titrated, hence less efficient. And because off-targets might be (moderately) repressed by the siRNA, they could trigger unwanted secondary effects. It is thus preferable to choose siRNAs that minimize the number of off-targets, and to minimize the number of off-targets whose modest down-regulation is most susceptible to trigger phenotypic consequences in humans.

The SARS-CoV-2 virus has an RNA genome, so it is theoretically possible to target both the genomic RNA (which is about 30 kb long) and individual mRNAs (there are 9 annotated ORFs - open reading frames- in the SARS-CoV-2 genome). mRNAs are more likely to be accessible to siRNAs, because genomic RNA is largely protected by encapsidation.

15

According to a particular embodiment, the unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 targets one or more viral sequences that belong to mature mRNAs of SARS-CoV-2.

According to a particular embodiment, the unmodified oligonucleotide of the invention targets the mRNA of the longest protein produced by the virus. The function of this protein is not yet known with precision; in the viral genome, its gene is called "ORF1ab". Said gene extends from position 266 to position 21555 of the viral genome shown in SEQ ID NO 7, and siRNA of the invention more preferably targets the region between positions 14790 to 14810 of said genome.

20

Up to date, these positions of the viral genome are, at the same time, 100% conserved between all the sequenced variants of the virus genome (more than one thousand variants described to date) and they are particularly accessible to siRNAs. (little folded in on themselves).

25

According to a particular embodiment, the siRNA of the invention presents a guide strand which comprises, or consists of, one of the following sequences:

SEQ ID NO 1: 5' P-UGAUAGUAGUCAUAAUCGCCUA 3';

30 SEQ ID NO 3: 5' P-UGACUUAAAAGUUCUUUAUGCG 3';

SEQ ID NO 5: 5' P-UUAGCUAAAAGACACGAAACCCGG 3';

SEQ ID NO 8: 5' P-UGACUUAAAAGUUCUUUAUGCUC 3';

mers, if only two base pairs on one strand of the invention (e.g. guide strand) can bind with the other, the oligonucleotide strand of the invention exhibits 10 percent complementarity. In the same way, if 20 base units of one 20 nt strand (e.g. guide strand) can be bond with 20 other base units of the target gene, the oligonucleotide strands of the invention exhibit 100 percent complementarity.

In a particular aspect, the oligonucleotides of the present invention is a RNA, typically a double-stranded RNA (or RNA duplexes), in particular a small interfering RNA (siRNA), with a guide strand and a passenger strand.

10 According to a more particular embodiment, the oligonucleotides of the present invention are synthetic RNA duplexes comprising or consisting of two unmodified 21-mer oligonucleotides annealed together to form short/small interfering RNAs (siRNAs).

15 The main limitation of anti-viral RNAi is the great mutability of viruses. It can be anticipated that siRNA-resistant virus variants will emerge rapidly. In order to delay as much as possible, the emergence of such variants, it is better to target constant regions of the virus genome (*i.e.*: regions that are exactly identical among the sequenced variants of the virus).

According to a particular embodiment, the unmodified oligonucleotide targeting one or more genes of the virus SARS-CoV-2 targets constant regions of the virus genome.

20 mRNA accessibility to RISC can be hindered by RNA-binding proteins, whose binding pattern is not known. But the 5' UTR and coding sequence of mRNAs are cleaned by ribosome scanning making them more sensitive to RISC than the 3' UTR: while scoring predicted off-targets, it is advisable to focus on those with a seed match in their 3' UTR.

25 According to a particular embodiment, the unmodified oligonucleotide targeting one or more genes of the virus SARS-CoV-2 targets the 5' UTR and coding sequence of the virus genome.

mRNA accessibility to RISC can also be inhibited by mRNA secondary structures, especially short-term interactions, which are likely to re-form rapidly after ribosome scanning.

According to a particular embodiment, the unmodified oligonucleotide targeting one or more genes of the virus SARS-CoV-2 targets poorly-structured regions of the mRNA of SARS-CoV-2 virus.

30

According to another aspect, the invention relates a pharmaceutical composition comprising at least one the siRNAs as defined above, in a pharmaceutically acceptable carrier or excipient.

#### 5 USE OF REVERSE MICELLES TO DELIVER UNMODIFIED NUCLEOTIDES TARGETING VIRUS SARS-CoV-2 GENES

The pharmaceutical composition of the present invention describes unmodified oligonucleotides, such as siRNAs, targeting one or more genes of the SARS-CoV-2 Coronavirus virus, formulated in the microemulsion (or reverse micelles) as described herein and intended for the treatment of patients contaminated by this virus.

10 The unmodified oligonucleotides, such as siRNAs, targeting one or more genes of the SARS-CoV-2 are present in the aqueous core of the reverse micelles.

The amount of unmodified oligonucleotides, such as siRNAs, targeting one or more genes of SARS-CoV-2 incorporated into the reverse micelle system is determined by their solubility in the hydrophilic phase (aqueous core). Preferably, the amount of unmodified oligonucleotides, such as siRNAs, targeting one or more genes of SARS-CoV-2 included in the reverse micelle system depends on their size.

15 The reverse micelles of the invention allow the oligonucleotide included therein to be administered and transported to cells with a high degree of protection in lipoprotein HDL and vHDL, in particular without affecting its stability.

20 It is known today that a reverse-micelle system can be used for the preparation of nanomaterials, which act as micro reactors. The activity and stability of bio molecules can be controlled, mainly by the concentration of water in this medium.

An object of the invention concerns a pharmaceutical composition comprising reverse micelles as defined above and at least a pharmaceutically acceptable carrier, excipient or support, more specifically for use in the treatment of COVID-19.

25 According to a particular embodiment, the pharmaceutical composition is in the form of airless bottle, a capsule, a caplet, an aerosol, a spray, a solution or a soft elastic gelatin capsule.

A further object of the invention concerns the use of reverse micelles as defined above for preparing a pharmaceutical composition intended for the treatment of COVID-19.

SEQ ID NO 10: 5' P-UAUAGCUAAAAGACACGAACCC 3';  
 SEQ ID NO 11: 5' P-AUAGCUAAAAGACACGAACCCG 3';  
 SEQ ID NO 12: 5' P-UUGAGUGCAUUAUCCAAG 3';  
 SEQ ID NO 13: 5' P-CUUGACUGCCGUCUCUGC 3';  
 SEQ ID NO 14: 5' P-GUUGAGUGCAUUAUCCAAC 3';  
 SEQ ID NO 15: 5' P-UCCUGAUUAUGUACAACACCG 3'.

Preferably, the siRNA of the invention presents a guide strand which comprises, or consists of, SEQ ID NO 1.

10 According to a preferred embodiment, the siRNA duplexes of the invention, with guide strand and passenger strand, comprises, or consists of, one of the following duplex sequences:

siRNA n°1

SEQ ID NO 1: guide strand: 5' P-UGAUAGUAGUCAUAUCCGUA 3';

SEQ ID NO 2: passenger strand: 5' GCGAUUAUGACUACUAUUUUA 3'.

siRNA n°2

SEQ ID NO 3: guide strand: 5' P-UGACUAAAAGUUCUUUAUCCG 3';

SEQ ID NO 4: passenger strand: 5' CAUAAAAGAACUUUAAAGUCCUC 3'.

siRNA n°3

SEQ ID NO 5: guide strand: 5' P-UUAGCUAAAAGACACGAACCCG 3';

SEQ ID NO 6: passenger strand: 5' GGUUCGUGUCUUUAGCUACUC 3'.

siRNA n°4

SEQ ID NO 8: guide strand: 5' P-UGACUAAAAGUUCUUUAUCCG 3';

SEQ ID NO 9: passenger strand: 5' GCAUAAAAGAACUUUAAAGUUUCU 3'.

25 Preferably, the siRNA of the invention comprises, or consists of, the duplex sequences of SEQ ID NO 1 and 2 (siRNA n°1).

30 The siRNA with guide strands corresponding to SEQ ID 10-15 also comprise passenger strands as to form effective siRNAs duplexes, as described above.

Schematic of said siRNA structures are dispatched in **FIGURE 1** (: Watson-Crick base pair; x: mismatch; : GU wobble).

According to a particular aspect, the invention relates to the siRNAs as identified above.



is sufficient to induce reduction of COVID-19 symptoms or to induce alleviation of one or more symptoms of COVID-19.

The subject (or patient) to treat is any mammal, preferably a human being. Preferably, the subject is a human patient, whatever its age or sex. New-borns, infants, children are included as well. More preferably, the patient or subject according to the invention is suspected to be infected by SARS-CoV-2 or has been diagnosed to have CoVID-19 or has been diagnosed as infected by SARS-CoV-2. The standard method of diagnosis is by reverse transcription polymerase chain reaction (rRT-PCR) from a nasopharyngeal swab or throat swab. The infection can also be diagnosed from a combination of symptoms, risk factors and a chest CT scan (Computer Tomography scan) showing features of pneumonia.

As used herein, the terms "mucosa" and "mucosal" refer to a mucous tissue such as of the respiratory, digestive, or genital tissue. "Mucosal delivery", "mucosal administration" and analogous terms as used herein refer to the administration of a composition through a mucosal tissue. "Mucosal delivery", "mucosal administration" and analogous terms include, but are not limited to, the delivery of a composition through preferably buccal administration, bronchi, gingival, lingual, buccal, vaginal, rectal, and gastro-intestinal mucosal tissue. Administration according to the invention is more preferably carried out via buccal mucosa or rectal mucosa.

#### **EXAMPLES**

The following examples are intended to exemplify the operation of the present invention but not to limit its scope.

#### **Example 1: Manufacture of a drug for the treatment of infectious pathologies linked to SARS-CoV-2 Coronavirus - siRNA #1**

The aim of this study was to evaluate by visual determination the formulation and the stability of the siRNA #1 targeting SARS-CoV-2 in the reverse microemulsion.

8.5 g of lecithin were dissolved in 6.8 g of absolute ethanol by magnetic stirring at 300 r/min for 15 minutes at room temperature. 1.4 g of sitosterol were added to the mixture and stirred in the same conditions. 32.6 g of glycerol monooleate was added thereto and magnetic stirring was carried out at 500 r/min for 45 minutes at 37°C.

168 mg of a siRNA aqueous solution containing 1,06 mg of siRNA #1 were added to 1148.0 mg of the oil mixture as prepared above and then stirred at room temperature by magnetic stirring at 700 r/min for 30 minutes.

The present invention further concerns a method for the treatment of COVID-19, wherein the method comprises the step of administering into a subject in need of such treatment a therapeutically efficient amount of one or more unmodified oligonucleotides as defined above.

5 More specifically, administration of one or more unmodified oligonucleotides in reverse micelle system as defined herein or pharmaceutical composition comprising the same is a mucosal delivery.

As pharmaceutically acceptable excipient, vehicle or carrier, any excipient, vehicle or carrier well-known to the person skilled in the art may be used. Other additives well-known to the person skilled in the art such as stabilisers, drying agents, binders or pH buffers may also be used. Preferred excipients in accordance with the invention promote adherence of the finished product to the mucosa.

10 The compositions of the invention can be administered in different ways, in particular via the oral, nasal, vaginal or rectal route, with a buccal, nasal, vaginal or digestive absorption, or more generally via mucosal tissue absorption. The composition of the invention is preferably administered by buccal route or rectal route, with a buccal mucosa or rectal mucosa absorption, respectively.

15 Within the context of the invention, the term treatment denotes curative, symptomatic, and preventive treatment. As used herein, the term "treatment" of COVID-19 refers to any act intended to extend life span of subjects (or patients) such as therapy and retardation of the disease progression. The treatment can be designed to eradicate the disease, to stop the progression of the disease, and/or to promote the regression of the disease. The term "treatment" of a disease also refers to any act intended to decrease one or more mild symptoms associated with the disease, including fever, cough, shortness of breath, muscle pain, sputum production and/or sore throat. The term "treatment" of the disease also refers to any act intended to decrease one or more severe symptoms associated with the disease, including pneumonia and/or multi-organ failure. More specifically, the treatment according to the invention is intended to delay the appearance of, alleviate, or hinder, the mild symptoms and more particularly the severe symptoms of COVID-19, such as COVID-19 associated pneumonia or multi-organ failure.

20 As used herein, the term "therapeutically effective amount" is intended an amount of unmodified oligonucleotides as defined above, administered to a patient that is sufficient to constitute a treatment of COVID-19 as defined above. In a particular embodiment, the therapeutically effective amount to be administered is an amount sufficient to down regulate or knock down the expression of a target nucleic acid of SARS-CoV-2 virus. The amount of unmodified oligonucleotides as defined above to be administered can be determined by standard procedure well known by those of ordinary skill in the art. Physiological data of the patient (e.g. age, size, and weight), the routes of administration and the disease to be treated have to be taken into account to determine the appropriate dosage. One skilled in the art will recognize that the amount of unmodified oligonucleotides to be administered will be an amount that

TEST ITEMS:

Name	NanosirRNA®#1	NanosirRNA®#2	NanosirRNA®#SCR
<b>Medesis Batch</b>	N° 210006	N°: 210007	N°: 210009
<b>Composition</b>	water in oil microemulsion containing 1,0 mg/mL of siRNA#1, targeting the SARS-Cov-2 coronavirus.	water in oil microemulsion containing 1,0 mg/mL of siRNA#2, targeting the SARS-Cov-2 coronavirus.	water in oil microemulsion containing 1,0 mg/mL of scrambled siRNA.
<b>Description</b>	Homogeneous yellow oily liquid Viscosity approx. 80 - 100 mPa.s at +25°C and approx 40-60 mPa.s at +37°C.		
<b>Storage conditions</b>	at room temperature (approximately +15 to +25°C).		
<b>Expiry date</b>	01/2022		
<b>Quantity</b>	6 mL (3 vials of 2 mL) Clear glass vials containing 2 mL of drug product		

ADMINISTRATION:

A constant dosage-volume of 1 mL/kg/day is used for all groups of animals. The quantity of dosage form administered to each animal is adjusted according to the bodyweight.

The dosage form is administered by rectal deposit without anesthesia, using a pipet tip with automatic pipetman.

Food and water will be removed before product administration and will be given 30 minutes after administration.

## ANIMALS:

**Strain:** Golden Syrian hamster model of SARS-COV-2

**Number:** 24 female SG hamsters

6-8 weeks old female SG hamsters of 90-120g are ear-tagged and randomized in the different treatment groups.

The microemulsion was limpid, monophasic and thermodynamically stable. These experiments show that the siRNA #1 is well formulated in the reverse micelle at 800 µg/ml and has no impact on the stability of the system.

Other microemulsions with close contents are also prepared and result in similar reverse micelle with a stable system.

**Example 2 : Manufacture of a drug for the treatment of infectious pathologies linked to SARS-CoV-2 Coronavirus - siRNA #2, siRNA #3, siRNA #4**

The aim of this study was to evaluate by visual determination the formulation and the stability of the siRNA #2, #3 et #4 targeting SARS-CoV-2 in the reverse microemulsion.

8.5 g of lecithin were dissolved in 6.8 g of absolute ethanol by magnetic stirring at 300 r/min for 15 minutes at room temperature. 1.4 g of sitosterol were added to the mixture and stirred in the same conditions. 32.6 g of glycerol monooleate was added thereto and magnetic stirring was carried out at 500 r/min for 45 minutes at 37°C.

168 mg of a siRNA aqueous solution were added to 1148.0 mg of the oil mixture as prepared above and then stirred at room temperature by magnetic stirring at 700 r/min for 30 minutes. The composition of each tested siRNA aqueous solution is in following table 1 :

Table 1 : tested siRNA aqueous solutions

#2	#3	#4
siRNA aqueous solution containing 1,06 mg of siRNA #2	siRNA aqueous solution containing 1,06 mg of siRNA #3	siRNA aqueous solution containing 1,06 mg of siRNA #4

Each microemulsion obtained was limpid, monophasic and thermodynamically stable. These experiments show that the siRNA #2, siRNA #3 and siRNA #4 are well formulated in the reverse micelles at 800 µg/ml and have no impact on the stability of the systems.

**Example 3: Efficacy study by rectal mucosa route in SG hamster model of SARS-COV-2**

OBJECTIVE:

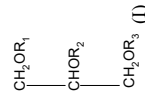
Evaluation of 2 reverse micelle systems according to the invention (with siRNA#1 and with siRNA#2) in hamster model of SARS-COV-2 after 4-day treatment.

## CLAIMS

1. Reverse micelle system comprising at least one sterol, acylglycerol, phospholipid, an alcohol, water and at least one unmodified oligonucleotide targeting one or more genes of SARS-CoV-2 virus.

5 2. The reverse micelle system according to claim 1, wherein the micelles present aqueous cores of around 4 nm, preferably from 3 to 5 nm, more preferably from 3.5 to 5 nm, in particular from 3.7 to 4.5 nm.

3. The reverse micelle system according to claim 1, wherein acylglycerol presents the following formula (I):



10

in which:

- R<sub>1</sub> is an acyl residue of a linear or branched, saturated or unsaturated fatty acid having between 14 and 24 carbon atoms, a hydrogen atom, or a mono-, di- or tri-galactose or glucose;

- R<sub>2</sub> is an acyl residue of a linear or branched, saturated or unsaturated fatty acid having between 2 and 18 carbon atoms;

15

- R<sub>3</sub> is an acyl residue of a linear or branched, saturated or unsaturated fatty acid having between 14 and 24 carbon atoms, or a hydrogen atom.

4. The reverse micelle system according to any one of claims 1-3, wherein the at least one sterol is sitosterol, and/or phospholipid is lecithin, and/or alcohol is ethanol, and/or acylglycerol is glycerol monooleate.

20

5. The reverse micelle system according to any one of claims 1-4, wherein the unmodified oligonucleotides are selected in the group consisting of antisense oligonucleotides, short interfering nucleic acid (siNA), short interfering RNA (siRNA), short interfering nucleic acid molecule, short interfering oligonucleotide molecule, miRNA, micro RNA, guide RNA (gRNA), short guide RNA (sgRNA) of a CRISPR system, short hairpin RNA (shRNA) and a mixture thereof.

25

6. The reverse micelle system according to any one of claims 1-5, wherein the unmodified oligonucleotides are at least 10, 15, 20 or 25 nucleotides (nt) long, more preferably in the range of 19 to 25 nucleotides long, or typically 19, 20, 21, 22, 23, 24 or 25 nt long.

## TREATMENT:

Group	hamsters	P 6 virus TCID <sub>50</sub> inoculum/50 $\mu$ L	Treatment	Dose	Frequency dosing	MOA
Group 1	6 WT	1.89E+06	siRNA 1	1,0 mg/mL	once	Rectal deposit
Group 2	6 WT	1.89E+06	siRNA 2	1,0 mg/mL	once	Rectal deposit
Group 3	6 WT	1.89E+06	Scramble	1,0 mg/mL	once	Rectal deposit
	18 hamsters					

## DURATION:

The dose formulations will be administrated once daily for a period of 4 days.

5

## COLLECTION AND ANALYSING OF SAMPLES:

Animals are sacrificed: lung and blood collection

Lung: 1) Quantification of viral load by real-time quantitative RT-qPCR

2) Quantification of infectious viral content by (end-point) titration

10 3) Histological examination for evaluation of inflammation in lung tissues.

SEQ ID NO 1: guide strand: 5' P-UGAUAGUAGUCAUAAUCCGCUA 3';  
 SEQ ID NO 3: guide strand: 5' P-UGACUUAAAAGUUCUUUAUGCG 3';  
 SEQ ID NO 5: guide strand: 5' P-UUAGCUAAAAGACACGAAACCGG 3';  
 SEQ ID NO 8: guide strand: 5' P-UGACUUAAAAGUUCUUUAUGCUC 3';  
 SEQ ID NO 10: guide strand: 5' P- UAUAGCUAAAGACACGAAACCC 3';  
 SEQ ID NO 11: guide strand: 5' P- AUAGCUAAAGACACGAAACCGG 3';  
 SEQ ID NO 12: guide strand: 5' P- UUGAGUGCAUCAUUAUCCAAG 3';  
 SEQ ID NO 13: guide strand: 5' P- CUUGACUGCGCCUCUCUCG 3';  
 SEQ ID NO 14: guide strand: 5' P- GUUGAGUGCAUCAUUAUCCAC 3';  
 SEQ ID NO 15: guide strand: 5' P- UCCUGAUUAUGUACAACACCG 3';

16. A siRNA duplex, with guide strand and passenger strand, which comprises, or consists of, one of the following duplex sequences:

siRNA n°1  
 SEQ ID NO 1: guide strand: 5' P-UGAUAGUAGUCAUAAUCCGCUA 3';  
 SEQ ID NO 2: passenger strand: 5' GCGAUUAUGACUACUUAUUUA 3'.  
 siRNA n°2  
 SEQ ID NO 3: guide strand: 5' P-UGACUUAAAAGUUCUUUAUGCG 3';  
 SEQ ID NO 4: passenger strand: 5' CAUAAAAGAACUUUAAAGUCCUC 3'.  
 siRNA n°3  
 SEQ ID NO 5: guide strand: 5' P-UUAGCUAAAAGACACGAAACCGG 3';  
 SEQ ID NO 6: passenger strand: 5' GGUUCGUGUCUUUAGCUACUC 3'.  
 siRNA n°4  
 SEQ ID NO 8: guide strand: 5' P-UGACUUAAAAGUUCUUUAUGCUC 3';  
 SEQ ID NO 9: passenger strand: 5' GCAUAAAGAACUUUAAAGUUUUCU 3'.

17. A pharmaceutical composition comprising at least one the siRNAs as defined in any one of claims 15 and 16, in a pharmaceutically acceptable carrier or excipient

30

7. The reverse micelle system according to any one of claims 1-6, wherein the unmodified oligonucleotides are synthetic RNA duplexes comprising or consisting of two unmodified 21-mer oligonucleotides annealed together to form siRNAs.

8. The reverse micelle system according to claim 8, wherein the siRNA presents a guide strand which comprises, or consists of, one of the following sequences:

SEQ ID NO 1: 5' P-UGAUAGUAGUCAUAAUCCGCUA 3';  
 SEQ ID NO 3: 5' P-UGACUUAAAAGUUCUUUAUGCG 3';  
 SEQ ID NO 5: 5' P-UUAGCUAAAAGACACGAAACCGG 3'.  
 SEQ ID NO 8: 5' P-UGACUUAAAAGUUCUUUAUGCUC 3';  
 SEQ ID NO 10: 5' P- UAUAGCUAAAGACACGAAACCC 3';  
 SEQ ID NO 11: 5' P- AUAGCUAAAGACACGAAACCGG 3';  
 SEQ ID NO 12: 5' P- UUGAGUGCAUCAUUAUCCAAG 3';  
 SEQ ID NO 13: 5' P- CUUGACUGCGCCUCUCUCG 3';  
 SEQ ID NO 14: 5' P- GUUGAGUGCAUCAUUAUCCAC 3';  
 SEQ ID NO 15: 5' P- UCCUGAUUAUGUACAACACCG 3';

9. The reverse micelle system according to claim 8, wherein the siRNA presents a guide strand comprising, or consisting of, siRNA n°1.

10. A method for the preparation of the reverse micelle system as defined in any one of the preceding claims, wherein it comprises the following steps

(a) Contacting (i) sterol, (ii) acylglycerol, preferably diacylglycerol of fatty acids, (iii) phospholipid, preferably phosphatidylcholine, (iv) alcohol, (v) water, preferably purified water, and (vi) at least one unmodified oligonucleotide capable of targeting one or more genes of SARS-CoV-2 virus,

(b) Stirring mixture obtained in step (a), at 40 °C or less, and for a time sufficient to obtain formation of reverse micelles, said stirring being carried out mechanically or by sonication.

11. A pharmaceutical composition comprising a reverse micelle system as defined in any one of claims 1-9, and at least a pharmaceutically acceptable carrier, excipient or support;

12. The composition according to claim 11 for use in the treatment of COVID-19.

13. The composition according to claim 11 for use in the treatment of COVID-19 associated pneumonia or multi-organ failure.

14. The composition according to any one of claims 11-13, wherein the composition is administered by oral route or rectal route, with a buccal mucosa or rectal mucosa absorption, respectively.

15. A siRNA presenting a guide strand which comprises, or consists of, one of the following sequences:

35

**ABSTRACT**

The present invention relates to specific reverse micelle system of the invention which allows the administration and intracellular delivery of unmodified oligonucleotide, such as siRNA, targeting one or more genes of the SARS-CoV-2 virus. The reverse micelle system of the invention is thus particularly useful for the treatment of the viral pathology linked to the SARS-CoV-2 virus.

# Bibliography

- Adams, B. D., Christine Parsons, and Frank J. Slack (June 2016). "The tumor-suppressive and potential therapeutic functions of miR-34a in epithelial carcinomas". In: *Expert Opin Ther Targets* 20.6, pp. 737–753. ISSN: 1744-7631. DOI: 10.1517/14728222.2016.1114102.
- Agarwal, V. et al. (Aug. 12, 2015). "Predicting effective microRNA target sites in mammalian mRNAs". In: *Elife* 4. ISSN: 2050-084X. DOI: 10.7554/eLife.05005.
- Agranat-Tamir, L. et al. (Apr. 2014). "Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome". In: *Nucleic Acids Res* 42.7, pp. 4640–4651. ISSN: 1362-4962. DOI: 10.1093/nar/gkt1413.
- Alerasool, N. et al. (Nov. 2020). "An efficient KRAB domain for CRISPRi applications in human cells". In: *Nat Methods* 17.11, pp. 1093–1096. ISSN: 1548-7105. DOI: 10.1038/s41592-020-0966-x.
- Alisch, R. S. et al. (Dec. 28, 2007). "Argonaute2 is essential for mammalian gastrulation and proper mesoderm formation". In: *PLoS Genet* 3.12, e227. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.0030227.
- Almeida, M. V., M. A. Andrade-Navarro, and R. F. Ketting (Jan. 15, 2019). "Function and Evolution of Nematode RNAi Pathways". In: *Noncoding RNA* 5.1, p. 8. ISSN: 2311-553X. DOI: 10.3390/ncrna5010008.
- Alon, S. et al. (Aug. 2012). "Systematic identification of edited microRNAs in the human brain". In: *Genome Res* 22.8, pp. 1533–1540. ISSN: 1549-5469. DOI: 10.1101/gr.131573.111.
- Altuvia, Y. et al. (2005). "Clustering and conservation patterns of human microRNAs". In: *Nucleic Acids Res* 33.8, pp. 2697–2706. ISSN: 1362-4962. DOI: 10.1093/nar/gki567.
- Alvarez-Saavedra, E. and H. R. Horvitz (Feb. 23, 2010). "Many families of *C. elegans* microRNAs are not essential for development or viability". In: *Curr Biol* 20.4, pp. 367–373. ISSN: 1879-0445. DOI: 10.1016/j.cub.2009.12.051.
- Ambros, V. (Apr. 7, 1989). "A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans*". In: *Cell* 57.1, pp. 49–57. ISSN: 0092-8674. DOI: 10.1016/0092-8674(89)90171-2.
- Ambros, V. and H. R. Horvitz (Oct. 26, 1984). "Heterochronic mutants of the nematode *Caenorhabditis elegans*". In: *Science* 226.4673, pp. 409–416. ISSN: 0036-8075. DOI: 10.1126/science.6494891.
- Ambros, V. and H. R. Horvitz (June 1987). "The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events". In: *Genes Dev* 1.4, pp. 398–414. ISSN: 0890-9369. DOI: 10.1101/gad.1.4.398.
- Ambros, V. et al. (Mar. 2003). "A uniform system for microRNA annotation". In: *RNA* 9.3, pp. 277–279. ISSN: 1355-8382. DOI: 10.1261/rna.2183803.
- Ameres, S. L. et al. (June 18, 2010). "Target RNA-directed trimming and tailing of small silencing RNAs". In: *Science* 328.5985, pp. 1534–1539. ISSN: 1095-9203. DOI: 10.1126/science.1187058.
- Arur, S. et al. (Mar. 24, 2009). "Multiple ERK substrates execute single biological processes in *Caenorhabditis elegans* germ-line development". In: *Proc Natl Acad Sci U S A* 106.12, pp. 4776–4781. ISSN: 1091-6490. DOI: 10.1073/pnas.0812285106.
- Arvey, A. et al. (Apr. 20, 2010). "Target mRNA abundance dilutes microRNA and siRNA activity". In: *Mol Syst Biol* 6, p. 363. ISSN: 1744-4292. DOI: 10.1038/msb.2010.24.

- Auyeung, V. C. et al. (Feb. 14, 2013). "Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing". In: *Cell* 152.4, pp. 844–858. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.01.031.
- Avesson, L. et al. (Oct. 2012). "MicroRNAs in Amoebozoa: deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated microRNAs". In: *RNA* 18.10, pp. 1771–1782. ISSN: 1469-9001. DOI: 10.1261/rna.033175.112.
- Axtell, M. J., Jakub O. Westholm, and Eric C. Lai (2011). "Vive la différence: biogenesis and evolution of microRNAs in plants and animals". In: *Genome Biol* 12.4, p. 221. ISSN: 1474-760X. DOI: 10.1186/gb-2011-12-4-221.
- Babiarz, J. E. et al. (Oct. 15, 2008). "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs". In: *Genes Dev* 22.20, pp. 2773–2785. ISSN: 0890-9369. DOI: 10.1101/gad.1705308.
- Baccarini, A. et al. (Mar. 8, 2011). "Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells". In: *Curr Biol* 21.5, pp. 369–376. ISSN: 1879-0445. DOI: 10.1016/j.cub.2011.01.067.
- Bader, A. G. (2012). "miR-34 - a microRNA replacement therapy is headed to the clinic". In: *Front Genet* 3, p. 120. ISSN: 1664-8021. DOI: 10.3389/fgene.2012.00120.
- Baek, D. et al. (Sept. 4, 2008). "The impact of microRNAs on protein output". In: *Nature* 455.7209, pp. 64–71. ISSN: 1476-4687. DOI: 10.1038/nature07242.
- Bagchi, A. and A. A. Mills (Apr. 15, 2008). "The quest for the 1p36 tumor suppressor". In: *Cancer Res* 68.8, pp. 2551–2556. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-07-2095.
- Barker, H. M. et al. (Jan. 13, 1994). "Three genes for protein phosphatase 1 map to different human chromosomes: sequence, expression and gene localisation of protein serine/threonine phosphatase 1 beta (PPP1CB)". In: *Biochim Biophys Acta* 1220.2, pp. 212–218. ISSN: 0006-3002. DOI: 10.1016/0167-4889(94)90138-4.
- Bartel, D. P. (Jan. 23, 2004). "MicroRNAs: genomics, biogenesis, mechanism, and function". In: *Cell* 116.2, pp. 281–297. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(04)00045-5.
- Bartel, D. P. (Jan. 23, 2009). "MicroRNAs: target recognition and regulatory functions". In: *Cell* 136.2, pp. 215–233. ISSN: 1097-4172. DOI: 10.1016/j.cell.2009.01.002.
- Bartel, D. P. (Mar. 22, 2018). "Metazoan MicroRNAs". In: *Cell* 173.1, pp. 20–51. ISSN: 1097-4172. DOI: 10.1016/j.cell.2018.03.006.
- Bartel, D. P. and CZ Chen (May 2004). "Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs". In: *Nat Rev Genet* 5.5, pp. 396–400. ISSN: 1471-0056. DOI: 10.1038/nrg1328.
- Barth, S. et al. (Feb. 2008). "Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5". In: *Nucleic Acids Res* 36.2, pp. 666–675. ISSN: 1362-4962. DOI: 10.1093/nar/gkm1080.
- Baskerville, S. and D. P. Bartel (Mar. 2005). "Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes". In: *RNA* 11.3, pp. 241–247. ISSN: 1355-8382. DOI: 10.1261/rna.7240905.
- Basyuk, E. et al. (Nov. 15, 2003). "Human let-7 stem-loop precursors harbor features of RNase III cleavage products". In: *Nucleic Acids Res* 31.22, pp. 6593–6597. ISSN: 0305-1048. DOI: 10.1093/nar/gkg855.
- Beg, M. S. et al. (Apr. 2017). "Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors". In: *Invest New Drugs* 35.2, pp. 180–188. ISSN: 1573-0646. DOI: 10.1007/s10637-016-0407-y.
- Berezikov, E. et al. (Oct. 26, 2007). "Mammalian mirtron genes". In: *Mol Cell* 28.2, pp. 328–336. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2007.09.028.

- Berezikov, Eugene et al. (Feb. 2011). "Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence". In: *Genome Res* 21.2, pp. 203–215. ISSN: 1549-5469. DOI: 10.1101/gr.116657.110.
- Bernstein, E. et al. (Jan. 18, 2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference". In: *Nature* 409.6818, pp. 363–366. ISSN: 0028-0836. DOI: 10.1038/35053110.
- Bernstein, E. et al. (Nov. 2003). "Dicer is essential for mouse development". In: *Nat Genet* 35.3, pp. 215–217. ISSN: 1061-4036. DOI: 10.1038/ng1253.
- Bertola, D. et al. (Mar. 2017). "The recurrent PPP1CB mutation p.Pro49Arg in an additional Noonan-like syndrome individual: Broadening the clinical phenotype". In: *Am J Med Genet A* 173.3, pp. 824–828. ISSN: 1552-4833. DOI: 10.1002/ajmg.a.38070.
- Bitetti, A. et al. (Mar. 2018). "MicroRNA degradation by a conserved target RNA regulates animal behavior". In: *Nat Struct Mol Biol* 25.3, pp. 244–251. ISSN: 1545-9985. DOI: 10.1038/s41594-018-0032-x.
- Bogerd, H. P. et al. (Jan. 15, 2010). "A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs". In: *Mol Cell* 37.1, pp. 135–142. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2009.12.016.
- Bohmert, K. et al. (Jan. 2, 1998). "AGO1 defines a novel locus of *Arabidopsis* controlling leaf development". In: *EMBO J* 17.1, pp. 170–180. ISSN: 0261-4189. DOI: 10.1093/emboj/17.1.170.
- Bohnsack, M. T., K. Czapinski, and D. Görlich (Feb. 2004). "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs". In: *RNA* 10.2, pp. 185–191. ISSN: 1355-8382. DOI: 10.1261/rna.5167604.
- Bommer, G. T. et al. (Aug. 7, 2007). "p53-mediated activation of miRNA34 candidate tumor-suppressor genes". In: *Curr Biol* 17.15, pp. 1298–1307. ISSN: 0960-9822. DOI: 10.1016/j.cub.2007.06.068.
- Bortolamiol-Becet, D. et al. (July 16, 2015). "Selective Suppression of the Splicing-Mediated MicroRNA Pathway by the Terminal Uridyltransferase Tailor". In: *Mol Cell* 59.2, pp. 217–228. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2015.05.034.
- Bosson, A. D., Jesse R. Zamudio, and Phillip A. Sharp (Nov. 6, 2014). "Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition". In: *Mol Cell* 56.3, pp. 347–359. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.09.018.
- Bracht, J. et al. (Oct. 2004). "Trans-splicing and polyadenylation of let-7 microRNA primary transcripts". In: *RNA* 10.10, pp. 1586–1594. ISSN: 1355-8382. DOI: 10.1261/rna.7122604.
- Brennecke, J. et al. (Mar. 2005). "Principles of microRNA-target recognition". In: *PLoS Biol* 3.3, e85. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0030085.
- Brenner, J. L. et al. (July 27, 2010). "Loss of individual microRNAs causes mutant phenotypes in sensitized genetic backgrounds in *C. elegans*". In: *Curr Biol* 20.14, pp. 1321–1325. ISSN: 1879-0445. DOI: 10.1016/j.cub.2010.05.062.
- Briskin, D., Peter Y. Wang, and David P. Bartel (July 28, 2020). "The biochemical basis for the cooperative action of microRNAs". In: *Proc Natl Acad Sci U S A* 117.30, pp. 17764–17774. ISSN: 1091-6490. DOI: 10.1073/pnas.1920404117.
- Brüning-Richardson, A. et al. (2011). "EB1 is required for spindle symmetry in mammalian mitosis". In: *PLoS One* 6.12, e28884. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0028884.
- Byrne, J. A. and Jana Christopher (Feb. 2020). "Digital magic, or the dark arts of the 21st century-how can journals and peer reviewers detect manuscripts and publications from paper mills?" In: *FEBS Lett* 594.4, pp. 583–589. ISSN: 1873-3468. DOI: 10.1002/1873-3468.13747.



- Béthune, J., Caroline G. Artus-Revel, and Witold Filipowicz (Aug. 2012). "Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells". In: *EMBO Rep* 13.8, pp. 716–723. ISSN: 1469-3178. DOI: 10.1038/embor.2012.82.
- Cai, X., C. H. Hagedorn, and B. R. Cullen (Dec. 2004). "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs". In: *RNA* 10.12, pp. 1957–1966. ISSN: 1355-8382. DOI: 10.1261/rna.7135204.
- Caygill, E. E. and L. A. Johnston (July 8, 2008). "Temporal regulation of metamorphic processes in *Drosophila* by the let-7 and miR-125 heterochronic microRNAs". In: *Curr Biol* 18.13, pp. 943–950. ISSN: 0960-9822. DOI: 10.1016/j.cub.2008.06.020.
- Cazalla, D. and J. A. Steitz (2010). "Down-regulation of a host microRNA by a viral noncoding RNA". In: *Cold Spring Harb Symp Quant Biol* 75, pp. 321–324. ISSN: 1943-4456. DOI: 10.1101/sqb.2010.75.009.
- Chang, J. et al. (July 2004). "miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1". In: *RNA Biol* 1.2, pp. 106–113. ISSN: 1555-8584. DOI: 10.4161/rna.1.2.1066.
- Chang, K. et al. (Jan. 7, 2013). "Creating an miR30-Based shRNA Vector". In: *Cold Spring Harb Protoc* 2013.7, pdb.prot075853. ISSN: 1940-3402, 1559-6095. DOI: 10.1101/pdb.prot075853. URL: <http://cshprotocols.cshlp.org/content/2013/7/pdb.prot075853>.
- Chang, T-C et al. (June 8, 2007). "Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis". In: *Mol Cell* 26.5, pp. 745–752. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2007.05.010.
- Chapat, C. et al. (May 23, 2017). "Cap-binding protein 4EHP effects translation silencing by microRNAs". In: *Proc Natl Acad Sci U S A* 114.21, pp. 5425–5430. ISSN: 1091-6490. DOI: 10.1073/pnas.1701488114.
- Cheloufi, S. et al. (June 3, 2010). "A dicer-independent miRNA biogenesis pathway that requires Ago catalysis". In: *Nature* 465.7298, pp. 584–589. ISSN: 1476-4687. DOI: 10.1038/nature09092.
- Chen, C-Y A. and A-B Shyu (2011). "Mechanisms of deadenylation-dependent decay". In: *Wiley Interdiscip Rev RNA* 2.2, pp. 167–183. ISSN: 1757-7004. DOI: 10.1002/wrna.40.
- Chen, S. and G. Gao (Oct. 2017). "MicroRNAs recruit eIF4E2 to repress translation of target mRNAs". In: *Protein Cell* 8.10, pp. 750–761. ISSN: 1674-8018. DOI: 10.1007/s13238-017-0444-0.
- Chen, Y. et al. (June 5, 2014a). "A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing". In: *Mol Cell* 54.5, pp. 737–750. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.03.034.
- Chen, YW et al. (Dec. 22, 2014b). "Systematic study of *Drosophila* microRNA functions using a collection of targeted knockout mutations". In: *Dev Cell* 31.6, pp. 784–800. ISSN: 1878-1551. DOI: 10.1016/j.devcel.2014.11.029.
- Chendrimada, T. P. et al. (Aug. 4, 2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing". In: *Nature* 436.7051, pp. 740–744. ISSN: 1476-4687. DOI: 10.1038/nature03868.
- Cheung, V. G. et al. (Mar. 2003). "Natural variation in human gene expression assessed in lymphoblastoid cells". In: *Nat Genet* 33.3, pp. 422–425. ISSN: 1061-4036. DOI: 10.1038/ng1094.
- Chi, S. W. et al. (July 23, 2009). "Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps". In: *Nature* 460.7254, pp. 479–486. ISSN: 1476-4687. DOI: 10.1038/nature08170.
- Chiang, H. R. et al. (May 15, 2010). "Mammalian microRNAs: experimental evaluation of novel and previously annotated genes". In: *Genes Dev* 24.10, pp. 992–1009. ISSN: 1549-5477. DOI: 10.1101/gad.1884710.

- Choi, Y. J. et al. (Oct. 23, 2011). "miR-34 miRNAs provide a barrier for somatic cell reprogramming". In: *Nat Cell Biol* 13.11, pp. 1353–1360. ISSN: 1476-4679. DOI: 10.1038/ncb2366.
- Choi, Y. J. et al. (Feb. 10, 2017). "Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells". In: *Science* 355.6325, eaag1927. ISSN: 1095-9203. DOI: 10.1126/science.aag1927.
- Chong, M. M. W. et al. (Sept. 1, 2008). "The RNaseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease". In: *J Exp Med* 205.9, pp. 2005–2017. ISSN: 1540-9538. DOI: 10.1084/jem.20081219.
- Chong, M. M. W. et al. (Sept. 1, 2010). "Canonical and alternate functions of the microRNA biogenesis machinery". In: *Genes Dev* 24.17, pp. 1951–1960. ISSN: 0890-9369. DOI: 10.1101/gad.1953310.
- Chung, WJ et al. (June 3, 2008). "Endogenous RNA interference provides a somatic defense against Drosophila transposons". In: *Curr Biol* 18.11, pp. 795–802. ISSN: 0960-9822. DOI: 10.1016/j.cub.2008.05.006.
- Cialek, C. A. et al. (Apr. 30, 2021). "Imaging translational control by Argonaute with single-molecule resolution in live cells". In: p. 2021.04.30.442135. DOI: 10.1101/2021.04.30.442135.
- Cifuentes, D. et al. (June 25, 2010). "A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity". In: *Science* 328.5986, pp. 1694–1698. ISSN: 1095-9203. DOI: 10.1126/science.1190809.
- Claycomb, J. M. (Aug. 4, 2014). "Ancient Endo-siRNA Pathways Reveal New Tricks". In: *Current Biology* 24.15, R703–R715. ISSN: 0960-9822. DOI: 10.1016/j.cub.2014.06.009.
- Cock, J. M. et al. (June 3, 2010). "The Ectocarpus genome and the independent evolution of multicellularity in brown algae". In: *Nature* 465.7298, pp. 617–621. ISSN: 1476-4687. DOI: 10.1038/nature09016.
- Cole, K. A. et al. (May 2008). "A functional screen identifies miR-34a as a candidate neuroblastoma tumor suppressor gene". In: *Mol Cancer Res* 6.5, pp. 735–742. ISSN: 1541-7786. DOI: 10.1158/1541-7786.MCR-07-2102.
- Concepcion, C. P. et al. (2012). "Intact p53-dependent responses in miR-34-deficient mice". In: *PLoS Genet* 8.7, e1002797. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1002797.
- Concordet, JP and M. Haeussler (July 2, 2018). "CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens". In: *Nucleic Acids Res* 46 (W1), W242–W245. ISSN: 1362-4962. DOI: 10.1093/nar/gky354.
- Corney, D. C. et al. (Sept. 15, 2007). "MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth". In: *Cancer Res* 67.18, pp. 8433–8438. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-07-1585.
- Czech, B. et al. (June 5, 2008). "An endogenous small interfering RNA pathway in Drosophila". In: *Nature* 453.7196, pp. 798–802. ISSN: 1476-4687. DOI: 10.1038/nature07007.
- D'Ambrogio, A. et al. (Dec. 27, 2012). "Specific miRNA stabilization by Gld2-catalyzed monoadenylation". In: *Cell Rep* 2.6, pp. 1537–1545. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2012.10.023.
- Davis, E. et al. (Apr. 26, 2005). "RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus". In: *Curr Biol* 15.8, pp. 743–749. ISSN: 0960-9822. DOI: 10.1016/j.cub.2005.02.060.
- de la Mata, M. et al. (Apr. 2015). "Potent degradation of neuronal miRNAs induced by highly complementary targets". In: *EMBO Rep* 16.4, pp. 500–511. ISSN: 1469-3178. DOI: 10.15252/embr.201540078.
- Denli, A. M. et al. (Nov. 11, 2004). "Processing of primary microRNAs by the Microprocessor complex". In: *Nature* 432.7014, pp. 231–235. ISSN: 1476-4687. DOI: 10.1038/nature03049.

- Denzler, R. et al. (June 5, 2014). "Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance". In: *Mol Cell* 54.5, pp. 766–776. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.03.045.
- Dexheimer, P. J., Jingkui Wang, and Luisa Cochella (Dec. 21, 2020). "Two MicroRNAs Are Sufficient for Embryonic Patterning in *C. elegans*". In: *Curr Biol* 30.24, 5058–5065.e5. ISSN: 1879-0445. DOI: 10.1016/j.cub.2020.09.066.
- Diebel, K. W., Anna L. Smith, and Linda F. van Dyk (Jan. 2010). "Mature and functional viral miRNAs transcribed from novel RNA polymerase III promoters". In: *RNA* 16.1, pp. 170–185. ISSN: 1469-9001. DOI: 10.1261/rna.1873910.
- Ding, L. et al. (Aug. 19, 2005). "The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in *C. elegans*". In: *Mol Cell* 19.4, pp. 437–447. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2005.07.013.
- Djuranovic, S., Ali Nahvi, and Rachel Green (Apr. 13, 2012). "miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay". In: *Science* 336.6078, pp. 237–240. ISSN: 1095-9203. DOI: 10.1126/science.1215691.
- Dou, D. and R. Joseph (June 3, 1996). "Cloning of human neuronatin gene and its localization to chromosome-20q 11.2-12: the deduced protein is a novel "proteolipid"". In: *Brain Res* 723.1, pp. 8–22. ISSN: 0006-8993. DOI: 10.1016/0006-8993(96)00167-9.
- Drake, M. et al. (Dec. 8, 2014). "A requirement for ERK-dependent Dicer phosphorylation in coordinating oocyte-to-embryo transition in *C. elegans*". In: *Dev Cell* 31.5, pp. 614–628. ISSN: 1878-1551. DOI: 10.1016/j.devcel.2014.11.004.
- Duchaine, T. F. and M. R. Fabian (Mar. 1, 2019). "Mechanistic Insights into MicroRNA-Mediated Gene Silencing". In: *Cold Spring Harb Perspect Biol* 11.3, a032771. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a032771.
- Dye, M. J., N. Gromak, and N. J. Proudfoot (Mar. 17, 2006). "Exon tethering in transcription by RNA polymerase II". In: *Mol Cell* 21.6, pp. 849–859. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2006.01.032.
- Ebert, M. S. and Phillip A. Sharp (Apr. 27, 2012). "Roles for microRNAs in conferring robustness to biological processes". In: *Cell* 149.3, pp. 515–524. ISSN: 1097-4172. DOI: 10.1016/j.cell.2012.04.005.
- Ecsedi, M., Magdalene Rausch, and Helge Großhans (Feb. 9, 2015). "The let-7 microRNA directs vulval development through a single target". In: *Dev Cell* 32.3, pp. 335–344. ISSN: 1878-1551. DOI: 10.1016/j.devcel.2014.12.018.
- Eichhorn, S. W. et al. (Oct. 2, 2014). "mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues". In: *Mol Cell* 56.1, pp. 104–115. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.08.028.
- Elbashir, S. M., W. Lendeckel, and T. Tuschl (Jan. 15, 2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs". In: *Genes Dev* 15.2, pp. 188–200. ISSN: 0890-9369.
- Elkayam, E. et al. (July 6, 2012). "The structure of human argonaute-2 in complex with miR-20a". In: *Cell* 150.1, pp. 100–110. ISSN: 1097-4172. DOI: 10.1016/j.cell.2012.05.017.
- Enchev, R. I., Brenda A. Schulman, and Matthias Peter (Jan. 2015). "Protein neddylation: beyond cullin-RING ligases". In: *Nat Rev Mol Cell Biol* 16.1, pp. 30–44. ISSN: 1471-0080. DOI: 10.1038/nrm3919.
- Fabian, M. R. et al. (Sept. 24, 2009). "Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation". In: *Mol Cell* 35.6, pp. 868–880. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2009.08.004.
- Fang, W. and D. P. Bartel (Oct. 1, 2015). "The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes". In: *Mol Cell* 60.1, pp. 131–145. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2015.08.015.

- Fang, W. and D. P. Bartel (Apr. 16, 2020). "MicroRNA Clustering Assists Processing of Suboptimal MicroRNA Hairpins through the Action of the ERH Protein". In: *Mol Cell* 78.2, 289–302.e6. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2020.01.026.
- Farh, K. KH et al. (Dec. 16, 2005). "The widespread impact of mammalian MicroRNAs on mRNA repression and evolution". In: *Science* 310.5755, pp. 1817–1821. ISSN: 1095-9203. DOI: 10.1126/science.1121158.
- Ferracin, M. et al. (Sept. 2011). "MicroRNA profiling for the identification of cancers with unknown primary tissue-of-origin". In: *J Pathol* 225.1, pp. 43–53. ISSN: 1096-9896. DOI: 10.1002/path.2915.
- Fire, A. et al. (Feb. 19, 1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*". In: *Nature* 391.6669, pp. 806–811. ISSN: 0028-0836. DOI: 10.1038/35888.
- Fletcher, C. E. et al. (Mar. 17, 2017). "A novel role for GSK3 $\beta$  as a modulator of Drosha microprocessor activity and MicroRNA biogenesis". In: *Nucleic Acids Res* 45.5, pp. 2809–2828. ISSN: 1362-4962. DOI: 10.1093/nar/gkw938.
- Forman, J. J., Aster Legesse-Miller, and Hilary A. Collier (Sept. 30, 2008). "A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence". In: *Proc Natl Acad Sci U S A* 105.39, pp. 14879–14884. ISSN: 1091-6490. DOI: 10.1073/pnas.0803230105.
- Friedman, R. C. et al. (Jan. 2009). "Most mammalian mRNAs are conserved targets of microRNAs". In: *Genome Res* 19.1, pp. 92–105. ISSN: 1088-9051. DOI: 10.1101/gr.082701.108.
- Fromm, B. et al. (Jan. 8, 2020). "MirGeneDB 2.0: the metazoan microRNA complement". In: *Nucleic Acids Res* 48 (D1), pp. D132–D141. ISSN: 1362-4962. DOI: 10.1093/nar/gkz885.
- Fukaya, T. and Y. Tomari (Nov. 25, 2011). "PABP is not essential for microRNA-mediated translational repression and deadenylation in vitro". In: *EMBO J* 30.24, pp. 4998–5009. ISSN: 1460-2075. DOI: 10.1038/emboj.2011.426.
- Förstemann, K. et al. (July 2005). "Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein". In: *PLoS Biol* 3.7, e236. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0030236.
- Förstemann, Klaus et al. (July 27, 2007). "Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1". In: *Cell* 130.2, pp. 287–297. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.05.056.
- Garcia, D. M. et al. (Sept. 11, 2011). "Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs". In: *Nat Struct Mol Biol* 18.10, pp. 1139–1146. ISSN: 1545-9985. DOI: 10.1038/nsmb.2115.
- Ghildiyal, M. and P. D. Zamore (Feb. 2009). "Small silencing RNAs: an expanding universe". In: *Nat Rev Genet* 10.2, pp. 94–108. ISSN: 1471-0064. DOI: 10.1038/nrg2504.
- Ghildiyal, M. et al. (May 23, 2008). "Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells". In: *Science* 320.5879, pp. 1077–1081. ISSN: 1095-9203. DOI: 10.1126/science.1157396.
- Ghini, F. et al. (Aug. 7, 2018). "Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation". In: *Nat Commun* 9.1, p. 3119. ISSN: 2041-1723. DOI: 10.1038/s41467-018-05182-9.
- Gilbert, L. A. et al. (July 18, 2013). "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes". In: *Cell* 154.2, pp. 442–451. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.06.044.
- Gingras, A. C., B. Raught, and N. Sonenberg (1999). "eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation". In: *Annu Rev Biochem* 68, pp. 913–963. ISSN: 0066-4154. DOI: 10.1146/annurev.biochem.68.1.913.

- Giraldez, A. J. et al. (May 6, 2005). "MicroRNAs regulate brain morphogenesis in zebrafish". In: *Science* 308.5723, pp. 833–838. ISSN: 1095-9203. DOI: 10.1126/science.1109020.
- Giraldez, A. J. et al. (Apr. 7, 2006). "Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs". In: *Science* 312.5770, pp. 75–79. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1122689.
- Goldspink, D. A. et al. (Sept. 1, 2013). "The microtubule end-binding protein EB2 is a central regulator of microtubule reorganisation in apico-basal epithelial differentiation". In: *J Cell Sci* 126 (Pt 17), pp. 4000–4014. ISSN: 1477-9137. DOI: 10.1242/jcs.129759.
- Gregory, R. I. et al. (Nov. 11, 2004). "The Microprocessor complex mediates the genesis of microRNAs". In: *Nature* 432.7014, pp. 235–240. ISSN: 1476-4687. DOI: 10.1038/nature03120.
- Griffiths-Jones, S. et al. (Jan. 2008). "miRBase: tools for microRNA genomics". In: *Nucleic Acids Res* 36 (Database issue), pp. D154–158. ISSN: 1362-4962. DOI: 10.1093/nar/gkm952.
- Grimson, A. et al. (July 6, 2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing". In: *Mol Cell* 27.1, pp. 91–105. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2007.06.017.
- Gripp, K. W. et al. (Sept. 2016). "A novel rasopathy caused by recurrent de novo missense mutations in PPP1CB closely resembles Noonan syndrome with loose anagen hair". In: *Am J Med Genet A* 170.9, pp. 2237–2247. ISSN: 1552-4833. DOI: 10.1002/ajmg.a.37781.
- Grishok, A. et al. (July 13, 2001). "Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing". In: *Cell* 106.1, pp. 23–34. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(01)00431-7.
- Groner, A. C. et al. (Mar. 5, 2010). "KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading". In: *PLoS Genet* 6.3, e1000869. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1000869.
- Grosswendt, S. et al. (June 19, 2014). "Unambiguous identification of miRNA:target site interactions by different types of ligation reactions". In: *Mol Cell* 54.6, pp. 1042–1054. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.03.049.
- Gu, S. et al. (Nov. 9, 2012). "The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo". In: *Cell* 151.4, pp. 900–911. ISSN: 1097-4172. DOI: 10.1016/j.cell.2012.09.042.
- Guo, H. et al. (Aug. 12, 2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels". In: *Nature* 466.7308, pp. 835–840. ISSN: 1476-4687. DOI: 10.1038/nature09267.
- Guo, Y. et al. (Feb. 27, 2015). "Characterization of the mammalian miRNA turnover landscape". In: *Nucleic Acids Res* 43.4, pp. 2326–2341. ISSN: 1362-4962. DOI: 10.1093/nar/gkv057.
- Guo, YJ et al. (Mar. 2020). "ERK/MAPK signalling pathway and tumorigenesis". In: *Exp Ther Med* 19.3, pp. 1997–2007. ISSN: 1792-0981. DOI: 10.3892/etm.2020.8454.
- Gutiérrez-Vázquez, C. et al. (June 2017). "3' Uridylation controls mature microRNA turnover during CD4 T-cell activation". In: *RNA* 23.6, pp. 882–891. ISSN: 1469-9001. DOI: 10.1261/rna.060095.116.
- Haase, A. D. et al. (Oct. 2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing". In: *EMBO Rep* 6.10, pp. 961–967. ISSN: 1469-221X. DOI: 10.1038/sj.embor.7400509.
- Hafner, M. et al. (Apr. 2, 2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP". In: *Cell* 141.1, pp. 129–141. ISSN: 1097-4172. DOI: 10.1016/j.cell.2010.03.009.
- Hagan, J. P., Elena Piskounova, and Richard I. Gregory (Oct. 2009). "Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells". In: *Nat Struct Mol Biol* 16.10, pp. 1021–1025. ISSN: 1545-9985. DOI: 10.1038/nsmb.1676.

- Hamilton, A. J. and D. C. Baulcombe (Oct. 29, 1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants". In: *Science* 286.5441, pp. 950–952. ISSN: 0036-8075. DOI: 10.1126/science.286.5441.950.
- Hammond, S. M. et al. (Mar. 16, 2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells". In: *Nature* 404.6775, pp. 293–296. ISSN: 0028-0836. DOI: 10.1038/35005107.
- Hammond, S. M. et al. (Aug. 10, 2001). "Argonaute2, a link between genetic and biochemical analyses of RNAi". In: *Science* 293.5532, pp. 1146–1150. ISSN: 0036-8075. DOI: 10.1126/science.1064023.
- Han, B. W. et al. (Nov. 22, 2011). "The 3'-to-5' exonuclease Nibbler shapes the 3' ends of microRNAs bound to *Drosophila* Argonaute1". In: *Curr Biol* 21.22, pp. 1878–1887. ISSN: 1879-0445. DOI: 10.1016/j.cub.2011.09.034.
- Han, J. et al. (Dec. 15, 2004). "The Drosha-DGCR8 complex in primary microRNA processing". In: *Genes Dev* 18.24, pp. 3016–3027. ISSN: 0890-9369. DOI: 10.1101/gad.1262504.
- Han, J. et al. (June 2, 2006). "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex". In: *Cell* 125.5, pp. 887–901. ISSN: 0092-8674. DOI: 10.1016/j.cell.2006.03.043.
- Han, J. et al. (Jan. 9, 2009). "Posttranscriptional crossregulation between Drosha and DGCR8". In: *Cell* 136.1, pp. 75–84. ISSN: 1097-4172. DOI: 10.1016/j.cell.2008.10.053.
- Han, J. et al. (Dec. 18, 2020). "A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming". In: *Science* 370.6523, eabc9546. ISSN: 1095-9203. DOI: 10.1126/science.abc9546.
- Hanahan, D. and R. A. Weinberg (Mar. 4, 2011). "Hallmarks of cancer: the next generation". In: *Cell* 144.5, pp. 646–674. ISSN: 1097-4172. DOI: 10.1016/j.cell.2011.02.013.
- Hansen, T. B. et al. (Sept. 30, 2011). "miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA". In: *EMBO J* 30.21, pp. 4414–4422. ISSN: 1460-2075. DOI: 10.1038/emboj.2011.359.
- Hansen, Thomas B. et al. (Mar. 21, 2013). "Natural RNA circles function as efficient microRNA sponges". In: *Nature* 495.7441, pp. 384–388. ISSN: 1476-4687. DOI: 10.1038/nature11993.
- Hatley, M E. et al. (Sept. 14, 2010). "Modulation of K-Ras-dependent lung tumorigenesis by MicroRNA-21". In: *Cancer Cell* 18.3, pp. 282–293. ISSN: 1878-3686. DOI: 10.1016/j.ccr.2010.08.013.
- He, L. et al. (June 9, 2005). "A microRNA polycistron as a potential human oncogene". In: *Nature* 435.7043, pp. 828–833. ISSN: 1476-4687. DOI: 10.1038/nature03552.
- He, L. et al. (June 28, 2007). "A microRNA component of the p53 tumour suppressor network". In: *Nature* 447.7148, pp. 1130–1134. ISSN: 1476-4687. DOI: 10.1038/nature05939.
- Helwak, A. et al. (Apr. 25, 2013). "Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding". In: *Cell* 153.3, pp. 654–665. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.03.043.
- Hendrickson, D. G. et al. (Nov. 2009). "Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA". In: *PLoS Biol* 7.11, e1000238. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1000238.
- Heo, I. et al. (Oct. 24, 2008). "Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA". In: *Mol Cell* 32.2, pp. 276–284. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2008.09.014.
- Heo, I. et al. (Aug. 21, 2009). "TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation". In: *Cell* 138.4, pp. 696–708. ISSN: 1097-4172. DOI: 10.1016/j.cell.2009.08.002.

- Heo, I. et al. (Oct. 26, 2012). "Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs". In: *Cell* 151.3, pp. 521–532. ISSN: 1097-4172. DOI: 10.1016/j.cell.2012.09.022.
- Herbert, K. M. et al. (Nov. 27, 2013). "Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile". In: *Cell Rep* 5.4, pp. 1070–1081. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2013.10.017.
- Hindson, C. M. et al. (Oct. 2013). "Absolute quantification by droplet digital PCR versus analog real-time PCR". In: *Nat Methods* 10.10, pp. 1003–1005. ISSN: 1548-7105. DOI: 10.1038/nmeth.2633.
- Hojo, H. et al. (Jan. 10, 2020). "The RNA-binding protein QKI-7 recruits the poly(A) polymerase GLD-2 for 3' adenylation and selective stabilization of microRNA-122". In: *J Biol Chem* 295.2, pp. 390–402. ISSN: 1083-351X. DOI: 10.1074/jbc.RA119.011617.
- Hong, D. S. et al. (May 2020). "Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours". In: *Br J Cancer* 122.11, pp. 1630–1637. ISSN: 1532-1827. DOI: 10.1038/s41416-020-0802-1.
- Horwich, M. D. et al. (July 17, 2007). "The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC". In: *Curr Biol* 17.14, pp. 1265–1272. ISSN: 0960-9822. DOI: 10.1016/j.cub.2007.06.030.
- Hu, S. et al. (Dec. 1, 1995). "The Drosophila abrupt gene encodes a BTB-zinc finger regulatory protein that controls the specificity of neuromuscular connections". In: *Genes Dev* 9.23, pp. 2936–2948. ISSN: 0890-9369. DOI: 10.1101/gad.9.23.2936.
- Huang, N. et al. (Oct. 14, 2010). "Characterising and predicting haploinsufficiency in the human genome". In: *PLoS Genet* 6.10, e1001154. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1001154.
- Huang, Tai-Chung et al. (Apr. 6, 2012). "Regulation of lipid metabolism by Dicer revealed through SILAC mice". In: *J Proteome Res* 11.4, pp. 2193–2205. ISSN: 1535-3907. DOI: 10.1021/pr2009884.
- Huntzinger, E. et al. (Dec. 15, 2010). "Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing". In: *EMBO J* 29.24, pp. 4146–4160. ISSN: 1460-2075. DOI: 10.1038/emboj.2010.274.
- Hutter, K. et al. (June 4, 2020). "SAFB2 Enables the Processing of Suboptimal Stem-Loop Structures in Clustered Primary miRNA Transcripts". In: *Mol Cell* 78.5, 876–889.e6. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2020.05.011.
- Hutvagner, G. and Phillip D. Zamore (Sept. 20, 2002). "A microRNA in a multiple-turnover RNAi enzyme complex". In: *Science* 297.5589, pp. 2056–2060. ISSN: 1095-9203. DOI: 10.1126/science.1073827.
- Hutvagner, G. et al. (Aug. 3, 2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA". In: *Science* 293.5531, pp. 834–838. ISSN: 0036-8075. DOI: 10.1126/science.1062961.
- Hutvagner, G. et al. (Apr. 2004). "Sequence-specific inhibition of small RNA function". In: *PLoS Biol* 2.4, E98. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0020098.
- Hébert, S. S. et al. (Oct. 15, 2010). "Genetic ablation of Dicer in adult forebrain neurons results in abnormal tau hyperphosphorylation and neurodegeneration". In: *Hum Mol Genet* 19.20, pp. 3959–3969. ISSN: 1460-2083. DOI: 10.1093/hmg/ddq311.
- Iimori, M. et al. (Mar. 31, 2016). "Phosphorylation of EB2 by Aurora B and CDK1 ensures mitotic progression and genome stability". In: *Nat Commun* 7, p. 11117. ISSN: 2041-1723. DOI: 10.1038/ncomms11117.
- Iwasaki, S. et al. (July 30, 2010). "Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes". In: *Mol Cell* 39.2, pp. 292–299. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2010.05.015.

- Janas, M. M. et al. (Oct. 2011). "Feed-forward microprocessing and splicing activities at a microRNA-containing intron". In: *PLoS Genet* 7.10, e1002330. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1002330.
- Jin, H. Y. et al. (2015). "Transfection of microRNA Mimics Should Be Used with Caution". In: *Front Genet* 6, p. 340. ISSN: 1664-8021. DOI: 10.3389/fgene.2015.00340.
- Jonas, S. and E. Izaurralde (Dec. 15, 2013). "The role of disordered protein regions in the assembly of decapping complexes and RNP granules". In: *Genes Dev* 27.24, pp. 2628–2641. ISSN: 1549-5477. DOI: 10.1101/gad.227843.113.
- Jonas, S. and E. Izaurralde (July 2015). "Towards a molecular understanding of microRNA-mediated gene silencing". In: *Nat Rev Genet* 16.7, pp. 421–433. ISSN: 1471-0064. DOI: 10.1038/nrg3965.
- Jones-Rhoades, M. W., David P. Bartel, and Bonnie Bartel (2006). "MicroRNAs and their regulatory roles in plants". In: *Annu Rev Plant Biol* 57, pp. 19–53. ISSN: 1543-5008. DOI: 10.1146/annurev.arplant.57.032905.105218.
- Joseph, R., D. Dou, and W. Tsang (June 30, 1994). "Molecular cloning of a novel mRNA (neuronatin) that is highly expressed in neonatal mammalian brain". In: *Biochem Biophys Res Commun* 201.3, pp. 1227–1234. ISSN: 0006-291X. DOI: 10.1006/bbrc.1994.1836.
- Juzenas, S. et al. (Sept. 19, 2017). "A comprehensive, cell specific microRNA catalogue of human peripheral blood". In: *Nucleic Acids Research* 45.16, p. 9290. DOI: 10.1093/nar/gkx706.
- Kagitani, F. et al. (Sept. 1, 1997). "Peg5/Neuronatin is an imprinted gene located on sub-distal chromosome 2 in the mouse". In: *Nucleic Acids Res* 25.17, pp. 3428–3432. ISSN: 0305-1048. DOI: 10.1093/nar/25.17.3428.
- Kahvejian, A. et al. (Jan. 1, 2005). "Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms". In: *Genes Dev* 19.1, pp. 104–113. ISSN: 0890-9369. DOI: 10.1101/gad.1262905.
- Kamenska, A. et al. (Mar. 2014). "Human 4E-T represses translation of bound mRNAs and enhances microRNA-mediated silencing". In: *Nucleic Acids Res* 42.5, pp. 3298–3313. ISSN: 1362-4962. DOI: 10.1093/nar/gkt1265.
- Kampmann, M. (Feb. 16, 2018). "CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine". In: *ACS Chem Biol* 13.2, pp. 406–416. ISSN: 1554-8929. DOI: 10.1021/acscchembio.7b00657. (Visited on 09/15/2021).
- Kanellopoulou, C. et al. (Feb. 15, 2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing". In: *Genes Dev* 19.4, pp. 489–501. ISSN: 0890-9369. DOI: 10.1101/gad.1248505.
- Karres, J. S. et al. (Oct. 5, 2007). "The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in Drosophila". In: *Cell* 131.1, pp. 136–145. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.09.020.
- Kasinski, A. L. and F. J. Slack (Nov. 1, 2012). "miRNA-34 prevents cancer initiation and progression in a therapeutically resistant K-ras and p53-induced mouse model of lung adenocarcinoma". In: *Cancer Res* 72.21, pp. 5576–5587. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-12-2001.
- Kasinski, A. L. et al. (July 2015). "A combinatorial microRNA therapeutics approach to suppressing non-small cell lung cancer". In: *Oncogene* 34.27, pp. 3547–3555. ISSN: 1476-5594. DOI: 10.1038/onc.2014.282.
- Kataoka, N., M. Fujita, and M. Ohno (June 2009). "Functional association of the Microprocessor complex with the spliceosome". In: *Mol Cell Biol* 29.12, pp. 3243–3254. ISSN: 1098-5549. DOI: 10.1128/MCB.00360-09.
- Kataoka, Y., M. Takeichi, and T. Uemura (Apr. 2001). "Developmental roles and molecular characterization of a Drosophila homologue of Arabidopsis Argonaute1, the founder of a



- novel gene superfamily". In: *Genes Cells* 6.4, pp. 313–325. ISSN: 1356-9597. DOI: 10.1046/j.1365-2443.2001.00427.x.
- Katoh, T., Hiroaki Hojo, and Tsutomu Suzuki (Sept. 3, 2015). "Destabilization of microRNAs in human cells by 3' deadenylation mediated by PARN and CUGBP1". In: *Nucleic Acids Res* 43.15, pp. 7521–7534. ISSN: 1362-4962. DOI: 10.1093/nar/gkv669.
- Katoh, T. et al. (Feb. 15, 2009). "Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2". In: *Genes Dev* 23.4, pp. 433–438. ISSN: 1549-5477. DOI: 10.1101/gad.1761509.
- Kawahara, Y. et al. (Feb. 23, 2007). "Redirection of silencing targets by adenosine-to-inosine editing of miRNAs". In: *Science* 315.5815, pp. 1137–1140. ISSN: 1095-9203. DOI: 10.1126/science.1138050.
- Kawahara, Y. et al. (Sept. 2008). "Frequency and fate of microRNA editing in human brain". In: *Nucleic Acids Res* 36.16, pp. 5270–5280. ISSN: 1362-4962. DOI: 10.1093/nar/gkn479.
- Kawamata, T., H. Seitz, and Y. Tomari (Sept. 2009). "Structural determinants of miRNAs for RISC loading and slicer-independent unwinding". In: *Nat Struct Mol Biol* 16.9, pp. 953–960. ISSN: 1545-9985. DOI: 10.1038/nsmb.1630.
- Kawamata, T., M. Yoda, and Y. Tomari (Sept. 1, 2011). "Multilayer checkpoints for microRNA authenticity during RISC assembly". In: *EMBO Rep* 12.9, pp. 944–949. ISSN: 1469-3178. DOI: 10.1038/embor.2011.128.
- Kawamura, Y. et al. (June 5, 2008). "Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells". In: *Nature* 453.7196, pp. 793–797. ISSN: 1476-4687. DOI: 10.1038/nature06938.
- Kerwin, J., I. Khan, and Reproducibility Project: Cancer Biology (Apr. 21, 2020). "Replication Study: A coding-independent function of gene and pseudogene mRNAs regulates tumour biology". In: *Elife* 9, e51019. ISSN: 2050-084X. DOI: 10.7554/eLife.51019.
- Ketting, R. F. et al. (Oct. 15, 2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*". In: *Genes Dev* 15.20, pp. 2654–2659. ISSN: 0890-9369. DOI: 10.1101/gad.927801.
- Kim, B., Kyowon Jeong, and V. Narry Kim (Apr. 20, 2017). "Genome-wide Mapping of DROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates". In: *Mol Cell* 66.2, 258–269.e5. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2017.03.013.
- Kim, B. et al. (July 2, 2015). "TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms". In: *EMBO J* 34.13, pp. 1801–1815. ISSN: 1460-2075. DOI: 10.15252/embj.201590931.
- Kim, K. et al. (July 2018). "SRSF3 recruits DROSHA to the basal junction of primary microRNAs". In: *RNA* 24.7, pp. 892–898. ISSN: 1469-9001. DOI: 10.1261/rna.065862.118.
- Kim, S. et al. (Aug. 20, 2021). "The regulatory impact of RNA-binding proteins on microRNA targeting". In: *Nat Commun* 12.1, p. 5057. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25078-5.
- Kim, Y. et al. (Nov. 6, 2014). "Deletion of human tarbp2 reveals cellular microRNA targets and cell-cycle function of TRBP". In: *Cell Rep* 9.3, pp. 1061–1074. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2014.09.039.
- Kim, YK, Boseon Kim, and V. Narry Kim (Mar. 29, 2016). "Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis". In: *Proc Natl Acad Sci U S A* 113.13, E1881–1889. ISSN: 1091-6490. DOI: 10.1073/pnas.1602532113.
- Kim, YK and V. N. Kim (Feb. 7, 2007). "Processing of intronic microRNAs". In: *EMBO J* 26.3, pp. 775–783. ISSN: 0261-4189. DOI: 10.1038/sj.emboj.7601512.
- Kingston, E. R. and David P. Bartel (Nov. 2019). "Global analyses of the dynamics of mammalian microRNA metabolism". In: *Genome Res* 29.11, pp. 1777–1790. ISSN: 1549-5469. DOI: 10.1101/gr.251421.119.

- Kleaveland, B. et al. (July 12, 2018). "A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain". In: *Cell* 174.2, 350–362.e17. ISSN: 1097-4172. DOI: 10.1016/j.cell.2018.05.022.
- Knight, S. W. and B. L. Bass (Sept. 21, 2001). "A Role for the RNase III Enzyme DCR-1 in RNA Interference and Germ Line Development in *Caenorhabditis elegans*". In: *Science* 293.5538, pp. 2269–2271. ISSN: 0036-8075. DOI: 10.1126/science.1062039.
- Kobayashi, H. and R. H. Singer (Apr. 30, 2021). "Single-molecule imaging of microRNA-mediated gene silencing in cells". In: p. 2021.04.30.442050. DOI: 10.1101/2021.04.30.442050.
- Kobayashi, H. and Y. Tomari (Jan. 2016). "RISC assembly: Coordination between small RNAs and Argonaute proteins". In: *Biochim Biophys Acta* 1859.1, pp. 71–81. ISSN: 0006-3002. DOI: 10.1016/j.bbagr.2015.08.007.
- Kobayashi, T. et al. (July 2015). "Early postnatal ablation of the microRNA-processing enzyme, Drosha, causes chondrocyte death and impairs the structural integrity of the articular cartilage". In: *Osteoarthritis Cartilage* 23.7, pp. 1214–1220. ISSN: 1522-9653. DOI: 10.1016/j.joca.2015.02.015.
- Krek, A. et al. (May 2005). "Combinatorial microRNA target predictions". In: *Nat Genet* 37.5, pp. 495–500. ISSN: 1061-4036. DOI: 10.1038/ng1536.
- Krol, J., Inga Loedige, and Witold Filipowicz (Sept. 2010). "The widespread regulation of microRNA biogenesis, function and decay". In: *Nat Rev Genet* 11.9, pp. 597–610. ISSN: 1471-0064. DOI: 10.1038/nrg2843.
- Krützfeldt, J. et al. (Dec. 1, 2005). "Silencing of microRNAs in vivo with 'antagomirs'". In: *Nature* 438.7068, pp. 685–689. ISSN: 1476-4687. DOI: 10.1038/nature04303.
- Kuzuoğlu-Öztürk, D. et al. (June 1, 2016). "miRISC and the CCR4-NOT complex silence mRNA targets independently of 43S ribosomal scanning". In: *EMBO J* 35.11, pp. 1186–1203. ISSN: 1460-2075. DOI: 10.15252/embj.201592901.
- Kwak, P. B. and Y. Tomari (Jan. 10, 2012). "The N domain of Argonaute drives duplex unwinding during RISC assembly". In: *Nat Struct Mol Biol* 19.2, pp. 145–151. ISSN: 1545-9985. DOI: 10.1038/nsmb.2232.
- Kwon, S. C. et al. (Feb. 7, 2019). "Molecular Basis for the Single-Nucleotide Precision of Primary microRNA Processing". In: *Mol Cell* 73.3, 505–518.e5. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2018.11.005.
- Kwon, S. C. et al. (Nov. 4, 2020). "ERH facilitates microRNA maturation through the interaction with the N-terminus of DGCR8". In: *Nucleic Acids Res* 48.19, pp. 11097–11112. ISSN: 1362-4962. DOI: 10.1093/nar/gkaa827.
- La Rocca, G. et al. (Aug. 31, 2021). "Inducible and reversible inhibition of miRNA-mediated gene repression in vivo". In: *eLife* 10, e70948. ISSN: 2050-084X. DOI: 10.7554/eLife.70948.
- Lagos-Quintana, M. et al. (Oct. 26, 2001). "Identification of novel genes coding for small expressed RNAs". In: *Science* 294.5543, pp. 853–858. ISSN: 0036-8075. DOI: 10.1126/science.1064921.
- Lai, E. C. (Apr. 2002). "Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation". In: *Nat Genet* 30.4, pp. 363–364. ISSN: 1061-4036. DOI: 10.1038/ng865.
- Landgraf, P. et al. (June 29, 2007). "A mammalian microRNA expression atlas based on small RNA library sequencing". In: *Cell* 129.7, pp. 1401–1414. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.04.040.
- Landthaler, M., A Yalcin, and T Tuschl (Dec. 14, 2004). "The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis". In: *Curr Biol* 14.23, pp. 2162–2167. ISSN: 0960-9822. DOI: 10.1016/j.cub.2004.11.001.
- Lasda, E. L. and T. Blumenthal (June 2011). "Trans-splicing". In: *Wiley Interdiscip Rev RNA* 2.3, pp. 417–434. ISSN: 1757-7012. DOI: 10.1002/wrna.71.

- Lau, N. C. et al. (Oct. 26, 2001). "An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*". In: *Science* 294.5543, pp. 858–862. ISSN: 0036-8075. DOI: 10.1126/science.1065062.
- Lazzaretti, D., Isabelle Tournier, and Elisa Izaurrealde (June 2009). "The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins". In: *RNA* 15.6, pp. 1059–1066. ISSN: 1469-9001. DOI: 10.1261/rna.1606309.
- Lebedeva, S. et al. (Aug. 5, 2011). "Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR". In: *Mol Cell* 43.3, pp. 340–352. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2011.06.008.
- Lee, E. J. et al. (Jan. 2008). "Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors". In: *RNA* 14.1, pp. 35–42. ISSN: 1469-9001. DOI: 10.1261/rna.804508.
- Lee, H Y. et al. (July 2013a). "Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing". In: *Nucleic Acids Res* 41.13, pp. 6568–6576. ISSN: 1362-4962. DOI: 10.1093/nar/gkt361.
- Lee, HC et al. (June 25, 2010). "Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi". In: *Mol Cell* 38.6, pp. 803–814. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2010.04.005. (Visited on 10/11/2021).
- Lee, JR et al. (Apr. 7, 2004). "An intramolecular interaction between the FHA domain and a coiled coil negatively regulates the kinesin motor KIF1A". In: *EMBO J* 23.7, pp. 1506–1515. ISSN: 0261-4189. DOI: 10.1038/sj.emboj.7600164.
- Lee, M. et al. (Dec. 4, 2014). "Adenylation of maternally inherited microRNAs by Wispy". In: *Mol Cell* 56.5, pp. 696–707. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.10.011.
- Lee, R. C. and V. Ambros (Oct. 26, 2001). "An extensive class of small RNAs in *Caenorhabditis elegans*". In: *Science* 294.5543, pp. 862–864. ISSN: 0036-8075. DOI: 10.1126/science.1065329.
- Lee, R. C., R. L. Feinbaum, and V. Ambros (Dec. 3, 1993). "The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*". In: *Cell* 75.5, pp. 843–854. ISSN: 0092-8674. DOI: 10.1016/0092-8674(93)90529-y.
- Lee, S. et al. (June 12, 2013b). "Selective degradation of host MicroRNAs by an intergenic HCMV noncoding RNA accelerates virus production". In: *Cell Host Microbe* 13.6, pp. 678–690. ISSN: 1934-6069. DOI: 10.1016/j.chom.2013.05.007.
- Lee, Y. et al. (Sept. 2, 2002). "MicroRNA maturation: stepwise processing and subcellular localization". In: *EMBO J* 21.17, pp. 4663–4670. ISSN: 0261-4189. DOI: 10.1093/emboj/cdf476.
- Lee, Y. et al. (Sept. 25, 2003). "The nuclear RNase III Drosha initiates microRNA processing". In: *Nature* 425.6956, pp. 415–419. ISSN: 1476-4687. DOI: 10.1038/nature01957.
- Lee, Y. et al. (Feb. 8, 2006). "The role of PACT in the RNA silencing pathway". In: *EMBO J* 25.3, pp. 522–532. ISSN: 0261-4189. DOI: 10.1038/sj.emboj.7600942.
- Lehrbach, N. J. et al. (Dec. 2012). "Post-developmental microRNA expression is required for normal physiology, and regulates aging in parallel to insulin/IGF-1 signaling in *C. elegans*". In: *RNA* 18.12, pp. 2220–2235. ISSN: 1469-9001. DOI: 10.1261/rna.035402.112.
- Levine, A. J. (Aug. 2020). "p53: 800 million years of evolution and 40 years of discovery". In: *Nat Rev Cancer* 20.8, pp. 471–480. ISSN: 1474-1768. DOI: 10.1038/s41568-020-0262-1.
- Lewis, B. P., Christopher B. Burge, and David P. Bartel (Jan. 14, 2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets". In: *Cell* 120.1, pp. 15–20. ISSN: 0092-8674. DOI: 10.1016/j.cell.2004.12.035.
- Lewis, B. P. et al. (Dec. 26, 2003). "Prediction of mammalian microRNA targets". In: *Cell* 115.7, pp. 787–798. ISSN: 0092-8674. DOI: 10.1016/s0092-8674(03)01018-3.

- Li, L. et al. (May 2013). "MiR-34a inhibits proliferation and migration of breast cancer through down-regulation of Bcl-2 and SIRT1". In: *Clin Exp Med* 13.2, pp. 109–117. ISSN: 1591-9528. DOI: 10.1007/s10238-012-0186-5.
- Li, X. et al. (Apr. 17, 2009). "A MicroRNA Imparts Robustness against Environmental Fluctuation during Development". In: *Cell* 137.2, pp. 273–282. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2009.01.058.
- Libri, V. et al. (Jan. 3, 2012). "Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target". In: *Proc Natl Acad Sci U S A* 109.1, pp. 279–284. ISSN: 1091-6490. DOI: 10.1073/pnas.1114204109.
- Lim, L. P. et al. (Apr. 15, 2003). "The microRNAs of *Caenorhabditis elegans*". In: *Genes Dev* 17.8, pp. 991–1008. ISSN: 0890-9369. DOI: 10.1101/gad.1074403.
- Lim, L. P. et al. (Feb. 17, 2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs". In: *Nature* 433.7027, pp. 769–773. ISSN: 1476-4687. DOI: 10.1038/nature03315.
- Lin, H-H et al. (Nov. 2010). "Neuronatin promotes neural lineage in ESCs via Ca(2+) signaling". In: *Stem Cells* 28.11, pp. 1950–1960. ISSN: 1549-4918. DOI: 10.1002/stem.530.
- Liu, J. et al. (Sept. 3, 2004). "Argonaute2 is the catalytic engine of mammalian RNAi". In: *Science* 305.5689, pp. 1437–1441. ISSN: 1095-9203. DOI: 10.1126/science.1102513.
- Liu, N. et al. (Nov. 22, 2011). "The exoribonuclease Nibbler controls 3' end processing of microRNAs in *Drosophila*". In: *Curr Biol* 21.22, pp. 1888–1893. ISSN: 1879-0445. DOI: 10.1016/j.cub.2011.10.006.
- Lu, J. et al. (June 9, 2005). "MicroRNA expression profiles classify human cancers". In: *Nature* 435.7043, pp. 834–838. ISSN: 1476-4687. DOI: 10.1038/nature03702.
- Luciano, D. J. et al. (Aug. 2004). "RNA editing of a miRNA precursor". In: *RNA* 10.8, pp. 1174–1177. ISSN: 1355-8382. DOI: 10.1261/rna.7350304.
- Lund, E. et al. (Jan. 2, 2004). "Nuclear export of microRNA precursors". In: *Science* 303.5654, pp. 95–98. ISSN: 1095-9203. DOI: 10.1126/science.1090599.
- Ma, JB et al. (Mar. 31, 2005). "Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein". In: *Nature* 434.7033, pp. 666–670. ISSN: 1476-4687. DOI: 10.1038/nature03514.
- Ma, X. et al. (June 21, 2011). "Loss of the miR-21 allele elevates the expression of its target genes and reduces tumorigenesis". In: *Proc Natl Acad Sci U S A* 108.25, pp. 10144–10149. ISSN: 0027-8424. DOI: 10.1073/pnas.1103735108. (Visited on 10/08/2021).
- Macrae, I. J. et al. (Jan. 13, 2006). "Structural basis for double-stranded RNA processing by Dicer". In: *Science* 311.5758, pp. 195–198. ISSN: 1095-9203. DOI: 10.1126/science.1121638.
- Marcet, B. et al. (June 2011). "Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway". In: *Nat Cell Biol* 13.6, pp. 693–699. ISSN: 1476-4679. DOI: 10.1038/ncb2241.
- Marcinowski, L. et al. (Feb. 2012). "Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo". In: *PLoS Pathog* 8.2, e1002510. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002510.
- Martin, R. et al. (Feb. 2009). "A *Drosophila pasha* mutant distinguishes the canonical microRNA and mirtron pathways". In: *Mol Cell Biol* 29.3, pp. 861–870. ISSN: 1098-5549. DOI: 10.1128/MCB.01524-08.
- Marzi, M. J. et al. (Apr. 2016). "Degradation dynamics of microRNAs revealed by a novel pulse-chase approach". In: *Genome Res* 26.4, pp. 554–565. ISSN: 1549-5469. DOI: 10.1101/gr.198788.115.

- Mathys, H. et al. (June 5, 2014). "Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression". In: *Mol Cell* 54.5, pp. 751–765. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.03.036.
- Mauxion, F. et al. (Dec. 2009). "BTG/TOB factors impact deadenylases". In: *Trends Biochem Sci* 34.12, pp. 640–647. ISSN: 0968-0004. DOI: 10.1016/j.tibs.2009.07.008.
- Meister, G. et al. (July 23, 2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs". In: *Mol Cell* 15.2, pp. 185–197. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2004.07.007.
- Memczak, S. et al. (Mar. 21, 2013). "Circular RNAs are a large class of animal RNAs with regulatory potency". In: *Nature* 495.7441, pp. 333–338. ISSN: 1476-4687. DOI: 10.1038/nature11928.
- Mishima, Y. et al. (Jan. 24, 2012). "Translational inhibition by deadenylation-independent mechanisms is central to microRNA-mediated silencing in zebrafish". In: *Proc Natl Acad Sci U S A* 109.4, pp. 1104–1109. ISSN: 1091-6490. DOI: 10.1073/pnas.1113350109.
- Miska, E. A. et al. (Dec. 2007). "Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability". In: *PLoS Genet* 3.12, e215. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.0030215.
- Mockly, S. and H. Seitz (2019). "Inconsistencies and Limitations of Current MicroRNA Target Identification Methods". In: *Methods Mol Biol* 1970, pp. 291–314. ISSN: 1940-6029. DOI: 10.1007/978-1-4939-9207-2\_16.
- Molnár, A. et al. (June 2007). "miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*". In: *Nature* 447.7148, pp. 1126–1129. ISSN: 1476-4687. DOI: 10.1038/nature05903. (Visited on 10/11/2021).
- Monteys, A. M. et al. (Mar. 2010). "Structure and activity of putative intronic miRNA promoters". In: *RNA* 16.3, pp. 495–505. ISSN: 1469-9001. DOI: 10.1261/rna.1731910.
- Moran, Y. et al. (Feb. 21, 2017). "The evolutionary origin of plant and animal microRNAs". In: *Nat Ecol Evol* 1.3, p. 27. ISSN: 2397-334X. DOI: 10.1038/s41559-016-0027.
- Moretti, F. et al. (May 27, 2012). "PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding". In: *Nat Struct Mol Biol* 19.6, pp. 603–608. ISSN: 1545-9985. DOI: 10.1038/nsmb.2309.
- Morin, R. D. et al. (Apr. 2008). "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells". In: *Genome Res* 18.4, pp. 610–621. ISSN: 1088-9051. DOI: 10.1101/gr.7179508.
- Morita, S. et al. (June 2007). "One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation". In: *Genomics* 89.6, pp. 687–696. ISSN: 0888-7543. DOI: 10.1016/j.ygeno.2007.01.004.
- Morlando, M. et al. (Sept. 2008). "Primary microRNA transcripts are processed co-transcriptionally". In: *Nat Struct Mol Biol* 15.9, pp. 902–909. ISSN: 1545-9993. DOI: 10.1038/nsmb.1475.
- Moss, E. G. and J. Romer-Seibert (Oct. 2014). "Cell-intrinsic timing in animal development". In: *Wiley Interdiscip Rev Dev Biol* 3.5, pp. 365–377. ISSN: 1759-7692. DOI: 10.1002/wdev.145.
- Mosse, Y. P. et al. (Oct. 2007). "Neuroblastomas have distinct genomic DNA profiles that predict clinical phenotype and regional gene expression". In: *Genes Chromosomes Cancer* 46.10, pp. 936–949. ISSN: 1045-2257. DOI: 10.1002/gcc.20477.
- Mourelatos, Z. et al. (Mar. 15, 2002). "miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs". In: *Genes Dev* 16.6, pp. 720–728. ISSN: 0890-9369. DOI: 10.1101/gad.974702.
- Mukherjee, N. et al. (Aug. 5, 2011). "Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability". In: *Mol Cell* 43.3, pp. 327–339. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2011.06.007.

- Mukherji, S. et al. (Aug. 21, 2011). "MicroRNAs can generate thresholds in target gene expression". In: *Nat Genet* 43.9, pp. 854–859. ISSN: 1546-1718. DOI: 10.1038/ng.905.
- Muller, H., Matteo Jacopo Marzi, and Francesco Nicassio (Sept. 29, 2014). "IsomiRage: From Functional Classification to Differential Expression of miRNA Isoforms". In: *Front Bioeng Biotechnol* 2, p. 38. ISSN: 2296-4185. DOI: 10.3389/fbioe.2014.00038. (Visited on 10/08/2021).
- Mullokandov, G. et al. (July 1, 2012). "High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries". In: *Nat Methods* 9.8, pp. 840–846. ISSN: 1548-7105. DOI: 10.1038/nmeth.2078.
- Nakanishi, K. (2016). "Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins?" In: *Wiley Interdiscip Rev RNA* 7.5, pp. 637–660. ISSN: 1757-7004. DOI: 10.1002/wrna.1356.
- Nakanishi, K. et al. (June 20, 2012). "Structure of yeast Argonaute with guide RNA". In: *Nature* 486.7403, pp. 368–374. ISSN: 1476-4687. DOI: 10.1038/nature11211.
- Navarro, F. and J. Lieberman (2015). "miR-34 and p53: New Insights into a Complex Functional Relationship". In: *PLoS One* 10.7, e0132767. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0132767.
- Newman, M. A., Vidya Mani, and Scott M. Hammond (Jan. 10, 2011). "Deep sequencing of microRNA precursors reveals extensive 3' end modification". In: *RNA* 17.10, pp. 1795–1803. ISSN: 1355-8382, 1469-9001. DOI: 10.1261/rna.2713611.
- Nguyen, T. A. et al. (June 4, 2015). "Functional Anatomy of the Human Microprocessor". In: *Cell* 161.6, pp. 1374–1387. ISSN: 1097-4172. DOI: 10.1016/j.cell.2015.05.010.
- Nishikura, K. (Feb. 2016). "A-to-I editing of coding and non-coding RNAs by ADARs". In: *Nat Rev Mol Cell Biol* 17.2, pp. 83–96. ISSN: 1471-0080. DOI: 10.1038/nrm.2015.4.
- Nishimura, T. et al. (June 9, 2015). "The eIF4E-Binding Protein 4E-T Is a Component of the mRNA Decay Machinery that Bridges the 5' and 3' Termini of Target mRNAs". In: *Cell Rep* 11.9, pp. 1425–1436. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2015.04.065.
- O'Donnell, K. A. et al. (June 9, 2005). "c-Myc-regulated microRNAs modulate E2F1 expression". In: *Nature* 435.7043, pp. 839–843. ISSN: 1476-4687. DOI: 10.1038/nature03677.
- Okada, C. et al. (Nov. 27, 2009). "A high-resolution structure of the pre-microRNA nuclear export machinery". In: *Science* 326.5957, pp. 1275–1279. ISSN: 1095-9203. DOI: 10.1126/science.1178705.
- Okada, N. et al. (Mar. 1, 2014). "A positive feedback between p53 and miR-34 miRNAs mediates tumor suppression". In: *Genes Dev* 28.5, pp. 438–450. ISSN: 1549-5477. DOI: 10.1101/gad.233585.113.
- Okamura, K. et al. (July 13, 2007). "The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in *Drosophila*". In: *Cell* 130.1, pp. 89–100. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.06.028. (Visited on 06/29/2021).
- Ota, A. et al. (May 1, 2004). "Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma". In: *Cancer Res* 64.9, pp. 3087–3095. ISSN: 0008-5472. DOI: 10.1158/0008-5472.can-03-3773.
- Oyang, E. L. et al. (2011). "Functional characterization of the dendritically localized mRNA neuronatin in hippocampal neurons". In: *PLoS One* 6.9, e24879. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0024879.
- Ozata, D. M. et al. (Feb. 2019). "PIWI-interacting RNAs: small RNAs with big functions". In: *Nat Rev Genet* 20.2, pp. 89–108. ISSN: 1471-0064. DOI: 10.1038/s41576-018-0073-3.
- Ozgur, S. et al. (Oct. 27, 2015). "Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex". In: *Cell Rep* 13.4, pp. 703–711. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2015.09.033.

- Ozsolak, F. et al. (Nov. 15, 2008). "Chromatin structure analyses identify miRNA promoters". In: *Genes Dev* 22.22, pp. 3172–3183. ISSN: 0890-9369. DOI: 10.1101/gad.1706508.
- Padgett, R. A. et al. (Aug. 31, 1984). "Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors". In: *Science* 225.4665, pp. 898–903. ISSN: 0036-8075. DOI: 10.1126/science.6206566.
- Panebianco, F. et al. (July 2019). "Delivery of biologically active miR-34a in normal and cancer mammary epithelial cells by synthetic nanoparticles". In: *Nanomedicine* 19, pp. 95–105. ISSN: 1549-9642. DOI: 10.1016/j.nano.2019.03.013.
- Park, C. Yon et al. (Apr. 19, 2012). "A resource for the conditional ablation of microRNAs in the mouse". In: *Cell Rep* 1.4, pp. 385–391. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2012.02.008.
- Park, J. H., Sang-Yoon Shin, and Chanseok Shin (Feb. 28, 2017). "Non-canonical targets destabilize microRNAs in human Argonautes". In: *Nucleic Acids Res* 45.4, pp. 1569–1583. ISSN: 1362-4962. DOI: 10.1093/nar/gkx029.
- Park, JE et al. (July 13, 2011). "Dicer recognizes the 5' end of RNA for efficient and accurate processing". In: *Nature* 475.7355, pp. 201–205. ISSN: 1476-4687. DOI: 10.1038/nature10198.
- Parker, J. S., S. M. Roe, and D. Barford (Mar. 31, 2005). "Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex". In: *Nature* 434.7033, pp. 663–666. ISSN: 1476-4687. DOI: 10.1038/nature03462.
- Paroo, Z. et al. (Oct. 2, 2009). "Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling". In: *Cell* 139.1, pp. 112–122. ISSN: 1097-4172. DOI: 10.1016/j.cell.2009.06.044.
- Parrish, S. et al. (Nov. 2000). "Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference". In: *Mol Cell* 6.5, pp. 1077–1087. ISSN: 1097-2765. DOI: 10.1016/s1097-2765(00)00106-4.
- Pasquinelli, A. E. et al. (Nov. 2, 2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA". In: *Nature* 408.6808, pp. 86–89. ISSN: 0028-0836. DOI: 10.1038/35040556.
- Patrick, D. M. et al. (Aug. 1, 2010). "Defective erythroid differentiation in miR-451 mutant mice mediated by 14-3-3zeta". In: *Genes Dev* 24.15, pp. 1614–1619. ISSN: 1549-5477. DOI: 10.1101/gad.1942810.
- Pedersen, J. S. et al. (Apr. 2006). "Identification and classification of conserved RNA secondary structures in the human genome". In: *PLoS Comput Biol* 2.4, e33. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.0020033.
- Pfeffer, S. et al. (Apr. 30, 2004). "Identification of virus-encoded microRNAs". In: *Science* 304.5671, pp. 734–736. ISSN: 1095-9203. DOI: 10.1126/science.1096781.
- Pfeffer, S. R., Chuan He Yang, and Lawrence M. Pfeffer (Sept. 2015). "The Role of miR-21 in Cancer". In: *Drug Dev Res* 76.6, pp. 270–277. ISSN: 1098-2299. DOI: 10.1002/ddr.21257.
- Pinzón, N. et al. (Feb. 2017). "microRNA target prediction programs predict many false positives". In: *Genome Res* 27.2, pp. 234–245. ISSN: 1549-5469. DOI: 10.1101/gr.205146.116.
- Piwecka, M. et al. (Sept. 22, 2017). "Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function". In: *Science* 357.6357, eaam8526. ISSN: 1095-9203. DOI: 10.1126/science.aam8526.
- Poliseno, L. et al. (June 24, 2010). "A coding-independent function of gene and pseudogene mRNAs regulates tumour biology". In: *Nature* 465.7301, pp. 1033–1038. ISSN: 1476-4687. DOI: 10.1038/nature09144.
- Posadas, D. M. and R. W. Carthew (Aug. 2014). "MicroRNAs and their roles in developmental canalization". In: *Curr Opin Genet Dev* 27, pp. 1–6. ISSN: 1879-0380. DOI: 10.1016/j.gde.2014.03.005.

- Pramanik, D. et al. (Aug. 2011). "Restitution of tumor suppressor microRNAs using a systemic nanovector inhibits pancreatic cancer growth in mice". In: *Mol Cancer Ther* 10.8, pp. 1470–1480. ISSN: 1538-8514. DOI: 10.1158/1535-7163.MCT-11-0152.
- Péllisson, A. et al. (Feb. 2007). "A Novel Repeat-Associated Small Interfering RNA-Mediated Silencing Pathway Downregulates Complementary Sense gypsy Transcripts in Somatic Cells of the *Drosophila* Ovary". In: *J Virol* 81.4, pp. 1951–1960. ISSN: 0022-538X. DOI: 10.1128/JVI.01980-06. (Visited on 09/06/2021).
- Qi, L. S. et al. (Feb. 28, 2013). "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression". In: *Cell* 152.5, pp. 1173–1183. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.02.022.
- Raj, A. and A. van Oudenaarden (Oct. 17, 2008). "Nature, nurture, or chance: stochastic gene expression and its consequences". In: *Cell* 135.2, pp. 216–226. ISSN: 1097-4172. DOI: 10.1016/j.cell.2008.09.050.
- Rajasethupathy, P. et al. (Sept. 24, 2009). "Characterization of small RNAs in *Aplysia* reveals a role for miR-124 in constraining synaptic plasticity through CREB". In: *Neuron* 63.6, pp. 803–817. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2009.05.029.
- Rasmussen, K. D. et al. (July 5, 2010). "The miR-144/451 locus is required for erythroid homeostasis". In: *J Exp Med* 207.7, pp. 1351–1358. ISSN: 1540-9538. DOI: 10.1084/jem.20100458.
- Raver-Shapira, N. et al. (June 8, 2007). "Transcriptional activation of miR-34a contributes to p53-mediated apoptosis". In: *Mol Cell* 26.5, pp. 731–743. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2007.05.017.
- Reichholf, B. et al. (Aug. 22, 2019). "Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis". In: *Mol Cell* 75.4, 756–768.e7. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2019.06.018.
- Reimão-Pinto, M. M. et al. (July 16, 2015). "Uridylation of RNA Hairpins by Tailor Confines the Emergence of MicroRNAs in *Drosophila*". In: *Mol Cell* 59.2, pp. 203–216. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2015.05.033.
- Reimão-Pinto, M. M. et al. (Nov. 15, 2016). "Molecular basis for cytoplasmic RNA surveillance by uridylation-triggered decay in *Drosophila*". In: *EMBO J* 35.22, pp. 2417–2434. ISSN: 1460-2075. DOI: 10.15252/embj.201695164.
- Reinhart, B. J. et al. (Feb. 24, 2000). "The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*". In: *Nature* 403.6772, pp. 901–906. ISSN: 0028-0836. DOI: 10.1038/35002607.
- Rissland, O. S., Sue-Jean Hong, and David P. Bartel (Sept. 16, 2011). "MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes". In: *Mol Cell* 43.6, pp. 993–1004. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2011.08.021.
- Rodriguez, A. et al. (Apr. 27, 2007). "Requirement of bic/microRNA-155 for normal immune function". In: *Science* 316.5824, pp. 608–611. ISSN: 1095-9203. DOI: 10.1126/science.1139253.
- Romano, N. and G. Macino (Nov. 1992). "Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences". In: *Mol Microbiol* 6.22, pp. 3343–3353. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.1992.tb02202.x.
- Rosenfeld, N. et al. (Apr. 2008). "MicroRNAs accurately identify cancer tissue origin". In: *Nat Biotechnol* 26.4, pp. 462–469. ISSN: 1546-1696. DOI: 10.1038/nbt1392.
- Rouya, C. et al. (Sept. 2014). "Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1". In: *RNA* 20.9, pp. 1398–1409. ISSN: 1469-9001. DOI: 10.1261/rna.045302.114.
- Ruby, J. G., C. H. Jan, and D. P. Bartel (July 5, 2007). "Intronic microRNA precursors that bypass Drosha processing". In: *Nature* 448.7149, pp. 83–86. ISSN: 0028-0836. DOI: 10.1038/nature05983.



- Rupaimoole, R. and F. J. Slack (Mar. 2017). "MicroRNA therapeutics: towards a new era for the management of cancer and other diseases". In: *Nat Rev Drug Discov* 16.3, pp. 203–222. ISSN: 1474-1784. DOI: 10.1038/nrd.2016.246.
- Ruskin, B. and M. R. Green (July 12, 1985). "An RNA processing activity that debranches RNA lariats". In: *Science* 229.4709, pp. 135–140. ISSN: 0036-8075. DOI: 10.1126/science.2990042.
- Saetrom, P. et al. (2007). "Distance constraints between microRNA target sites dictate efficacy and cooperativity". In: *Nucleic Acids Res* 35.7, pp. 2333–2342. ISSN: 1362-4962. DOI: 10.1093/nar/gkm133.
- Saini, H. K., Sam Griffiths-Jones, and Anton James Enright (Nov. 6, 2007). "Genomic analysis of human microRNA transcripts". In: *Proc Natl Acad Sci U S A* 104.45, pp. 17719–17724. ISSN: 0027-8424. DOI: 10.1073/pnas.0703890104.
- Saito, Ku. et al. (July 2005). "Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells". In: *PLoS Biol* 3.7, e235. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0030235.
- Salomon, W. E. et al. (July 2, 2015). "Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides". In: *Cell* 162.1, pp. 84–95. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.06.029. (Visited on 08/13/2021).
- Schirle, N. T. and I. J. MacRae (May 25, 2012). "The crystal structure of human Argonaute2". In: *Science* 336.6084, pp. 1037–1040. ISSN: 1095-9203. DOI: 10.1126/science.1221551.
- Schirle, N. T., Jessica Sheu-Gruttadauria, and Ian J. MacRae (Oct. 31, 2014). "Structural basis for microRNA targeting". In: *Science* 346.6209, pp. 608–613. ISSN: 1095-9203. DOI: 10.1126/science.1258040.
- Schirle, N. T. et al. (Sept. 11, 2015). "Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets". In: *Elife* 4. ISSN: 2050-084X. DOI: 10.7554/eLife.07646.
- Schmiedel, J. M. et al. (Apr. 3, 2015). "Gene expression. MicroRNA control of protein expression noise". In: *Science* 348.6230, pp. 128–132. ISSN: 1095-9203. DOI: 10.1126/science.aaa1738.
- Schopp, I. M. et al. (June 6, 2017). "Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes". In: *Nat Commun* 8, p. 15690. ISSN: 2041-1723. DOI: 10.1038/ncomms15690.
- Schwanhäusser, B. et al. (May 19, 2011). "Global quantification of mammalian gene expression control". In: *Nature* 473.7347, pp. 337–342. ISSN: 1476-4687. DOI: 10.1038/nature10098.
- Schwarz, D. S. et al. (Sept. 8, 2006). "Designing siRNA that distinguish between genes that differ by a single nucleotide". In: *PLoS Genet* 2.9, e140. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.0020140.
- Seitz, H. et al. (Sept. 2004). "A Large Imprinted microRNA Gene Cluster at the Mouse Dlk1-Gtl2 Domain". In: *Genome Res* 14.9, pp. 1741–1748. ISSN: 1088-9051. DOI: 10.1101/gr.2743304.
- Selbach, M. et al. (Sept. 4, 2008). "Widespread changes in protein synthesis induced by microRNAs". In: *Nature* 455.7209, pp. 58–63. ISSN: 1476-4687. DOI: 10.1038/nature07228.
- Sheu-Gruttadauria, J. et al. (Sept. 19, 2019). "Structural Basis for Target-Directed MicroRNA Degradation". In: *Mol Cell* 75.6, 1243–1255.e7. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2019.06.019.
- Shi, C. Y. et al. (Dec. 18, 2020). "The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation". In: *Science* 370.6523, eabc9359. ISSN: 1095-9203. DOI: 10.1126/science.abc9359.
- Shin, C. et al. (June 25, 2010). "Expanding the microRNA targeting code: functional sites with centered pairing". In: *Mol Cell* 38.6, pp. 789–802. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2010.06.005.

- Shkumatava, A. et al. (Feb. 15, 2009). "Coherent but overlapping expression of microRNAs and their targets during vertebrate development". In: *Genes Dev* 23.4, pp. 466–481. ISSN: 1549-5477. DOI: 10.1101/gad.1745709.
- Siciliano, V. et al. (2013). "MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise". In: *Nat Commun* 4, p. 2364. ISSN: 2041-1723. DOI: 10.1038/ncomms3364.
- Siddiqui, N. et al. (Aug. 24, 2007). "Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein". In: *J Biol Chem* 282.34, pp. 25067–25075. ISSN: 0021-9258. DOI: 10.1074/jbc.M701256200.
- Silva, J. M. et al. (Nov. 2005). "Second-generation shRNA libraries covering the mouse and human genomes". In: *Nat Genet* 37.11, pp. 1281–1288. ISSN: 1546-1718. DOI: 10.1038/ng1650.
- Slack, F. J. et al. (Apr. 2000). "The lin-41 RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor". In: *Mol Cell* 5.4, pp. 659–669. ISSN: 1097-2765. DOI: 10.1016/s1097-2765(00)80245-2.
- Soleimani, S. et al. (July 29, 2020). "Small regulatory noncoding RNAs in *Drosophila melanogaster*: biogenesis and biological functions". In: *Briefings in Functional Genomics* 19.4, pp. 309–323. ISSN: 2041-2657. DOI: 10.1093/bfpg/ela005.
- Song, JJ et al. (Sept. 3, 2004). "Crystal structure of Argonaute and its implications for RISC slicer activity". In: *Science* 305.5689, pp. 1434–1437. ISSN: 1095-9203. DOI: 10.1126/science.1102514.
- Song, R. et al. (June 5, 2014). "miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110". In: *Nature* 510.7503, pp. 115–120. ISSN: 1476-4687. DOI: 10.1038/nature13413.
- Sood, P. et al. (Feb. 21, 2006). "Cell-type-specific signatures of microRNAs on target mRNA expression". In: *Proc Natl Acad Sci U S A* 103.8, pp. 2746–2751. ISSN: 0027-8424. DOI: 10.1073/pnas.0511045103.
- Stark, A. et al. (Dec. 2003). "Identification of *Drosophila* MicroRNA targets". In: *PLoS Biol* 1.3, E60. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0000060.
- Su, L. K. and Y. Qi (Jan. 15, 2001). "Characterization of human MAPRE genes and their proteins". In: *Genomics* 71.2, pp. 142–149. ISSN: 0888-7543. DOI: 10.1006/geno.2000.6428.
- Subkhankulova, T., Michael J. Gilchrist, and Frederick J. Livesey (June 3, 2008). "Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells". In: *BMC Genomics* 9, p. 268. ISSN: 1471-2164. DOI: 10.1186/1471-2164-9-268.
- Sullivan, C. S. et al. (June 2, 2005). "SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells". In: *Nature* 435.7042, pp. 682–686. ISSN: 1476-4687. DOI: 10.1038/nature03576.
- Sulston, J. E. and H. R. Horvitz (Feb. 1, 1981). "Abnormal cell lineages in mutants of the nematode *Caenorhabditis elegans*". In: *Developmental Biology* 82.1, pp. 41–55. ISSN: 0012-1606. DOI: 10.1016/0012-1606(81)90427-9.
- Takimoto, K., Motoaki Wakiyama, and Shigeyuki Yokoyama (June 2009). "Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression". In: *RNA* 15.6, pp. 1078–1089. ISSN: 1469-9001. DOI: 10.1261/rna.1363109.
- Tam, O. H. et al. (May 22, 2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes". In: *Nature* 453.7194, pp. 534–538. ISSN: 1476-4687. DOI: 10.1038/nature06904.
- Tan, G. C. et al. (Aug. 2014). "5' isomiR variation is of functional and evolutionary importance". In: *Nucleic Acids Res* 42.14, pp. 9424–9435. ISSN: 1362-4962. DOI: 10.1093/nar/gku656.

- Tang, X. et al. (Oct. 2010). "Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization". In: *Nucleic Acids Res* 38.19, pp. 6610–6619. ISSN: 1362-4962. DOI: 10.1093/nar/gkq547.
- Tang, X. et al. (2011). "Glycogen synthase kinase 3 beta (GSK3 $\beta$ ) phosphorylates the RNAase III enzyme Drosha at S300 and S302". In: *PLoS One* 6.6, e20391. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0020391.
- Tang, X. et al. (2013). "Acetylation of drosha on the N-terminus inhibits its degradation by ubiquitination". In: *PLoS One* 8.8, e72503. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0072503.
- Tarasov, V. et al. (July 1, 2007). "Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest". In: *Cell Cycle* 6.13, pp. 1586–1593. ISSN: 1551-4005. DOI: 10.4161/cc.6.13.4436.
- Tarver, J. E. et al. (July 27, 2015). "microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*". In: *Nucleic Acids Research* 43.13, pp. 6384–6398. ISSN: 0305-1048. DOI: 10.1093/nar/gkv578.
- Taylor, G. A. et al. (Oct. 2008). "Behavioral characterization of P311 knockout mice". In: *Genes Brain Behav* 7.7, pp. 786–795. ISSN: 1601-183X. DOI: 10.1111/j.1601-183X.2008.00420.x.
- Tazawa, H. et al. (Sept. 25, 2007). "Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells". In: *Proc Natl Acad Sci U S A* 104.39, pp. 15472–15477. ISSN: 0027-8424. DOI: 10.1073/pnas.0707351104.
- Thornton, J. E. et al. (Oct. 2012). "Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7)". In: *RNA* 18.10, pp. 1875–1885. ISSN: 1469-9001. DOI: 10.1261/rna.034538.112.
- Tian, Y. et al. (Feb. 20, 2014). "A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer". In: *Mol Cell* 53.4, pp. 606–616. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.01.003.
- Till, S. et al. (Oct. 2007). "A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain". In: *Nat Struct Mol Biol* 14.10, pp. 897–903. ISSN: 1545-9993. DOI: 10.1038/nsmb1302.
- Tokumaru, S. et al. (Nov. 2008). "let-7 regulates Dicer expression and constitutes a negative feedback loop". In: *Carcinogenesis* 29.11, pp. 2073–2077. ISSN: 1460-2180. DOI: 10.1093/carcin/bgn187.
- Tomari, Y., T. Du, and P. D. Zamore (July 27, 2007). "Sorting of *Drosophila* small silencing RNAs". In: *Cell* 130.2, pp. 299–308. ISSN: 0092-8674. DOI: 10.1016/j.cell.2007.05.057.
- Trang, P. et al. (June 2011). "Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice". In: *Mol Ther* 19.6, pp. 1116–1122. ISSN: 1525-0024. DOI: 10.1038/mt.2011.48.
- Treiber, T. et al. (Apr. 20, 2017). "A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis". In: *Mol Cell* 66.2, 270–284.e13. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2017.03.014.
- Triboulet, R. et al. (June 2009). "Post-transcriptional control of DGCR8 expression by the Microprocessor". In: *RNA* 15.6, pp. 1005–1011. ISSN: 1469-9001. DOI: 10.1261/rna.1591709.
- Tsutsumi, A. et al. (Sept. 18, 2011). "Recognition of the pre-miRNA structure by *Drosophila* Dicer-1". In: *Nat Struct Mol Biol* 18.10, pp. 1153–1158. ISSN: 1545-9985. DOI: 10.1038/nsmb.2125.
- Tuschl, T. et al. (Dec. 15, 1999). "Targeted mRNA degradation by double-stranded RNA in vitro". In: *Genes Dev* 13.24, pp. 3191–3197. ISSN: 0890-9369. DOI: 10.1101/gad.13.24.3191.
- Ulitsky, I. et al. (Dec. 23, 2011). "Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution". In: *Cell* 147.7, pp. 1537–1550. ISSN: 1097-4172. DOI: 10.1016/j.cell.2011.11.055.

- Ustianenko, D. et al. (Dec. 2013). "Mammalian DIS3L2 exoribonuclease targets the uridylylated precursors of let-7 miRNAs". In: *RNA* 19.12, pp. 1632–1638. ISSN: 1469-9001. DOI: 10.1261/rna.040055.113.
- Vasquez-Rifo, A. et al. (2012). "Developmental characterization of the microRNA-specific *C. elegans* Argonautes alg-1 and alg-2". In: *PLoS One* 7.3, e33750. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0033750.
- Ventura, A. et al. (Mar. 7, 2008). "Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters". In: *Cell* 132.5, pp. 875–886. ISSN: 1097-4172. DOI: 10.1016/j.cell.2008.02.019.
- Vesely, C. et al. (Oct. 29, 2014). "ADAR2 induces reproducible changes in sequence and abundance of mature microRNAs in the mouse brain". In: *Nucleic Acids Res* 42.19, pp. 12155–12168. ISSN: 1362-4962. DOI: 10.1093/nar/gku844.
- Viviani, B. (Oct. 2006). "Preparation and coculture of neurons and glial cells". In: *Curr Protoc Cell Biol* Chapter 2, Unit 2.7. ISSN: 1934-2616. DOI: 10.1002/0471143030.cb0207s32.
- Volinia, S. et al. (Feb. 14, 2006). "A microRNA expression signature of human solid tumors defines cancer gene targets". In: *Proc Natl Acad Sci U S A* 103.7, pp. 2257–2261. ISSN: 0027-8424. DOI: 10.1073/pnas.0510565103.
- Wada, T., Jiro Kikuchi, and Yusuke Furukawa (Feb. 1, 2012). "Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8". In: *EMBO Rep* 13.2, pp. 142–149. ISSN: 1469-3178. DOI: 10.1038/embor.2011.247.
- Waghray, S. et al. (July 2015). "Xenopus CAF1 requires NOT1-mediated interaction with 4E-T to repress translation in vivo". In: *RNA* 21.7, pp. 1335–1345. ISSN: 1469-9001. DOI: 10.1261/rna.051565.115.
- Wang, L. et al. (Jan. 1, 2015a). "The analysis of microRNA-34 family expression in human cancer studies comparing cancer tissues with corresponding pericarcinoma tissues". In: *Gene* 554.1, pp. 1–8. ISSN: 1879-0038. DOI: 10.1016/j.gene.2014.10.032.
- Wang, X. et al. (Mar. 2015b). "Tumor suppressor miR-34a targets PD-L1 and functions as a potential immunotherapeutic target in acute myeloid leukemia". In: *Cell Signal* 27.3, pp. 443–452. ISSN: 1873-3913. DOI: 10.1016/j.cellsig.2014.12.003.
- Wang, Y. et al. (Mar. 2007). "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal". In: *Nat Genet* 39.3, pp. 380–385. ISSN: 1061-4036. DOI: 10.1038/ng1969.
- Wang, Y. et al. (Dec. 18, 2008). "Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex". In: *Nature* 456.7224, pp. 921–926. ISSN: 1476-4687. DOI: 10.1038/nature07666.
- Wang, Y. et al. (Oct. 8, 2009). "Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes". In: *Nature* 461.7265, pp. 754–761. ISSN: 1476-4687. DOI: 10.1038/nature08434.
- Warner, M. J. et al. (Nov. 16, 2016). "S6K2-mediated regulation of TRBP as a determinant of miRNA expression in human primary lymphatic endothelial cells". In: *Nucleic Acids Res* 44.20, pp. 9942–9955. ISSN: 1362-4962. DOI: 10.1093/nar/gkw631.
- Watanabe, T. et al. (May 22, 2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes". In: *Nature* 453.7194, pp. 539–543. ISSN: 1476-4687. DOI: 10.1038/nature06908.
- Wee, L. M. et al. (Nov. 21, 2012). "Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties". In: *Cell* 151.5, pp. 1055–1067. ISSN: 1097-4172. DOI: 10.1016/j.cell.2012.10.036.
- Weinberg, R. A. (Nov. 22, 1991). "Tumor suppressor genes". In: *Science* 254.5035, pp. 1138–1146. ISSN: 0036-8075. DOI: 10.1126/science.1659741.

- Welch, C., Y. Chen, and R. L. Stallings (July 26, 2007). "MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells". In: *Oncogene* 26.34, pp. 5017–5022. ISSN: 0950-9232. DOI: 10.1038/sj.onc.1210293.
- Wen, J. et al. (Sept. 2015). "Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates". In: *PLoS Comput Biol* 11.9, e1004441. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004441.
- Westholm, J. O. et al. (Feb. 2012). "Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs". In: *RNA* 18.2, pp. 177–192. ISSN: 1469-9001. DOI: 10.1261/rna.030627.111.
- Wienholds, E. et al. (Nov. 2003). "The microRNA-producing enzyme Dicer1 is essential for zebrafish development". In: *Nat Genet* 35.3, pp. 217–218. ISSN: 1061-4036. DOI: 10.1038/ng1251.
- Wiggins, J. F. et al. (July 15, 2010). "Development of a lung cancer therapeutic based on the tumor suppressor microRNA-34". In: *Cancer Res* 70.14, pp. 5923–5930. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-10-0655.
- Wightman, B., I. Ha, and G. Ruvkun (Dec. 3, 1993). "Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*". In: *Cell* 75.5, pp. 855–862. ISSN: 0092-8674. DOI: 10.1016/0092-8674(93)90530-4.
- Wightman, B. et al. (Oct. 1991). "Negative regulatory sequences in the *lin-14* 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development". In: *Genes Dev* 5.10, pp. 1813–1824. ISSN: 0890-9369. DOI: 10.1101/gad.5.10.1813.
- Wilson, R. C. et al. (Feb. 5, 2015). "Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis". In: *Mol Cell* 57.3, pp. 397–407. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2014.11.030.
- Xiao, C. et al. (Oct. 5, 2007). "MiR-150 controls B cell differentiation by targeting the transcription factor *c-Myb*". In: *Cell* 131.1, pp. 146–159. ISSN: 1097-4172. DOI: 10.1016/j.cell.2007.07.021.
- Xie, J. et al. (Mar. 4, 2012). "Long-term, efficient inhibition of microRNA function in mice using rAAV vectors". In: *Nat Methods* 9.4, pp. 403–409. ISSN: 1548-7091. DOI: 10.1038/nmeth.1903. (Visited on 09/06/2021).
- Xie, M. et al. (Dec. 19, 2013). "Mammalian 5'-capped microRNA precursors that generate a single microRNA". In: *Cell* 155.7, pp. 1568–1580. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.11.027.
- Yamakuchi, M., Marcella Ferlito, and Charles J. Lowenstein (Sept. 9, 2008). "miR-34a repression of SIRT1 regulates apoptosis". In: *Proc Natl Acad Sci U S A* 105.36, pp. 13421–13426. ISSN: 1091-6490. DOI: 10.1073/pnas.0801613105.
- Yamamura, S. et al. (2012). "MicroRNA-34a modulates c-Myc transcriptional complexes to suppress malignancy in human prostate cancer cells". In: *PLoS One* 7.1, e29722. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0029722.
- Yang, D., H. Lu, and J. W. Erickson (Oct. 5, 2000). "Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in *Drosophila* embryos". In: *Curr Biol* 10.19, pp. 1191–1200. ISSN: 0960-9822. DOI: 10.1016/S0960-9822(00)00732-6.
- Yang, JS et al. (Aug. 24, 2010). "Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis". In: *Proc Natl Acad Sci U S A* 107.34, pp. 15163–15168. ISSN: 1091-6490. DOI: 10.1073/pnas.1006432107.
- Yang, Q. et al. (Feb. 19, 2015). "Stress induces p38 MAPK-mediated phosphorylation and inhibition of Drosha-dependent cell survival". In: *Mol Cell* 57.4, pp. 721–734. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2015.01.004.
- Yekta, S., I-Hung Shih, and David P. Bartel (Apr. 23, 2004). "MicroRNA-directed cleavage of HOXB8 mRNA". In: *Science* 304.5670, pp. 594–596. ISSN: 1095-9203. DOI: 10.1126/science.1097434.

- Yeom, KH et al. (Sept. 2006). "Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing". In: *Nucleic Acids Res* 34.16, pp. 4622–4629. ISSN: 0305-1048. DOI: 10.1093/nar/gkl458.
- Yi, R. et al. (Dec. 15, 2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs". In: *Genes Dev* 17.24, pp. 3011–3016. ISSN: 0890-9369. DOI: 10.1101/gad.1158803.
- Yoda, M. et al. (Jan. 2010). "ATP-dependent human RISC assembly pathways". In: *Nat Struct Mol Biol* 17.1, pp. 17–23. ISSN: 1545-9985. DOI: 10.1038/nsmb.1733.
- Yoda, M. et al. (Nov. 14, 2013). "Poly(A)-specific ribonuclease mediates 3'-end trimming of Argonaute2-cleaved precursor microRNAs". In: *Cell Rep* 5.3, pp. 715–726. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2013.09.029.
- Youngman, E. M. and J. M. Claycomb (Nov. 27, 2014). "From early lessons to new frontiers: the worm as a treasure trove of small RNA biology". In: *Front Genet* 5, p. 416. ISSN: 1664-8021. DOI: 10.3389/fgene.2014.00416.
- Yu, S. and V. N. Kim (Sept. 2020). "A tale of non-canonical tails: gene regulation by post-transcriptional RNA tailing". In: *Nat Rev Mol Cell Biol* 21.9, pp. 542–556. ISSN: 1471-0080. DOI: 10.1038/s41580-020-0246-8.
- Yuan, S. et al. (Feb. 26, 2019). "Motile cilia of the male reproductive system require miR-34/miR-449 for development and function to generate luminal turbulence". In: *Proc Natl Acad Sci U S A* 116.9, pp. 3584–3593. ISSN: 1091-6490. DOI: 10.1073/pnas.1817018116.
- Zambrano, R. M. et al. (Feb. 2017). "Further evidence that variants in PPP1CB cause a rasopathy similar to Noonan syndrome with loose anagen hair". In: *Am J Med Genet A* 173.2, pp. 565–567. ISSN: 1552-4833. DOI: 10.1002/ajmg.a.38056.
- Zamore, P. D. et al. (Mar. 31, 2000). "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals". In: *Cell* 101.1, pp. 25–33. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)80620-0.
- Zekri, L., Duygu Kuzuoğlu-Öztürk, and Elisa Izaurralde (Apr. 3, 2013). "GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation". In: *EMBO J* 32.7, pp. 1052–1065. ISSN: 1460-2075. DOI: 10.1038/emboj.2013.44.
- Zeng, Y., R. Yi, and B. R. Cullen (Jan. 12, 2005). "Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha". In: *EMBO J* 24.1, pp. 138–148. ISSN: 0261-4189. DOI: 10.1038/sj.emboj.7600491.
- Zhang, H. et al. (July 9, 2004). "Single processing center models for human Dicer and bacterial RNase III". In: *Cell* 118.1, pp. 57–68. ISSN: 0092-8674. DOI: 10.1016/j.cell.2004.06.017.
- Zhao, J. et al. (Sept. 2013). "TP53-independent function of miR-34a via HDAC1 and p21(CIP1/WAF1.)" In: *Mol Ther* 21.9, pp. 1678–1686. ISSN: 1525-0024. DOI: 10.1038/mt.2013.148.
- Zisoulis, D. G. et al. (Feb. 2010). "Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*". In: *Nat Struct Mol Biol* 17.2, pp. 173–179. ISSN: 1545-9985. DOI: 10.1038/nsmb.1745.

## ABSTRACT

### Reciprocal regulation between microRNAs and their targets

Over the last 15 years, the several hundreds of identified microRNAs have been proposed to control numerous biological processes in healthy conditions or diseases. We studied two aspects of microRNA biology: the biological role of a perceived microRNA as a tumor suppressor; and the control of microRNA stability. Some microRNAs have been presented to act as pro-oncogenic or tumor-suppressor due to their role in controlling critical cellular pathways in the establishment of cancer. Nevertheless, a consensual definition of tumor-suppressing microRNAs is still missing. Similar to coding genes, we propose that tumor suppressor microRNAs must show evidence of genetic or epigenetic inactivation in cancers and exhibit an anti-proliferative activity under endogenous expression levels. In a first project, we tried out this definition with the miR-34a microRNA, which has attracted a lot of attention because it is regulated by the tumor-suppressor transcription factor p53 and became the first microRNA-based drug reaching clinical trial phase 1 in oncology. We used cancer genetics data to assess the expression level and the genetic status of miR-34a in multiple cancer types. We also performed genetic ablation to measure the endogenous function of this microRNA on cell proliferation in cancer cell lines and investigated in-depth previous over-expression studies showing its anti-proliferative effect to explain discrepancies with our results. Browsing a large diversity of cancer types, it appears that miR-34a is not down-regulated in primary tumors relative to normal adjacent tissues, and its gene does not accumulate mutations in cancer. Our work also shows that the established anti-proliferative action of miR-34a was based on over-expression experiments, leading to unrealistically high microRNA levels. Our data indicate that endogenous miR-34a levels do not have such an effect; therefore argue against a tumor-suppressive function for miR-34a. MicroRNAs repress mRNAs, but reciprocally, target mRNAs can also modulate microRNA stability. In a second project, we considered the endogenous regulation of microRNAs by mRNAs through target RNA-directed microRNA degradation (TDMD). Artificial targets as well as in vivo examples (viral transcripts, long non-coding RNAs *libra* in zebrafish and *Cyano* in mouse) have shown that extensive complementarity between a target RNA and a microRNA triggers the proteolysis of the microRNA-Induced Silencing Complex by the ubiquitin-proteasome pathway, which exposes the microRNA for degradation. Based on published data about target RNA patterns leading to TDMD and phylogenetic conservation, we developed a computational tool for the in silico identification of RNA sites that induce microRNA degradation through TDMD. Supplemented with published RNA-seq and small-RNA-seq data, our software allows to focus on cell-specific TDMD inducer candidates. Our search uncovered several convincing candidates in mouse neurons and their molecular characterization has been initiated.

---

### Régulation réciproque entre les microARN et leurs cibles

Au cours des 15 dernières années, des centaines de microARN ont été identifiés et proposés comme impliqués dans le contrôle de nombreux processus biologiques, en condition saine ou dans des maladies. Nous avons étudié deux aspects de la biologie des microARN : le rôle biologique d'un microARN perçu comme suppresseur de tumeur, et le contrôle de la stabilité des microARN. Certains microARN ont été définis comme pro-oncogéniques ou suppresseurs de tumeur de par leur rôle dans le contrôle de voies cellulaires critiques dans l'établissement de cancer. Néanmoins, une définition consensuelle d'un microARN suppresseur de tumeur fait toujours défaut. Comme pour les gènes codants, nous proposons que les microARN suppresseurs de tumeur doivent démontrer des signes d'inactivation génétique ou épigénétique dans les cancers et présenter une activité anti-proliférative à des niveaux d'expression endogènes. Dans un premier projet, nous avons testé cette définition avec le microRNA miR-34a, qui a attiré beaucoup d'attention puisqu'il est directement régulé par le facteur de transcription suppresseur de tumeur p53 et est devenu le premier miARN testé en tant que médicament atteignant la phase 1 des essais cliniques en oncologie. Nous avons utilisé des données de séquençage de cancers pour évaluer le niveau d'expression et le statut génétique de miR-34a dans plusieurs types de cancer. Nous avons également réalisé une ablation génétique dans des lignées de cellules cancéreuses afin de mesurer la fonction endogène de ce microRNA sur la prolifération cellulaire et avons examiné en profondeur les études antérieures de surexpression montrant son effet anti-prolifératif pour expliquer les divergences avec nos résultats. En parcourant une grande diversité de types de cancer, il apparaît que miR-34a n'est pas inhibé dans les tumeurs primaires par rapport aux tissus normaux adjacents, et que le locus exprimant ce microARN ne présente pas d'accumulation de mutations dans les cancers. Notre travail montre également que l'activité anti-proliférative établie du miR-34a est basée sur des expériences de surexpression conduisant à des niveaux de microRNA irréaliment élevés. Nos données indiquent finalement que les niveaux endogènes de miR-34a n'ont pas un tel effet et plaident donc contre une fonction tumeur-suppressive pour miR-34a. Les microARN répriment les ARNm, mais réciproquement, les ARNm cibles peuvent également moduler la stabilité des microARN. Dans un second projet, nous avons examiné la régulation endogène des microARN par les ARNm via la dégradation des microARN dirigée par les ARN cibles ou "target RNA-directed microRNA degradation" (TDMD). Des cibles artificielles ainsi que des exemples in vivo (des transcrits viraux, l'ARNlnc *libra* chez le Poisson-zèbre, l'ARNlnc *Cyano* chez la Souris) ont permis de démontrer qu'une complémentarité étendue entre un ARN cible et un microARN entraîne la protéolyse du complexe protéique associé au microARN ou "microRNA-Induced Silencing Complex" par la voie ubiquitine-protéasome, ce qui expose le microARN à la dégradation. Sur la base des données publiées sur les modèles d'ARN cibles conduisant au TDMD et à l'analyse de la conservation phylogénétique des génomes, nous avons développé un outil informatique permettant l'identification in silico des sites d'ARN qui induisent la dégradation des microARN par TDMD. En le complétant avec des données de RNA-seq et de small-RNA-seq publiées, notre logiciel permet de faire ressortir les candidats inducteurs de TDMD spécifiques à un type cellulaire. Nos résultats ont permis d'identifier plusieurs candidats convaincants dans les neurones de souris et leur caractérisation moléculaire a été initiée.