

# Functional characterization of C/D snoRNA-derived microRNAs

Dissertation

for the award of the degree:

**„Doctor of Philosophy” (Ph.D.)**

of the Georg-August-University Göttingen

within the doctoral program Molecular Biology

of the Georg-August University School of Sciences (GAUSS)

Submitted by

Gustavo Nicolás Lemus Díaz M.D.

From Bogotá, Colombia

Göttingen 2017



PhD Thesis Committee:

Jens Gruber Ph.D  
Primate Genetics Laboratory  
Junior Research Group Medical RNA Biology  
German Primate Center, Göttingen

Prof. Dr. Reinhard Lührmann  
Max Planck Institute for Biophysical Chemistry, Göttingen  
Department of Cellular Biochemistry

PD. Dr. Halyna Shcherbata  
Max-Planck-Institute for Biophysical Chemistry, Göttingen  
Gene Expression And Signaling Group

Extended Evaluation Committee (alphabetically):

Dr. Alexander Hahn  
Infection Biology Unit  
Junior Research Group Herpesviruses  
German Primate Center, Göttingen

Prof. Steven Johnsen Ph.D  
Clinic for General, Visceral and Pediatric Surgery  
University Medical Center Göttingen

Prof. Dr. Henning Urlaub  
Max-Planck-Institute for Biophysical Chemistry, Göttingen  
Bioanalytical Mass Spectrometry Group

Date of the oral examination: 8.12.2017



## Affidavit

Herewith I declare, that I prepared this thesis "Functional characterization of C/D snoRNA-derived miRNAs" independently and with no other sources and aids than quoted.

Gustavo Nicolas Lemus Diaz

Göttingen 14.10.2017



## TABLE OF CONTENT

Table of content.....	I
List of Figures.....	III
List of Abbreviations: .....	V
Acknowledgments .....	VII
Abstract.....	XI
<b>1 Introduction.....</b>	<b>1</b>
<b>Small RNAs: petite genome regulators .....</b>	<b>1</b>
1.1.1 small nuclear RNA: cut and paste to diversify Expression .....	2
1.1.2 microRNAs: tiny regulators of gene expression.....	3
1.1.3 small nucleolar RNA: modifying to diversify RNAs .....	8
<b>2 Aims of the thesis .....</b>	<b>14</b>
<b>3 Materials &amp; Methods .....</b>	<b>15</b>
<b>Molecular biology .....</b>	<b>15</b>
3.1.1 miRNA Dual Fluorescent reporter generation .....	15
3.1.2 GAS5 construct generation .....	16
3.1.3 bacterial transformation and competent cells generation and plasmid isolation ....	16
3.1.4 Polymerase chain reaction and Restriction endonuclease reactions.....	16
3.1.5 RNA Isolation, size and integrity evaluation .....	17
3.1.6 cDNA synthesis and Quantitative PCR analysis.....	18
3.1.7 NGS library preparation .....	18
Next generation sequencing .....	18
<b>Cell biology .....</b>	<b>19</b>
3.1.8 Cell culture .....	19
3.1.9 DNA and siRNA Transfection. ....	19
3.1.10 Luciferase Reporter Assays.....	19
3.1.11 Fluorescence microscopy and flow cytometry .....	20
<b>Protein and RNA biochemistry .....</b>	<b>20</b>
3.1.12 Western Blot.....	20
3.1.13 Northern Blot.....	21
<b>Bioinformatics and statistical analyses.....</b>	<b>22</b>
3.1.14 Next generation sequencing data analysis.....	22
3.1.15 Graphs generation .....	23
3.1.16 Analytical Flow Cytometry.....	24
3.1.17 Other statistical analysis.....	25
<b>Materials .....</b>	<b>25</b>
3.1.18 Machines .....	25
3.1.19 Buffers and Solutions.....	26
3.1.20 Chemicals and Materials.....	27
3.1.21 Bacterial strains information .....	28

3.1.22	Cell lines .....	29
3.1.23	Cell culture media and additives .....	29
<b>4</b>	<b>Results .....</b>	<b>29</b>
	<b>NGS characterization of snoRNA-derived miRNAs .....</b>	<b>30</b>
4.1.1	snoRNA-derived miRNA characterization from small RNA libraries. ....	30
	<b>Scrutinizing miRNAs repression activity at the single cell level .....</b>	<b>52</b>
4.1.2	Generation of dual fluorescence plasmids: Proof of Principle.....	53
4.1.3	Single cell level analysis: molecular titration model for miRNA induced postranscriptional regulation.....	55
4.1.4	Measuring Endogenous miRNA activity at the single cell level.....	69
	<b>Endogenous activity of C/D snoRNA-derived miRNAs .....</b>	<b>75</b>
4.1.5	Analytical flow cytometry of C/D snoRNA-derived miRNAs Threshold modulation. ....	75
4.1.6	Analytical flow cytometry of C/D snoRNAs: Outcompeting miRNAs. ....	76
4.1.7	CLASH chimeras for U3 5' domain derived miRNAs. ....	79
	<b>Insight into snoRNA-derived miRNAs biogenesis .....</b>	<b>80</b>
4.1.8	hMTR4 knock-down increases C/D snoRNA expression in Hek293 cells. ....	80
4.1.9	Exogenous genomic Gas5 increases U78, U44 and U74 derived miRNAs. ....	81
<b>5</b>	<b>Discussion .....</b>	<b>86</b>
	<b>C/D snoRNA-derived miRNAs-like molecules .....</b>	<b>86</b>
5.1.1	C/D snoRNA-derived miRNAs differ from canonical miRNAs.....	87
5.1.2	5' U3 Domain source of miRNAs .....	90
5.1.3	C/D miRNAs degradation vs. functionality. ....	93
5.1.4	Analytical Flow cytometry high-resolution platform for miRNAs functionality.....	94
<b>6</b>	<b>Summary.....</b>	<b>98</b>
<b>7</b>	<b>Bibliography.....</b>	<b>100</b>
<b>8</b>	<b>Appendix.....</b>	<b>108</b>
	<b>Supplementary figures. ....</b>	<b>108</b>
	<b>Plasmid used and generated in this Thesis. ....</b>	<b>119</b>
	<b>Oligo sequences used for miRNA analysis at single cell level .....</b>	<b>120</b>



## LIST OF FIGURES

Figure 1. Biogenesis of small nuclear RNPs .....	3
Figure 2. Biogenesis of microRNAs .....	4
Figure 3. Structure of small nucleolar RNAs .....	9
Figure 4. Biogenesis of intronic small nucleolar RNAs.....	10
Figure 5. Biogenesis of U3 small nucleolar RNA .....	12
Figure 6 QR code to replicate on line numerical simulation for miRNA titration model. ....	25
Figure 7. NGS Characterization of snoRNA-derived miRNAs .....	30
Figure 8. small RNA seq libraries revealed expected distribution for snoRNA and miRNAs. ....	32
Figure 9 snoRNA-derived miRNAs expression differs from mature counterparts.....	33
Figure 10 Reads hierarchical clustering analysis sorts snoRNA-processing patterns .....	34
Figure 11. snoRNA-derived miRNA show heterogeneous processing. ....	35
Figure 12. U78 derived miRNA detection by northern blot.....	36
Figure 13. Outline of raw data recovery for size separated RNA. ....	37
Figure 14. snoRNA-derived miRNAs content is lower than mature snoRNAs and their expression positively correlates. ....	38
Figure 15. snoRNA-derived miRNAs locate in the nucleolus .....	39
Figure 16. CLIP and RIP-seq analyzed data. ....	41
Figure 17. Hek293 small RNA, RIP and CLIP libraries recover comparable miRNA profiles. ....	42
Figure 18. Argonaute CLIP and RIP identifies 3 bound snoRNA-derived miRNAs .....	43
Figure 19. Dicer PAR-CLIP recovers U3 snoRNAs fragments. ....	44
Figure 20. DGCR8 HITS-CLIP recovers snoRNA and miRNAs.....	45
Figure 21. snoRNPs PAR-CLIP recover known snoRNAs. ....	46
Figure 22. Nop56 and DGCR8 snoRNA recovery resembles snoRNA-derived miRNA clusters.....	48
Figure 23. snoRNP and miRNPs CLIP-seq recover U3 and ACA45. ....	49
Figure 24. U3 snoRNA 5'-domain interacts with miRNPs .....	51
Figure 25. Advanced dual fluorescent reporter for the comprehensive, high-resolution analysis of miRNA based regulation. ....	53
Figure 26 Untranslated trans assay: theoretical principle .....	54
Figure 27 Dual Fluorescence Reporter characterized in-vivo miRNA activity .....	55
Figure 28 Small RNAs gene regulation: the simple picture. ....	57
Figure 29 Threshold linear response of a target mRNA .....	58
Figure 30. Titration model of miRNAs induced regulation .....	59
Figure 31 Numerical simulations describe miRNA-induced thresholding .....	61
Figure 32 Dual fluorescent reporter analytical flow cytometry: Function generation from single cell measurements .....	62
Figure 33 Exogenous expression of siRNA shift analytical functions <i>in vivo</i> . ....	63
Figure 34. Increasing reporter binding sites sharpens siRNA induced threshold <i>in vivo</i> .....	64
Figure 35. <i>In vivo</i> -derived analytical functions describe three functional groups.....	65
Figure 36 .Threshold transition to linearity depend on miRNA-mRNA complex formation.....	67
Figure 37 Threshold limit is controlled by miRNA expression. ....	68
Figure 38 miRNA Analytical functions uncouple expression from functionality .....	69
Figure 39 miRNA titration model integrates expression and binding <i>in vivo</i> .....	70
Figure 40 miRNAs Subcellular localization influences their analytical output .....	71
Figure 41. Threshold sharpening for endogenous miRNAs.....	74
Figure 42. Threshold shift by outcompeting miRNA with siRNA <i>in vivo</i> .....	75
Figure 43. snoRNA-derived miRNAs analytical resemble low functional miRNAs.....	77
Figure 44. U3 derived miRNA function resembles low functional miRNAs. ....	78
Figure 45. Outcompeting miRNA-derived snoRNA shift the ACA45 and 5'-U3 domain analytical function. ....	79

Figure 46. 5'-U3 domain derived miRNA's chimeras identify mRNA targets .....	80
Figure 47. hMTR4 knock-down increases mature C/D snoRNA expression in Hek293 cells. ....	82
Figure 48. snoRNA-derived miRNAs are unaffected by hMTR4 knock-down C/D snoRNA expression in Hek293 cells. ....	83
Figure 49. Genomic GAS5 (gGAS) exogenous expression increases C/D derived miRNAs in Hek293 cells ...	84
Figure 50. U3 structure and 5'domain miRNP interaction .....	90
Figure 51. Hypothetical model for U3 derived miRNAs.....	92

## **LIST OF ABBREVIATIONS:**

ncRNA: non-coding RNAs

sRNA: smallRNAs

miRNA: microRNAs

snRNA: Small nuclear RNAs

snoRNA: Small nucleolar RNAs

RISC: RNA-induced silencing complex

NGS: Next generation sequencing

5' ETS : 5' External transcribed spacer

CLIP: Cross linking immunoprecipitation

PAR-CLIP: Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation

HITS-CLIP: High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation

CLASH: Cross-linking ligation and sequencing of hybrids

CRAC: Cross-linking and analysis of cDNAs

RIP:RNA Immunoprecipitation

RNA-seq: RNA sequencing

C/D-miRNAs: C/D snoRNA-derived miRNAs

FACS: Fluorescence-activated cell sorting

UTA: Untranslated trans assay for analytical flow cytometry

AGO: Argonaute

snoRNP: Small nucleolar ribonucleoproteins

miRNP: miRNA ribonucleoproteins



## ACKNOWLEDGMENTS

First I would like to thank the German taxpayers, without their contribution and political decision I couldn't have achieved this doctoral dissertation, at the same time I appreciate the Colombian taxpayers because without their support to the public education I wouldn't have evolved academically to be admitted here.

I thank Jens Gruber for his support and supervision, the freedom he gave to develop my ideas and primarily to teach me that science is relevant, but family, values, and to be honest give a meaningful life and the strength to deal with the vicissitudes of the scientific entrepreneur.

I thank my thesis committee for their support Reinhard Lührmann and Halyna Shcherbata, thank you professor Lührmann for helping me finding the focus I needed to develop this work. To professor Markus Bohnsack for the long brainstorming session that enriched my ideas and helped me enormously.

To Lutz Walter for allowing me to use his lab the first years and to Christian Roos thanks for the beers and good times, thank you Angela Nool for supporting me on my Bioinformatics path and helping me with my shyni app.

I like to thank Kai Böker for our extended stays in the lab discussing and setting things up from scratch gave us a valuable friendship. Also, I big thank you to Rafael Rinaldi who became a good colleague and brought the Latin vibe we needed in the lab.

To Stefan Lüdke and Stefan Schneider the best post-doc I could ever have, thank you for showing me the high idealism that brings satisfaction; thanks to Lara Shiller for helping me correcting my lazy dyslexia. To Ellen Eckermann-Felkl and Angelina Schuder for their excellent technical support, Angie I also appreciate your friendship and love for work that brings happiness and efficiency at the same time.

To my rotation students, Michael Mitter, Jasmin Preis and Liezel Tamon, your presence in the lab was perfect to develop my team skills. Jasmin thank you for being the sister I never had; I hope we see together in the academic world again. Ignacio Rodriguez gracias por estar aquí y trabajar juntos.

To my closest friends Diego Giraldo, Lina Jaime, Sharlen Moore, Burak Bali, Tomas Offner and Stefan Guse thank you for your support and companion. Shar thanks for showing me the strong

Latin American impulse towards perfection; it is never easy, but it brings great satisfaction. Stefan without you I wouldn't have evolved as a philosopher thank you very much.

Gracias Nora Cascante, tu amor por el castellano y la alegría que emanabas trajo la luz en mis días oscuros, gracias por ser mi hermana mayor, and of course thank you Mike I enjoy your never ending positive vibe, Barbara Flix mi catalana favorita gracias por tu apoyo y amistad. Jeniffer Ospina te agradezco estar aquí, una década no fue suficiente para ser amigos pero el exilio Alemán lo fue.

Ana María Perdomo gracias por estar siempre allí y quererme por quien soy, gracias Carlos Clavijo por enseñarme todo lo que sé y ser mi modelo a seguir. Gracias Fabio Vilardi por estar allí, científicamente y personalmente, no puedo tener nadie más importante en ambos campos, gracias por enseñarme a amar la ciencia, el pasado y la vida.

Thank you, Stefan Schneider, Nora Cascante and Derek Victor for the kind and thoughtful comments in the document and your support.

To Steffen Burkhardt and Kerstin Grüniger, your support and work have been not just important but essential. Thank you for teaching me to love a job and do it with passion; you are the central point for the Molbio family, to the GGNB staff thank you for your support, it was a pleasure working with you, Gracias Kirsten por ser mi amiga fan de Alexander Humboldt.

## ORIGINAL PUBLICATIONS

Hoffmann, D. Gruber, J. Böker K.O. Deppe D. Sehmisch S. Arndt Schilling A. **Lemus-Diaz, N.** Komrakova M, and Schneider S (2018) "Effects of RANKL-Knockdown by Virus-like Particle mediated RNA-interference in a rat model of osteoporosis" *Molecular Therapy - Nucleic Acids. In press*

Schiller, L. T\* and **Lemus-Diaz, N\***. et al. (2018) Enhancing exosome-associated AAV production by CD9 over-expression. *MolTher Methods & Clinical Developments*, DOI: 10.1016/j.omtm.2018.03.008

Böker, K. O\* and **Lemus-Diaz, N\***. et al.(2017). The impact of the CD9 tetraspanin on exosome secretion and lentivirus infectivity. *Mol Ther.* 2017 Nov 16. pii: S1525-0016(17)30567-1.

**Lemus-Diaz, N.** et al. (2017). Dissecting miRNA gene repression on single cell level with an advanced fluorescent reporter system. *Scientific reports* **7**, 45197, doi:10.1038/srep45197

\*Contributed equally to this work

In Preparation:

**Lemus-Diaz, N.**, Rinaldi Ferreira, R. Tamon, L, Gruber, J., (2017) U3 C/D snoRNA 5' domain is source of miRNAs.

Schneider, S., Schiller, L. T., Böker, K. O., **Lemus-Diaz, N.**, Gruber, J., (2017). Retargeted JC polyoma derived virus-like particles as efficient and specific cell transduction tool.





## ABSTRACT

miRNAs are essential regulators of cell fate and involved in several human diseases, although their biogenesis is a well-accepted paradigm (microprocessor, Exp5, Dicer, and Argonaute). New evidence revealed different biogenesis for miRNAs distinct to canonical miRNA pathway; this new source of miRNAs including mirtrons, Exp1 dependent and Dicer-independent miRNAs, furthermore also others ncRNAs like tRNAs and snoRNAs produce fragments that reflects Dicer processing and Argonaute incorporation in sRNA profiles, however, functional evaluation and mechanisms are still poorly described. Studying C/D snoRNA-derived miRNAs using NGS profiles (sRNA-seq and CLIP of ribonucleoproteins) called small RNA fragments derived from C/D snoRNA loci poorly recovered from miRNP (Argonaute, Dicer) but strongly associated to snoRNPs (Nop56, Nop58, and FBL). However, U3 (C/D snoRNA) suited as a *bona fide* source of miRNAs including defined mRNA targets. To test functionality in-vivo at high resolution, I generated, implemented and validated a single cell assay for miRNA posttranscriptional regulation. Using this system U3 derived fragments perform as low expressed miRNAs in vivo in the Hek293 model and hold mRNA endogenous targets.

In conclusion, using NGS data and the high-resolution reporter system, this study showed that non-canonical U3 C/D snoRNA is a *bona fide* miRNA source, while intronic C/D snoRNA-derived fragments rather might be degradation products.



## 1 INTRODUCTION

RNA biology paradigms redefined themselves and evolved quickly, bringing together fascinating ideas and scenarios where the flexibility and multi-faceted RNA nature surprises us with unconceived mechanisms<sup>1</sup>.

However, this increased flexibility and boost of information have created different "schools"<sup>2,3</sup> around particular phenomena, lacking consensus perspectives or paradigms. Building applications and research based on incomplete or poorly understood mechanisms. For example, Gunther Meister's lab revealed that snoRNAs and tRNAs could generate miRNA-like molecules and described their biogenesis path<sup>3-5</sup>. But deeper understanding and systematicity around the mechanism are inadequate. Since then many studies reported tRNA or snoRNA fragments suggesting association with pathological scenarios, assuming Meister's evidence as a paradigm and not as proof of principle that requires replication and further validation, as many reviews claim<sup>6-11</sup>.

As my central research interest, I want to characterize the function of snoRNA-derived miRNAs. So, I will introduce small RNAs current paradigms, including function and biogenesis, examining common trafficking factors and functional differences. I will focus on current miRNA models of transcription, transport, and processing. Also, I will discuss recent experimental advances, concentrating on the lack of high-resolution sensors.

Finally, I will outline the canonical function and biogenesis of snoRNAs, including a description of new features, evidence for them, and the flaws that require experimental clarification or validation.

### SMALL RNAs: PETITE GENOME REGULATORS

Small RNAs are the best-studied non-coding RNAs (ncRNAs) and what we know about their genesis, processing, and mechanisms exemplify RNA diversity. Interestingly their trafficking principle is not unique and probably governed by the requirements of the cell rather than RNA subtype exclusive pathway<sup>12</sup>.

Current consensus classifies ncRNAs arbitrarily by size creating two types of ncRNA: long non-coding RNAs (lncRNA) and small RNAs (sRNAs), while the firsts are larger than 200 nt, the

seconds are shorter<sup>13</sup>. Besides the different functions, the small RNAs have in common that they use base-pairing interactions to modify their target RNAs, for what they associate with RNA binding proteins (RNP).

The small RNAs include snRNA, snoRNA, and miRNA: small nuclear RNAs (snRNAs) aid intron removal, small nucleolar RNAs (snoRNAs) guide rRNA chemical modifications and miRNAs regulate gene expression by inducing degradation or translational repression of mRNAs.

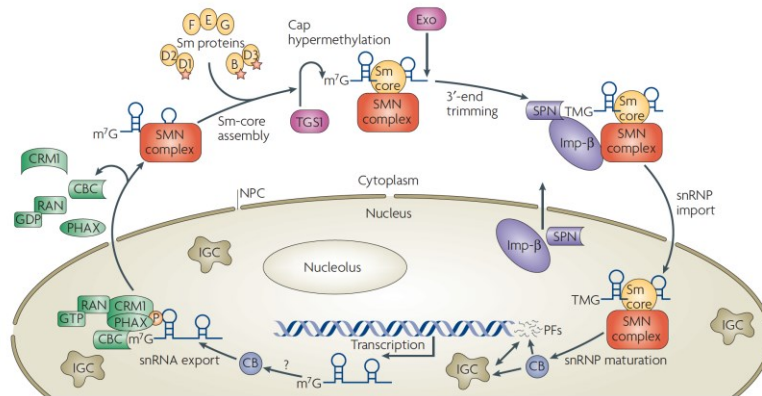
### *1.1.1 SMALL NUCLEAR RNA: CUT AND PASTE TO DIVERSIFY EXPRESSION*

Small nuclear RNAs are abundant and non-polyadenylated RNAs located mainly at the nucleoplasm; they are the spliceosome core and catalyze intronic removal. Within snRNA there are two distinct classes: Sm and Lsm classes, divided accordingly to sequences motif and interacting proteins, the Sm snRNA classes include U1, U2, U4, U5, U7, U11, and U12, while U6 is a Lsm snRNA.

Lsm class snRNAs have a monomethyl phosphate cap, a 3' stem-loop, and a uridine stretch that serves as the binding site for the Lsm proteins, while the Sm-Class snRNA have a 5'-trimethylguanosine cap, a 3' stem loop, and the Sm proteins binding motif<sup>12</sup>.

snRNA splicing activity requires ribonucleoproteins to organize it with the pre-mRNA, the snRNA-mRNA interaction resembles the conformation of the RNA-based catalytic center of group II introns. The snRNPs flexibility allows a precise regulation that enables interaction with different intronic sizes and control of alternative splicing. Reviews from Lührmann and Shi<sup>14,15</sup> describe mechanistic details on the structure and chemistry of catalytic steps during splicing.

RNA polymerase II transcribes all Sm snRNAs and couples their 3' processing. After transcription three proteins (Cap binding protein, adaptor PHAX and Exportin 1) bring snRNAs to the cytoplasm; a process that is dependent on GTP and PHAX phosphorylation state. In the cytoplasm, the SMN complex assembles the Sm core (a ring of 7 sm proteins), while TSG1 sets the snRNA three methyl cap. Finally, Snurportin1 (SPN) and importin B return the snRNAs to the nucleus, where in the Cajal bodies snRNAs get further modified and end in the nucleoplasm to be part of the spliceosome (Figure 1).



Reproduced with permission from Matera et al.<sup>12</sup>

### Figure 1. Biogenesis of small nuclear RNPs

Cartoon depicts the transcription, trafficking and modification of SM snRNA.

U6 has a different processing; it remains in the nucleus, is not trimethylated and is transcribed by polymerase III. Furthermore, in contrast to other Sm snRNAs, U7 has a distinct SM motif<sup>12</sup>.

#### 1.1.2 MICRORNAs: TINY REGULATORS OF GENE EXPRESSION

miRNAs are by far one of the greatest revolutions in RNA and cell biology; it is probably the "millennial" discovery and quickly scaled as a new paradigm that we take for granted and even included in high school curricula, meaning that miRNAs override the specialized knowledge and enter human kind collective consciousness.

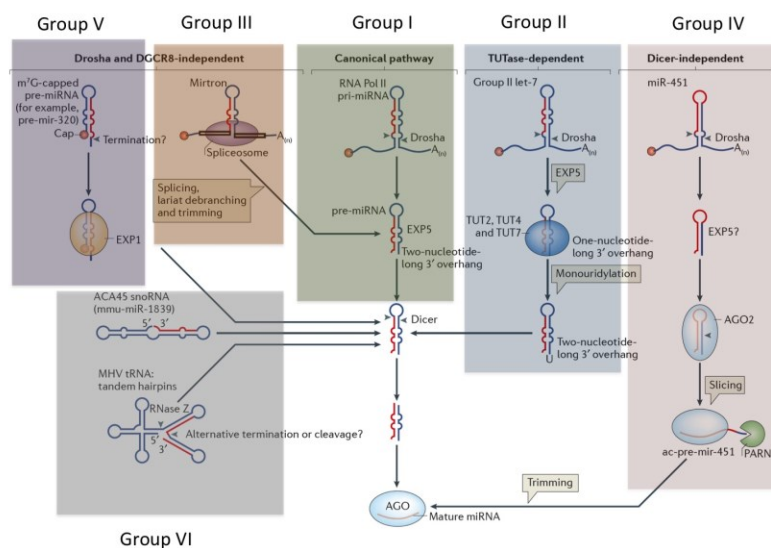
microRNAs regulate gene expression by binding to mRNAs by sequences complementarity. Initially, described as an odd discovery in *C.elegans*<sup>16,17</sup>, and considered relevant until 2001 when several groups reported numerous 21 to 25 nt RNAs that were unambiguously conserved in evolution and relevant for development; that year the term miRNA was born<sup>18-20</sup>.

miRNA genes yield bigger precursors than the final mature counterparts; Polymerase II transcribes the pri-miRNA as a polyadenylated and 7mG capped transcript. Then the microprocessor, which includes DROSHA and DGCR8, processes it and produces a shorter version (~70 nt) with a stem-loop structure (pre-miRNA), later it is transported to the cytoplasm by Exportin 5. In the cytoplasm, the endonuclease DICER, which interacts with the TRBP protein, cuts the stem loop and generates a RNA duplex. Afterward, the duplex is released, and the heat shock cognate (HSC70) and the heat shock protein 90 (HSP90) hydrolyse ATP to load the RNA duplex in the Argonaute proteins, which select the guide strand by 5' stability of the duplex<sup>8</sup>.

miRNAs repress gene expression by four different ways: they degrade the transcript, inhibit translation elongation, induce early termination of translation or avoid its initiation. miRNAs cause degradation by removing the m7G cap and reducing the length of the poly A tail. As well, they sequester transcripts in P-bodies inhibiting translation<sup>21</sup>.

### 1.1.2.1 microRNAs Biogenesis

The cell generates the vast majority of miRNAs through the pathway described above. However several studies point to alternative routes for miRNAs biogenesis, and the paths can be grouped into 6 clusters according to their genomic location, processing, and trafficking proteins. I organized the groups according to the similarities to the canonical miRNAs; meaning that group II is almost identical to the textbook miRNA biogenesis and VI is essentially a different path (Figure 2).



Reproduced and modified with permission from Ha and Kim<sup>8</sup>

### Figure 2. Biogenesis of microRNAs

miRNA groups defined by biogenesis: I. canonical miRNAs, II. let-7 that requires a extra mono uridylation, III. Mirtrons that do not require the microprocessor, IV. Dicer independent but require Argonaute 2 catalytic activity, V. m7-Cap at the 5' miRNA that do not requires DGCR8\*, and are XPO1 dependent, VI. Other source of miRNAs.

\*Group V XPO1 dependent could have m7-Cap or TMG-Cap.

Group I corresponds to the canonical pathway (m7G-cap and poly A, microprocessor-Exp5-Dicer-Ago)<sup>8</sup>. The second group is similar to canonical miRNAs, but it holds an extra step; usually, Dicer substrates require a 3' overhang of 2 nucleotides and the group II let-7 family only possesses one nucleotide, so mono-uridyl transferases (TUT) add the missing nucleotide, allowing processing by Dicer<sup>22</sup>. The third group comprises the Mirtrons (intronic miRNAs) that require the spliceosome, the debranching enzyme activity and trimming by nucleases before Exp5 ships them out the nucleus to continue with miRISC formation<sup>23,24</sup>. The group IV is Dicer-independent and requires the catalytic activity of Ago2, since its stem is too short to be processed by Dicer. Ago2 cuts the miRNAs while the nuclease PARN trims it, producing a RISC complex with a miRNA of 23 nt<sup>25,26</sup>.

Finally, group V miRNAs are transcribed similarly to snRNAs. Hence, they have a 7mG cap and probably are subjected to 3' terminal processing. As snRNA they are transported to the cytoplasm by PHAX and XPO1; because the cap, just the 3' branch of the hairpin is effectively loaded in Ago<sup>27,28</sup>. Furthermore, a subset of these miRNAs is expressed in starvation conditions, because Exportin V expression is decreased while the Exportin I transports this subset of miRNAs<sup>29</sup>. Summarizing the group V, which mainly described by Joan Steitz lab excellent studies, is a collection of miRNAs that require XPO1 transport and have either 7mG-cap or TMG-cap depending on the metabolic cell status<sup>27,29</sup>.

The group VI includes miRNA derived from another ncRNA like snoRNAs (See below) or tRNAs. Their function is still debatable due to their low expression, and probably they are the result of leaky processing. For example, tRNAs interact with La proteins avoiding tRNAs to join the miRNA biogenesis pathway<sup>4,5</sup>.

### **1.1.2.2 Experimental strategies for microRNA functionality**

For the evaluation of miRNAs several biochemical and computational techniques have been developed in the field. These allow us to unravel the details of new mechanisms and associations that are applicable to clinical scenarios. For the evaluation of miRNAs several biochemical and computational techniques have been developed in the field. These allow us to unravel the details of new mechanisms and associations that are applicable to clinical scenarios. I will concentrate on high-throughput procedures, including profiling methods, *in vivo* target validations and single-molecule imaging. Moreover, I will further describe the gap between biochemical and the functional resolutions: while the biochemical techniques include quantitative high-throughput data, functionality *in vivo* is still qualitative<sup>21</sup>.

To have working definitions, I refer here as functionality to any experimental approach that evaluates the reduction of targets, and its phenotypic consequences, i.e., luciferase-based assays that quantify sensor changes in response to miRNAs *in vivo*. I describe here as *in vivo* to any test performed in a living system including cell culture; and I call biochemical approaches to miRNAs or target molecules quantification using amplification, cDNA synthesis, immunoprecipitation and others.

Steinkraus<sup>21</sup> describes in detail the experimental advances in miRNA discovery; however, I will summarize this review and include newly developed techniques that are not cited or poorly explained.

#### *1.1.2.2.1 miRNA profiling and capture-based techniques*

Deep sequencing and mass spectrometry have increased the possibilities to evaluate miRNAs and gene expression. However, The first and probably most used inquiry for miRNAs is primarily miRNA profiling. Briefly, it counts the number of miRNA molecules and runs statistical analyses to get differential expression either of miRNAs or mRNA, connecting the outputs (miRNA-mRNA targets) by computational dissection. Together, these profiles generate hypothesis that need further experimental evaluation.

Using miRNA exogenous expression or inhibition, scientists can evaluate and quantify miRNA targets. Depending on the mechanism they are eager to study, miRNA can either induce mRNA degradation or translation inhibition (see 1.1.2). Consequently, mRNA analysis on cDNA level (qPCR or next generation sequencing) evaluates RNAs degradation, mass spectrometry defines the affected coding targets, and ribosome profiling determines which miRNAs targets are translationally repressed.

Both strategies collect the total miRNA content, including then Argonaute bound molecules, miRNA in biogenesis, differentially localized or bound to free mRNA. So biochemist developed bait-based techniques in which miRNAs are linked to the effector molecule, narrowing down to the functional miRNA pool.

Primary profiling capture-based recoveries rely on cDNA synthesis and analysis (qPCR, microarrays or NGS), and differ on protein purification and RNA stability. Therefore, they are dependent on several factors: if the RNA can be crosslinked to Argonaute or not, or if it is performed via RNA immunoprecipitation (RIP) or crosslink-immunoprecipitation (CLIP).



CLIP approaches differ on their cross-linking mechanism, either they use photoactivatable ribonucleosides or the natural RNA trend to react with specific amino acids using 254 nm light. Moreover, including a ligation RNA-RNA, it is possible to sequence miRNA-mRNA targets, calling those hybrids chimeric reads.<sup>30-32</sup>

Hybrids can be found in several data sets, and some chimeric calls do not require an extra ligation step<sup>32</sup>. Recently, Amy Pasquinelli's lab developed a PCR amplification of chimeras, allowing cheaper and focused evaluation of *in vivo* derived chimeras (ChimP)<sup>33</sup>.

#### 1.1.2.2.2 *Systems Biology and in-vitro single molecules studies*

Recently FRET-based *in vitro* assays calculated kinetic parameters for AGO-miRNA binding and RISC formation, increasing the understanding of miRNA target recognition and the structural basis that allows better insights into the miRNA targeting mechanism from mRNA-miRNA hybrids and structural studies<sup>34,35</sup>. Since laboratory scientists use different techniques and the classic biochemistry approaches have difficulties to integrate data. Computational scientists develop algorithm to understand the complicated miRNA networks<sup>36,37</sup>.

To solve this issue, many system biologists use high-throughput sequencing data, phenotyping studies and *in vitro* kinetics to build a better look at the whole miRNA functional network. Thanks to system biology we understand miRNAs fine-tuning behavior and their switch-like control of gene regulation<sup>38-40</sup>.

Modeling miRNA networks allows to understand in a broad sense the miRNA induced regulation; these models generate confident predictions. Sadly, just few scientists use the models and apply them to new experimental questions, making a prospective analysis of their predictions. Rephrasing, we use data to create models, confirm the models retrospectively and often without testing their predictions experimentally.

Therefore, I decided to use the mathematical models for miRNAs to test empirically the group VI miRNAs derived from C/D snoRNAs (see section 0).

### 1.1.2.2.3 *Functional sensors of miRNA activity: High-resolution sequencing in a weak functional read outs world*

The increasing complexity and detail of information yielded by miRNA profiling and Argonaute FRET did not result in a boost of better *in vivo* reporters.

Commonly, scientists involved in exhaustive biochemical characterizations (Bait profiling) tested functional miRNA-mRNA interaction using luciferase reporter assays 5,21,30. These experiments require two proteins, the response element for miRNAs and a reference protein, to create quantitative results (ratios) for miRNA regulation and target validation for mammalian cells. But luciferase outputs dismissed the intricate functional network behind miRNAs.

miRNAs with comparable cellular concentration (as detected in miRNA profiling) show a broad range of functional outputs. For instance, high-throughput assays for miRNA activity (based on lentiviral libraries of miRNA sensors) showed that only highly abundant miRNAs repressed mRNA target and reported that around the 60% of miRNAs displayed no discernable activity <sup>41</sup>. Moreover, some highly expressed miRNAs have only weak regulatory action. Consequently, many studies on miRNAs inhibition suggested that particular miRNA expression thresholds are required to obtain a detectable repression <sup>42,43</sup>.

I consider that high-throughput functional studies have low resolution similarly to luciferase assays because such libraries require cell selection (antibiotic) and described the whole population. Also, their inability to discern down regulation -for up to 60% of the evaluated miRNAs- contradicts the fact that inhibition of mildly expressed miRNAs displayed discernible phenotypes. (See regulation of non-canonical targets of miR92a in CLASH <sup>30</sup>)

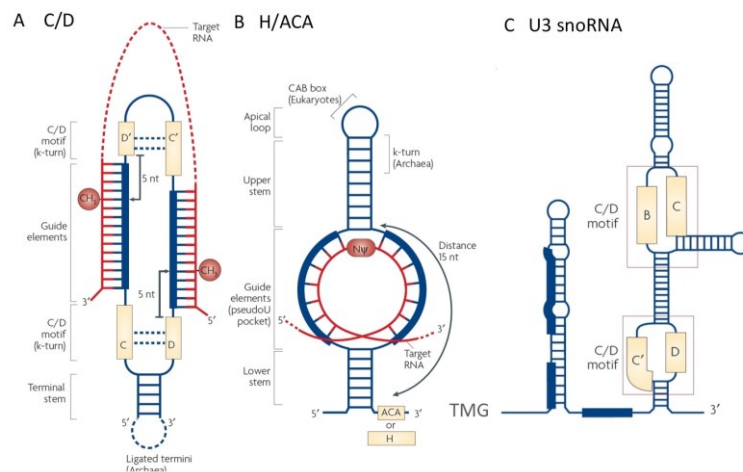
For that reason, I generated a single plasmid reporter assay and validated it for miRNAs. To increase resolution I used fluorescent proteins instead of luciferase and quantified them at single cell level using flow cytometry. To analyze the enormous amount of data points from individual cell measurements: I adapted a titration model of miRNAs regulation<sup>44</sup>, tested its predictions prospectively for canonical miRNAs and applied it for non-canonical miRNAs or Group VI. (See scope of the thesis section 0).

### 1.1.3 *SMALL NUCLEOLAR RNA: MODIFYING TO DIVERSIFY RNAS*

Small nucleolar RNAs (snoRNA) are an abundant set of noncoding RNA that mainly guide chemical modification of ribosomal RNA. The two classes of snoRNAs are H/ACA and C/D

snoRNAs: each type has a particular sequence, structure, associated RNPs, and guide pseudo-uridylation (H/ACA) and 2' O methylation (C/D) of rRNA. Some snoRNAs chaperone rRNA and participate in the external and internal transcribed spacers (ETS and ITS) cleavage of the pre-rRNA<sup>7,12</sup>.

The H/ACA box snoRNAs structure contains two hairpins separated by a hinge holding the H motif (ANANNA, N=any base), with the ACA motif at their 3' end. The stem loop's bottom, embedded in the hairpins, interacts with the rRNA and sets the nucleoside to isomerize. The H/ACA snoRNPs are dyskerin (the pseudouridine synthase), GAR1 (nucleolar protein family A, member 1, NOLA1), NHP2 (NOLA2) and NOP10 (NOLA3). The C/D box snoRNAs hold the conserved boxes C (RUGAUGA, R=purine) and D (CUGA) located at their 5' and 3' end, respectively; the two motifs interaction forms a kink-turn structure with a terminal stem (~ 4 to 5 bp). Frequently, C/D snoRNAs contain incomplete C and D box copies, denominated C' and D' and located internally<sup>6,9,45</sup> (Figure 3).



Reproduced with permission from Matera and Terns<sup>12</sup>

### Figure 3. Structure of small nucleolar RNAs

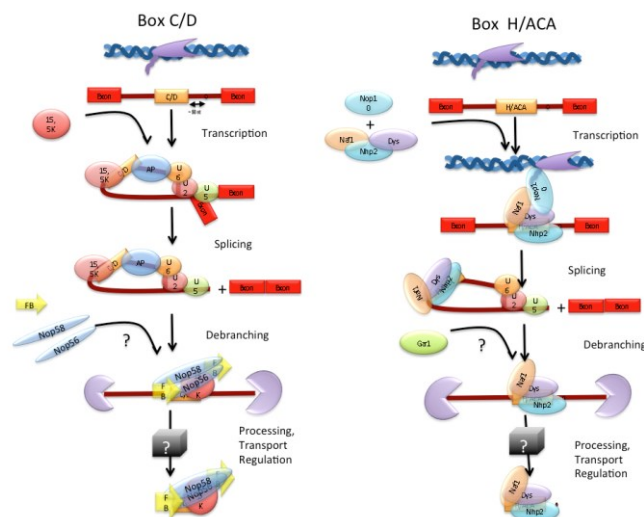
Figure depicts the anatomy of (A) C/D snoRNAs (B) H/ACA snoRNAs and U3 C/D snoRNA. It shows in yellow rectangles the sequence motifs, blue lines represent the snoRNAs while the red lines depict target RNA.

Looking in direction 5' to 3' the canonical snoRNA motif and guide sequences (gs) order are C - gs - D' - C'-gs-D. The 2'O-ribose methylation of target RNAs occurs precisely five nucleotides from the D or D' box. The C/D snoRNPs are Fibrillarin (FBL, the methyltransferase), NOP56 (or NOL5A, nucleolar protein 5A), NOP5/NOP58 and 15.5K<sup>6,9,45</sup>.

### 1.1.3.1 snoRNA biogenesis and control

snoRNAs evolved differently, generating several gene organization and transcription strategies. Overall snoRNAs could be intronic or not, polycistronic or monocistronic, and pol-II or pol-III dependent (For details see <sup>46</sup>). In humans, snoRNAs are primarily intronic (one snoRNA per intron), included mainly in 5' TOP transcripts that codify for ribosomal proteins, but can also be found in noncoding transcripts, i.e. Gas5 or SHG loci.

Pol II transcribes H/ACA snoRNAs, and the snoRNPs assembly occurs co-transcriptionally or early after transcription. Their hosting introns are longer, and the splicing and the snoRNP are separate events. After debranching and trimming the snoRNPs end up in the nucleolus <sup>45,47</sup> (Figure 4).



Adapted and modify from Figure 2 in Kiss <sup>45</sup>.

#### Figure 4. Biogenesis of intronic small nucleolar RNAs

Figure depicts snoRNA biogenesis and associated proteins; RNA is depicted in red and snoRNAs are the orange squares, the arrow describes the distance between snRrNA gene and the branching points (Black dot).

In contrast to H/ACA snoRNA, C/D snoRNAs require the splicing machinery for their correct processing. The majority of snoRNAs have a distance of approximately 50 nt between the snoRNA gene and their hosting intron's branching point. After the first catalytic step of splicing occurs and the complex C1 is formed (Lariat and 3' exon) <sup>48,49</sup>, the protein IBP160 couples the protein 15.5K with the intronic snoRNAs allowing a correct folding. Then the second catalytic step yields two components: the hosting snoRNA lariat and the exon-exon spliced mRNA. <sup>50</sup>.

Afterward, the debranching enzyme opens the lariat, and the exosome endonuclease DIS3 trims the snoRNAs, which requires the NEXT complex and probably occurs in the nucleoplasm. Then snoRNAs are transported into the nucleolus<sup>51</sup>. In yeast, the debranching step determines the snoRNA mature formation and function, but no evidence has been reported in the mammalian system<sup>52,53</sup>.

The mature snoRNA steady state levels are controlled in yeast by the TRAMP complex and the nuclear exosome<sup>54</sup>, while in humans, EXOC10 (the nucleolar exosomal catalytic subunit) regulates the mature snoRNA levels<sup>51</sup>. hMTR4 is the primary adaptor for the TRAMP and NEXT complex of the human nuclear RNA exosome<sup>55,56</sup>, in yeast mtr4 is the adaptor that controls mature snoRNA<sup>54</sup>. However the human TRAMP show, in contrast to yeast, differential association with exosomal factors in a subnuclear fashion<sup>57,58</sup>, suggesting a different evolutionary path.

It seems that the NEXT complex (hMTR4- ZCCHC8 and RBM7) adapts the processing of immature snoRNAs after debranching, while the TRAMP homologs (hMTR4, ZCCHC7, and hTRF4-2) regulate the mature form. Furthermore, DGCR8, part of the pri-miRNAs processing machinery, works as an adaptor of the exosome for mature snoRNAs in the nucleolus<sup>59</sup>. It will be interesting to see how the two exosomal adaptors hMTR4 and DGCR8 control snoRNAs mature steady state levels (See results 2.1.31 hMTR Knock-down).

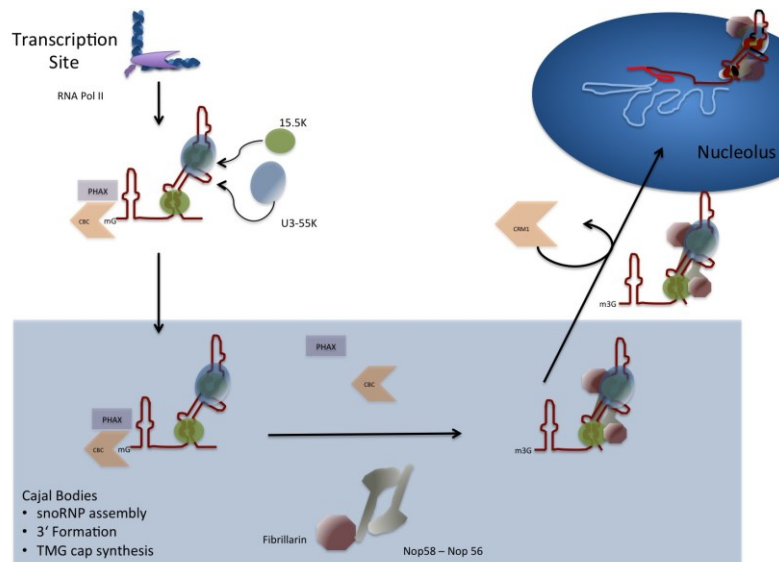
### 1.1.3.2 U3 typical non-intronic C/D snoRNAs

U3, U8, and U13 are not-intronic snoRNA transcribed by Polymerase II, and their biogenesis uses various trans acting factors. In contrast to other snoRNAs, these participate in rRNA processing rather than its methylation, and U3 processing and biogenesis are relatively well studied<sup>60-62</sup>.

U3 has two unique domain: the 5', that contains the conserved sequences: CAG, A' and A boxes, and the 5' and 3' hinges, the latter resembles other C/D snoRNAs carrying a C', C and D boxes (**Figure 3**). In contrast to other snoRNAs, U3 has a conserved motif B interacting with the C box and is necessary for the interaction with its unique protein U3-55K. The A and A' motifs base-pair with the 18S rRNA and the hinges interact with the pre-rRNA 5' ETS, which is necessary for the accurate pre-rRNA processing. The 3' domain interacts with the snoRNPs (Nop58, Nop56, FBL, 15.5K, and U3-55K), and it keeps the U3 located in the nucleolus. Mutations of these motifs result in U3 mislocation<sup>60-62</sup>.

Interestingly, regulation of U3-55K seems to control ribosome biogenesis; cell differentiation reduces U3-55k expression and snoRNA U3 levels, reducing pre-rRNA processing<sup>63</sup>. Recently, the Voit group showed that stalled ribosome biogenesis during stress induces U3-55K hyperacetylation, which lessens its interaction with U3, reducing pre-rRNA processing<sup>64</sup>.

After transcription, U3 interacts with CBC and PHAX, and like snoRNAs it has a trimethylated cap. But it appears that a cytoplasmic phase is not required because TGS1 (see snRNAs) holds two isoforms differentiated by an NLS<sup>65,66</sup>. For U3 the TGS1 nuclear isoform modifies the cap, while Exportin 1 transports it to the nucleolus<sup>60,67</sup>. Intriguingly, U3 can be located in the cytoplasm in starving cells, at the end of the cell cycle, and sometimes during biogenesis<sup>61,68,69</sup>. However, reviews skipped this conflicting evidence and settled that U3 has no cytoplasmic phase (For details I recommend the discussion from Watkins in U8 and U3 biogenesis publication<sup>69</sup>) (**Figure 5**).



**Figure 5. Biogenesis of U3 small nucleolar RNA**

U3 is transcribed by RNA polymerase II and undergoes through a set of multi-dynamic protein-RNA interactions in the nucleoplasm.

### 1.1.3.3 snoRNA sources of miRNA like molecules

*De novo* sequencing of ncRNAs displayed an accumulation of 19-40 nt RNA<sup>70</sup>, including segments from snoRNAs and tRNAs, and several clinical studies show differential processing of those derived miRNAs. However, an agreement for functionality is unclear and contradictory. Below I will describe the experimental evidence for each proposed feature and the conceptual holes that require empirical testing and are the center of this dissertation.

In 2008, Miester's lab Ago RIP-seq experiments showed that the H/ACA snoRNA 45 produces miRNAs in a Drosha-independent manner but require Dicer activity <sup>4</sup>. Then, Michelle Scott described similar processing for C/D and H/ACA box and how annotated miRNAs show snoRNA motifs, always using computational analysis, next generation sequencing and northern blots, without any functional read-outs <sup>9,71-73</sup>.

In 2011, Brameier et al., described the processing of C/D snoRNAs resembling guide and passenger strand distribution of miRNAs. They included several human cell lines and used luciferase assays that showed modulation upon insertion of miRNA-like complementarity<sup>74</sup>, However, no deeper statistical analyses or confirmation tests were applied (See Figure 3 from<sup>74</sup>). For that reason, we considered studying those C/D snoRNA-derived miRNA using a systematic approach (For the definition of systematicity I recommend Paul Hoyningen-Huene's paper <sup>75</sup>).

Many reviews described as interesting that snoRNAs could generate miRNA-like molecules, defining as priority the need of experimental validation and advice to keep caution working with the group VI miRNAs (see 1.1.2.1) especially, due to their low conservation and low expression <sup>6-11</sup>.

#### **1.1.3.4 snoRNA novel functions**

Several studies proposed that snoRNA-derived miRNA molecules can control alternative splicing. Still, it is not clear whether the mature or tiny forms perform this activity <sup>10,11,76</sup>. Alternatively, other sources proposed that snoRNA-derived miRNAs work together with Piwi proteins and control transcription epigenetically <sup>77,78</sup>.

Finally, sequencing of RNA-RNA interactions show that snoRNAs control mRNA steady state demonstrating that snoRNA can control coding genes expression <sup>79</sup>. Scott and colleagues noted similar effect for exogenous snoRNAs expression vectors, where replacing sequences between the C and D boxes reduced the expression of complementary mRNA <sup>80</sup>.

## AIMS OF THE THESIS

To rule out that C/D snoRNA-derived miRNAs show functional output as canonical miRNAs, I reasoned that a deductive approach, where I look for disproving rather than confirming a hypothesis, might generate better evidence.

As introduced before, the current paradigm for canonical miRNAs predicts several structural, biochemical and functional behaviors. I decided to use these predictions for miRNAs as references and controls to show whether C/D snoRNA-derived miRNAs can be ruled out as canonical miRNAs.

To characterize structurally and biochemically C/D snoRNA-derived miRNAs, the current paradigm predicts that they should have a size of 19 to 22, they should show a consistent processing pattern - a defined start and end on their genomic locus, and they should interact with Argonaute proteins, as well as Dicer and the microprocessor substrates.

From the functional perspective, they should downregulate any mRNA that contains its target sequence, have endogenous regulated genes, and functional relation between them. For instance, a neuron-expressed miRNA has targets that inhibit glial phenotypes and favors neuron-related gene expression.

Experimental biochemical and bioinformatic advances allow in-depth miRNA characterization regarding expression and structure, but functional assays still have low resolution, are poorly developed and do not allow *in vivo* miRNA activity quantification.

To tackle these two issues, C/D snoRNA-derived miRNA, and low *in vivo* resolution, I used high-throughput data, mainly NGS from small RNA, RIP and CLIP datasets, to evaluate structurally and biochemically miRNAs and applied it to snoRNAs. Second, I generated, tested and validated a high-resolution tool for miRNA activity quantification *in vivo*, and used it for C/D snoRNA-derived miRNAs. Finally, I used CLASH-acquired data to sort out endogenous targets of putative snoRNA-derived miRNA.



## 2 MATERIALS & METHODS


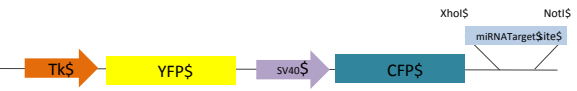

All molecular, microbiological and cell biology methods were performed, if not otherwise stated, as described in Current Protocols in Molecular Biology<sup>81</sup>.

### MOLECULAR BIOLOGY

#### 2.1.1 MIRNA DUAL FLUORESCENT REPORTER GENERATION

Plasmids were constructed using the PsiCheck2 (Promega, Wisconsin, USA) dual luciferase reporter backbone: promoters, XhoI and NotI sites were kept, and fluorescent proteins (GFP, RFP or CFP) replaced the luciferase genes. Small RNA sequences were derived from miRBase (www.mirbase.org) and Oligonucleotides (Sigma-Aldrich, Munich Germany) were designed with NotI and XhoI overhangs annealed and ligated into fluorescence plasmid (**Supplementary Table 4**), Table 1 summarized the step of the produced the plasmids.

**Table 1.** Generation of plasmids for Analytical Flow cytometry.

miRNA sensing plasmid	Cloning Strategy
<p><b>UTA-1</b></p> 	<p>1. GFP from pAC-GFP-N + psiCHECK-2 backbone (ApaI-XbaI). 2. Residual hluc+codon removal by PCR. 3. CFP from pcDNA-CFP replacing hRluc (NheI+AsiSI)</p>
<p><b>UTA-2</b></p> 	<p>1. PCR YFP extraction pLenti-CaMKIIa-eArchT 3.0-EYFP PCR (Addgene Plasmid #35513) + UTA1.0 (ApaI and XbaI).</p>
<p><b>UTA-3</b></p> 	<p>1. RFP pLemiR UTA1.0 (NheI+AsiSI)</p>

Three different constructs were generated using the psiCHECK-2 backbone and replacing the luciferase with fluorescent proteins.

Annealing reactions were performed in 10 uL final volume including 1 uL of each strand (100uM), 1 uL T4 ligase buffer (New England Biolabs, MA, USA) and 0.5 ul T4 polynucleotide

kinase (New England Biolabs, MA, USA); then incubated at 37°C for 1 h and cooled down from 95 to 25°C at 5°C/min. The aligned oligonucleotides were ligated to an empty fluorescent vector digested with XhoI and SacI (New England Biolabs, MA, USA) for 2 h or overnight at 16 °C then transform into *E. coli* Top10. Used oligos are listed in **Supplementary Table 4**.

Plasmids are public available In Addgene <https://www.addgene.org/browse/article/22044/>

### 2.1.2 *GAS5 CONSTRUCT GENERATION*

Hek293 cells genomic DNA was isolated using DNeasy Blood & Tissue Kits (Qiagen, Germany) and used as a template for GAS5 PCR amplification using Phusion polymerase (New England Biolabs, MA, USA), PCR reactions were performed using 3 % DMSO and two steps PCR. PCR products were A tailed using Taq polymerase and 2 mM ATP for 25 min at 72 °C, and inserted into pGEM®-T Easy and pTARGET™(Promega, Wisconsin, USA) following the manufacturer's instructions; however, the ligated plasmid were transformed in Top10 instead of JM109. To generate pAcGAS5-GFP construct, pAcGFP1-N1 (Takara Bio USA, CA, USA) and pGEM-Gas5 were digested with Not I, ligated with T4 ligase (New England Biolabs, MA, USA) and transformed into Top10 *E.coli* strains. For primers check Table 2.

### 2.1.3 *BACTERIAL TRANSFORMATION AND COMPETENT CELLS GENERATION AND PLASMID ISOLATION*

Competent *E.coli* Top10 or Stbl3 were prepared using the Mix & Go *E. coli* transformation buffer set from Zymo (Zymo Research Corp. CA, USA), and transformations were performed according to manufacturer's instruction. Bacteria cultures were kept at 37 °C on LB-plates or in LB-media. Plasmids were isolated using miniprep isolation custom protocol while for transfections plasmids were purified using the QIAquick Plasmid Midi Kit (Qiagen, Germany). Using the following Antibiotic concentrations: Ampicillin, 100 µg/mL (H<sub>2</sub>O) (Carl Roth, Germany) and Kanamycin 50 µg/mL (H<sub>2</sub>O) (Applichem, Germany).

### 2.1.4 *POLYMERASE CHAIN REACTION AND RESTRICTION ENDONUCLEASE REACTIONS*

Amplification of specific DNA fragments was achieved by polymerase chain reaction using Taq (Biotherm™, Germany) or Phusion polymerase (New England Biolabs, MA, USA). According to manufacturer's instructions. Results were analyzed by DNA electrophoresis in 1% - 2 % Agarose (w/v), standard electrophoresis was done at 120 V for 45 min in 1X TAE buffer. For purification 1/4 of the PCR reaction was analysed by gel electrophoresis and the remaining 3/4 purified using QIAquick PCR purification kit. Used Primers are listed in Table 2

**Table 2.** Standard PCR Primers

<i>Primer Name</i>	<i>Sequences 5'-3'</i>
Gas1-Exon-1 FWD	<i>agtcgactcctgtgaggtatg</i>
Gas5-Exon-6 FWD	<i>aacttgctggaccagctta</i>
Gas5-Exon-11 FWD	<i>gcatgcagcttactgcttgaa</i>
Gas5 Genom FWD	<i>cacactccagatggattgcaa</i>
Gas5 Genom RV	<i>cccaggcaccatactgtct</i>
GFP FWD	<i>ctgatcgagctgaatggcga</i>
GFP REV	<i>catcagcgatcttggtggtt</i>
ECFP-FWD	<i>ctaccccgaccacatgaagc</i>
ECFP-RV	<i>tattttattgcgccagcgg</i>
Luc Out FWD	<i>ctaagaacattaagaagg</i>
Luc Out RV	<i>ggctttaccaacagtacc</i>
YFP FWD	<i>gggcccatggtgagcaagggcgaggag</i>
YFP RV	<i>tctagatcactgtacagctcgtccatgc</i>
IncGas5 FWD	<i>gagctctttcgaggtaggagtcgac</i>
IncGas5 RV	<i>ccgcggttgattgcaaaaatttatta</i>

Restriction endonuclease reactions were performed according to NEB instructions, Briefly, restrictions were performed for 2 h at 37 °C. PCR products or desired endonuclease fragments were separated by agarose gel electrophoresis and the desired product was purified using the QIAquick Gel extraction Kit (Qiagen, Germany) according to manufacturer's directions. DNA concentration was measured using Synergy 2 System (Biotek, Winooski, USA) or NanoDrop™ 2000 (Thermo Fisher Scientific, CA, USA)

PCR products were separated by agarose gel electrophoreses as described in 2.1.4 Size separated products were analyzed under UV light and desired product was cut. Excised DNA was purified by QIAquick gel extraction kit (Qiagen, Germany). After purification DNA was analyzed by Synergy 2 System (Biotek, VT, USA) and 1200 ng of plasmid DNA was diluted in 15 µL nuclease free water and send for Sanger sequencing in SEQLAB Sequence Laboratories GmbH, Göttingen, Germany or Macrogen, Inc. Seoul, South Korea.

#### 2.1.5 RNA ISOLATION, SIZE AND INTEGRITY EVALUATION

Total RNA was isolated using Phenol/Chloroform extraction according to producer's instructions (Trizol, Thermo Fisher Scientific, CA, USA). RNA was measured by Synergy system (BioTek, VT, USA) or NanoDrop™ 2000 (Thermo Fisher Scientific, CA, USA). RNA amount, size distribution

and integrity number (RIN) were analysed on a BioAnalyzer System (Agilent, CA, USA) via RNA 6000 Pico Kit (Agilent, CA, USA).

### 2.1.6 cDNA SYNTHESIS AND QUANTITATIVE PCR ANALYSIS

Total RNA (1000 ng) was reverse transcribed using Sensifast cDNA Synthesis Kit (Bioline, London, UK) and diluted (1:5 or 1:10). After reverse transcription 25 ng of cDNA were used for quantitative analysis on ABI StepOnePlus system (Applied Biosystems, CA, USA). Relative expression was calculated via  $\Delta\Delta C_T$ -method using GAPDH as housekeeping gene; primers used for quantification are listed in **Table 3**.

**Table 3** qPCR Primers

<i>Primer Name</i>	<i>Sequences 5'-3'</i>	<i>Primer Name</i>	<i>Sequences 5'-3'</i>
Gas-5-FW	<i>agctggaagtgaatgg</i>	40-9 RV	<i>ttgacaccgctaacttat</i>
Gas-5-RV-1	<i>caagccgactctccatacc</i>	40-13 RV	<i>cattttgaaggaacggtag</i>
Snord75 FWD	<i>cctgtgatgctttaagag</i>	40-14 RV	<i>accacctaagtactccagcaa</i>
Snord 75 RV	<i>cctcagaatagaatttcag</i>	40-16RV	<i>ctggtcataccctggattta</i>
40-9 FWD	<i>ggcatctggactagaatgaa</i>	40-18RV	<i>gcaaagttggagacaagaac</i>
40-13 FWD	<i>ggaaatagtgagaaaagca</i>	40-31RV	<i>atgtagtgacacctgcaca</i>
40-14 FWD	<i>taacacactttcccagac</i>	40-38RV	<i>cattctggatgcatgtgtag</i>
40-16 FWD	<i>atccatcatctcaccacatt</i>	40-52RV	<i>ggtcgtttgagtggactaac</i>
40-18 FWD	<i>tttacctccctctctgttca</i>	hMTR4 FWD	<i>tccactgagcagcttatccac</i>
40-31 FWD	<i>tgacaatttcacacctaga</i>	hMTR4 RV	<i>aaagtagtcgagtcgcgag</i>
40-38 FWD	<i>ctactgatcacaccctctc</i>	40-2b FWD	<i>gggagctcaaggaaaaggag</i>
40-52 FWD	<i>agttccaagaaccacacac</i>	40-2b RV	<i>cagtgaaggagagcgtatc</i>
GAPDH FWD	<i>gtcagccgcatcttcttttg</i>	40-33FWD	<i>ctggcctagctaaagtctca</i>
GAPDH RV	<i>gcgcccaatacgaccaaadc</i>	40-33 RV	<i>tctgtcctagctctcagtc</i>

### 2.1.7 NGS LIBRARY PREPARATION

Total Isolated RNA was analysed for quality and quantity using BioAnalyzer System (BioAnalyzer RNA Pico Kit, Agilent, CA, USA). Total RNA (2 µg) was purified with the magnetic bead clean-up module (Life Technologies, CA, USA) for small RNAs. Small RNA was ligated with sequencing adapters, reverse transcribed and purified (Ion Total RNA-Seq Kit v2, Life Technologies, CA, USA). Finally, barcodes were added, and library size and amount were detected via BioAnalyzer HS Chip (Agilent, Santa Clara, USA).

### NEXT GENERATION SEQUENCING

Libraries were diluted (18pM) and clonal amplified by emulsion PCR on Ion-Torrent OneTouch System using the on PGM Template OT2 200 Kit following producer protocol (Life Technologies, CA, USA). Amplified libraries were purified using on OneTouch ES System, and loaded into Ion-Torrent 316 Chip and sequenced via Ion-Torrent PGM System.

## CELL BIOLOGY

### 2.1.8 CELL CULTURE

All cell lines were maintained according to ATCC instructions in humidified environment at 37°C and 5% CO<sub>2</sub>. HeLa cells were cultured in RPMI (Roswell Park Memorial Institute) 1640 supplemented with 10% FCS, whereas DMEM (Dulbecco's modified Eagle's Medium) with 15% FCS was used for Hek293 cells.

### 2.1.9 DNA AND siRNA TRANSFECTION.

Hek293 cells (100,000 cells/well in 24-well) were transfected using Lipofectamine 2000 (Thermofisher, MA, USA). Typical transfection was performed with 1 µg plasmidic DNA per well, 2 µl transfection reagent per 1 µg of DNA in 100 µl OptiMEM incubated for 20 min; then the mixture was added to the cells and incubated for 72h at 37°C and 5% CO<sub>2</sub>. For co-transfection experiments plasmid and siRNA (GL2 or LMNA) (Dharmacon, CO, USA) were transfected into Hek293 cells (100,000 cells/well in 24 well plate) with Lipofectamine 2000, for out-competing experiments, 20 nM siRNA were added to 1 µg plasmidic DNA with 2 µl Lipofectamine.

siRNA single transfections were performed using 200 nM siRNA and 1 µl lipofectamine, incubated for 72 h and RNA was isolated; for double siRNA transfection cells were transfected again after 48 h and incubated for two more days previously to RNA isolation.

**Table 4** siRNA used in this work.

siRNA	Catalog Dharmacon
GL-2	D-001100-01-20
LMNA	L-004978-00-000
hMTR4	J-031902-09-0002

### 2.1.10 LUCIFERASE REPORTER ASSAYS

Hek293 cells were transfected as described in 2.1.9 and Luciferase activities were determined using the Dual-Glo<sup>®</sup> Luciferase Assay (Promega, WI, USA) in accordance to manufacturer's

protocols, luminescence was assayed on a Synergy 2 microplate reader (Biotek, VT, USA). Luminescence intensities for the Renilla luciferase ( $R_{luc}$ ) were normalized against the simultaneously examined firefly luciferase ( $F_{luc}$ ). Normalized  $R_{luc} / F_{luc}$  values were calculated according to manufacturer's guidelines.

#### 2.1.11 FLUORESCENCE MICROSCOPY AND FLOW CYTOMETRY

Transfected Hek293 cells were evaluated using fluorescence microscopy; pictures were taken using an Axio Observer microscope (Zeiss, Jena, Germany). Exposure time was 90 ms for GFP (Filter BP 525/50) and 900 ms for RFP (Filter BP 605/70).

Cells were washed three times with cold PBS, fixed in 2% paraformaldehyde for 45 min and washed again in cold PBS and re-suspended in 200  $\mu$ L PBS; cells were read afterwards or kept at 4 °C for a maximum of 72 h before FACS analysis.

Flow cytometry was performed using a BD LSR II instrument (BD, New Jersey, USA). The filters employed 550LP- BP575/26 and BP450/50 for YFP and CFP respectively. YFP positive cells were selected using the FACSdiva™ software and FCS files were exported for analysis in R (See Bioinformatics).

### PROTEIN AND RNA BIOCHEMISTRY

#### 2.1.12 WESTERN BLOT

Cells were washed twice with cold PBS and resuspended in 150  $\mu$ L RIPA buffer. Then 50  $\mu$ L TCA added (50 % w/v TCA) and incubated 10 minutes on ice and centrifuged at 13.000 rpm for 5 minutes at 4 °C and the pellet was washed with 4 volumes cold acetone and centrifuged. The pellet was dried out for 10 min at 37 °C and resuspended in Laemmli buffer<sup>82</sup> and incubated at 95 °C for 10 min and loaded on SDS gel.

Protein separation was performed on sodium dodecyl sulfate (SDS-) gel electrophoresis<sup>83</sup> and gels were prepared in Novex™ cassettes according to the following standard recipe:

**Table 5.** Recipe for separation gel

Component	Amount (10% gels)	Amount (12.5 % gels)
H <sub>2</sub> O	2,5 mL	2.19 mL
40% Acrylamide/bis Acrylamid	1.25 mL	1.56 mL
1,5 M Tris, pH 8,8	1.25 mL	1.25 mL
10% ammonium persulfate (APS)	50 $\mu$ L	50 $\mu$ L
TEMED	5 $\mu$ L	5 $\mu$ L

**Table 6.** Stacking gel recipe

Component	Amount
H <sub>2</sub> O	3.13 mL
40% Acrylamide/bis Acrylamid	0.62 mL
1,5 M Tris, pH 6,8	1.25 mL
10% ammoniumpersulfate (APS)	50 µL
TEMED	5 µL

The used ladders were Page ruler pre-stained Protein Ladder (Fermentas, MA, USA) or Magic Mark XP protein Ladder (Thermo Fischer, MA, USA), Electrophoresis was performed at 35 mA for 2 h. Afterward proteins were transferred onto a nitrocellulose membrane using the wet system XCell II Blot Module (Thermo Fischer, MA, USA) using 3 Whatman papers and 2 sponges soaked in transfer buffer for 1 h at 25 V. Transfer was checked using Panceau red and membrane was blocked with NETT-G buffer for 1h at room temperature.

Incubated with primary antibody over night at 4 °C and next day membranes were washed three times with TBS-T and incubated with HRP coupled anti-Rabbit (G-21234 Life Technologies, Carlsbad, USA) diluted 1:10.000 in TBS-T for 2h at room temperature and then washed three times with TBS-T. The chemiluminescence signal was developed with HRP substrate (Luminata Forte, Merck, Germany) and detected by ECL machine (Chemocam Imager, INTAS, Germany). Primary antibodies used were anti-hMTR4 (ab70551 Abcam, Cambridge, UK) at dilution 1:5000 and anti-FBL (EPR10823(B)Abcam, Cambridge, UK ) at a dilution 1:5000.

#### 2.1.13 NORTHERN BLOT

RNA was diluted in 2X Novex TM TBE-Urea Sample Buffer (Thermo Fisher Scientific, MA, USA) and incubated for 5 min at 95 °C. Following separation in 15% Polyacrylamide/Urea gel Electrophoresis at 180 V for 75 min using the XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific, MA, USA) in TBE buffer. Transfer into Bright Start Nylon membrane was performed using the wet system XCell II Blot Module (Thermo Fischer, MA, USA) at 25 V for 30 minutes in 0.5X TBE buffer and the RNA was crosslinked for 3 minutes at 302 nm UV transilluminator.

The hybridization and detection were performed using the North2South® Chemiluminescent hybridization and detection kit following manufacturer's instructions. 1 nM of custom prepared LNA biotin labeled probe (Exiqon, (Qiagen, Germany)) at 45 °C in a hybridization incubation

chamber and the signal was detected by ECL machine (Chemocam Imager, INTAS, Germany). The target U78 sequence used was UGAGCAUGUAGACAAAGGUAACACUGAA.

#### BIOINFORMATICS AND STATISTICAL ANALYSES

All the bioinformatics and statistical analysis were performed using Biolinux open source<sup>84</sup>, R environment with R studio<sup>85,86</sup> and Bioconductor<sup>87</sup>.

#### 2.1.14 NEXT GENERATION SEQUENCING DATA ANALYSIS

##### 2.1.14.1 Ion torrent Data Analysis

Raw reads were trimmed and quality filtered employing the Ion -Torrent Suite Software (Life Technologies, CA, USA). FASTQ outputs from the suite were changed to FASTA format and collapsed using FASTX-toolkit (<http://hannonlab.cshl.edu/>); next collapsed FASTA files were aligned to the miRbase v 21<sup>88</sup> and snoRNA base v3<sup>89</sup> with Bowtie2<sup>90</sup> employing the following command line arguments:

```
-D 15 -R 3 -N 0 -L 16 -k 20 --local -i S,1,0.50 --score-min L,18,0 --ma 1 --np 0 --mp 2,2 --rdg 5,1 --rfg 5,1
```

The .fa format files from miRbase and snoRNA base v3 were used to create a mapping database with the command bowtie2-build. Then SAM outputs were changed to Blast format using *sam2blast*, later significant hits by E-value were selected using *mtophits\_blast* and duplicates were removed with *remove\_duplicate\_hits\_blast.pl*, scripts were adapted or modified from Grzegorz Kudla (<https://github.com/gkudla>)<sup>91</sup>.

Results files (Called here Filtered-Blast) were further analyzed by custom made R-scripts and normalization was performed using DEseq package<sup>85,92</sup>.

Alignment to the reference genome (Hg19) was performed using the Ion -Torrent Suite Software and BAM files were employed for genomic track analyses in IGV genome browser<sup>93</sup>.

##### 2.1.14.2 GEO datasets analysis

The GEO data were retrieved using the SRA-toolkit<sup>94</sup> and transformed from binaries to FASTQ using the *fastq-dump* tool, all the downloaded libraries were analyzed with FASTQC and just data that fit the original publication description were kept. According to the platform and the library preparation demultiplexing and adaptor trim were applied using *Flexbar*<sup>95</sup> or *FASTX-toolkit* (<http://hannonlab.cshl.edu/>), for paired end data the two FASTQ files were merged using PEAR<sup>96</sup> with the default command line. The data set analyzed were:



**Table 7.** GEO retrived raw data.

Description	GEO accession number	Reference	Comment
Hela cellular fractionation	GSE50057	( <sup>97</sup> ) Bai et al.	Ion Torrent library
snoRNPs Hek293	GSE43666	( <sup>98</sup> ) Kishore et al.	PAR-CLIP endogenous NOP56, Fibrillarin, Dyskerin and FLAG-NOP58. Size selection of small RNAs
DGCR8 Hek293	GSE39086	( <sup>99</sup> ) Macias et al.	HITS-CLIP FLAG and endogenous DGCR8
DICER Hek293	GSE55333	( <sup>100</sup> ) Rybak-Wolf et al.	PAR-CLIP FLAG-Dicer1
Argonaute Hek293	GSE21918	( <sup>101</sup> ) Hafner et al.	PAR-CLIP FLAG-AGO 1 to 4
	GSE50452	(30) Helwak et al	CLASH Libraries E1, E4 and E7
	GSE58127	( <sup>102</sup> ) Flores et al.	RIP and Cellular small RNAs. RIP seq of 2 antibodies against Argonaute

Filtered and demultiplexed Fastq files were processed as Ion Torrent Fastq files see 2.1.14.1.

#### 2.1.15 GRAPHS GENERATION

All the graphs were performed in the R environment, if is not otherwise stated, using the default graphics and the ggplot2 package<sup>97,103</sup>

##### 2.1.15.1 Fraction calculation of mature vs small snoRNAs

To compare RNA purification methods, 20-200 aligned reads ( in blast format) were subsetted according to read length, then separated between (mature and processed) using the threshold of 40 nt. Finally, they were compared with the Ion-Torrent libraries.

##### 2.1.15.2 Hierarchical clustering from binary matrices

To define processing pattern filtered blast files were refined for individual snoRNAs or miRNAs. The reported mature or pre-miRNA length was used as column number while each read corresponds to a row, forming a matrix of nucleotides as column and reads as rows. For each base on the read that matches a position on the transcript; the value 1 was given, while no match was set to 0. Creating a binary matrix with columns showing each base on the transcript, while rows represent the reads with 1 for bases presence and 0 for absent bases.

For more than 100.000 reads the total read number was simplified by calculating fraction of the total amount, and multiplied by a 10 multiplier

to get a number greater than 1 and transformed into an integer before generating the binary matrix.

The matrix was organized using hierarchical clustering of rows using the *hclust* command of basic R and plot using the *Heatmap2* command from gplot package<sup>104</sup>

### 2.1.15.3 Density plots

To define the density of reads over the transcripts similarly to Grannemann on CRAC analysis<sup>105</sup> Blast files were filtered for individual snoRNAs or miRNAs and transcript size was used as vector length. The sum per the column of binary matrices was calculated and replaced in the density vector, later the fraction of total number of reads was calculated. Tidy data was generated for each library and RNA, then plotted using ggplots2<sup>103</sup>

### 2.1.15.4 snoRNA chimera calls from CLASH data set

To call chimeras from CLASH Ago1 data the Hyb pipeline was implemented in Biolinux as recommended by the Authors<sup>91</sup>, snoRNA based chimeras were called using the following command:

```
hyb analyse in=data_comp_hOH7.blast format=blast align=none type=sno pref=none hval=0.1 hmax=10 gmax=4
```

Blast files were produced as described for Ion-torrent using the default data base in the Hyb pipeline for the GSE50452 (libraries E1,E4 and E7). Then the U3 chimeras were collected for the three libraries and checked manually.

## 2.1.16 ANALYTICAL FLOW CYTOMETRY

### 2.1.16.1 Simulations and parameters estimation

For small RNA induced regulation in the bacterial system, a randomly generated  $X$  values were used as input in a custom wrote R function using Equation 4 and parameters  $Fr$  and  $\phi$  were arbitrary selected. For miRNA simulation similar approaches was performed but using Equation 7, then siRNA experiments or endogenous miRNA on Hek293 cells were used to estimate  $\lambda$  and  $\theta$  by nonlinear (weighted) least squares using Equation 7. For detailed equation explanations and descriptions see 2.1.26.2

Then a new set of simulations were performed using the range of YFP intensities were used for simulations; the simulations can be reproduced at <https://medicalrnabiology.shinyapps.io/pUTApp/>



**Figure 6** QR code to replicate on line numerical simulation for miRNA titration model.

### 2.1.16.2 Analytical function generation

Positive YFP cells data was extracted on FCS 2.0 or 3.0 format and transferred into the R environment using the FlowCore Bioconductor package<sup>87,106,107</sup>. Analytical functions generation from raw data was generated similarly to<sup>31,44,108,109</sup>. YFP Log transformed relative intensities were binned at 0.05 interval, then descriptive statistical values for relative log CFP intensities were calculated, all the plots represented here show the average CFP intensity per bin (**Figure 32**)

### 2.1.17 OTHER STATISTICAL ANALYSIS

Some Non parametric ANOVA analysis, Dunn`s multiple comparison tests and linear analysis were performed using the PRISM software (GraphPad, CA, USA)

## MATERIALS

### 2.1.18 MACHINES

**Table 8: Used machines and manufacturer information**

Machine	Model	Company
Balance	EG 620-3NM	Kern & Sohn GmbH, Germany
Chemiluminescence System	ChemoCam Imager	INTAS, Germany
Centrifuge	Heraeus Megafuge 8R	Thermo Fisher Scientific, MA, USA
Centrifuge	Heraeus Fresco21	Thermo Fisher Scientific, MA, USA
Centrifuge	Avanti J-30I	Beckman Coulter, CA, USA
E-Gel system	E-Gel iBase System	Life Technologies, CA, USA
Fluorometer	Qubit 2.0	Life Technologies, CA, USA
Freezer (-20°C)		Bosch, Germany
Freezer (-80°C)	MDF-DU500VH-PE	Panasonic, Japan
Freezer (-150°C)	ULT7150-9-D	Thermo Fisher Scientific, MA, USA
Flow Cytometer	LSR II Flow Cytometer	BD Bioscience, NJ, USA
Gel documentation system	Gel iX Imager	INTAS, Germany
Hamilton pipette	Microliter™ Syringes	Hamilton, NV, USA

Ice machine	ZBE 70-35	Ziegra, Germany
Incubator	Heracell VIOS 160i	Thermo Fisher Scientific, MA, USA
Incubator	IN75	Memmert, Germany
Incubator	Ecotron	Infors HT, Switzerland
Hybridization chamber	MS Hybridization oven	Major Science Co.Taiwan
Ion Torrent NGS	PGM	Life Technologies, CA, USA
Ion Torrent OneTouch2	OneTouch2™	Life Technologies, CA, USA
Ion Torrent OneTouch ES	OneTouch ES	Life Technologies, CA, USA
Microscope	Axio Vert. A1	Zeiss, Germany
PCR Thermocycler	Labcycler	Sensoquest, Germany
PCR Thermocycler	2720 Thermal cycler	Applied Biosystems, MA, USA
pH meter	SevenCompact™ S210	Mettler Toledo, OH, USA
Plate Reader	Synergy 2	BioTek, VT, USA
Platform Rocker	PMR-30	Grant bio, Cambridge, UK
Power supply	EV231	Consort bvba, Belgium
Safety Cabinet	Safe 2020	Thermo Fisher Scientific, MA, USA
Thermoblock	ThermoStat plus	Eppendorf, Germany
Thermoblock	CTM	HTA-BioTec, Germany
Vortexer	Vortex-Genie™ 2	Scientific Industries, NY, USA
Water bath	WNB10	Memmert, Germany
Western Blot chamber	XCell SureLock® MiniCell	Life Technologies, CA, USA

### 2.1.19 BUFFERS AND SOLUTIONS

Table 9: Information about buffers and solutions

Buffer and solutions	Recipe
Electrophoresis buffer (SDS-PAGE) (10x)	250 mM Tris 1,9 M Glycin 1 % SDS
Laemmli buffer	0,125 M Tris-HCl pH 6,8 4 % SDS 20 % Glycerol 10 % $\beta$ -Mercaptoethanol 0,004 % Bromphenolblau
LB media	Yeast extract (5 g/L) Trypton (10 g/L) NaCl (5 g/L)
RIPA buffer	25 mM Tris-HCl pH 7,6 150 mM NaCl 1 % NP-40 1 % Natrium-Deoxycholat 0,1 % SDS
TAE buffer (Agarose gel electrophoreses)	40 mM Tris-HCl 20 mM acetic acid 1 mM EDTA pH 7,0
TBS	500 mM Tris 1,5 M NaCl pH 7,6
TBS-T	500 mM Tris 1,5 M NaCl 0,05 % Tween20
Western blot blocking buffer (NETT-G)	1 L NETT (1 x) 0,25 % Gelatine

Western blot transferbuffer	12 mM Tris 96 mM Glycin 20 % Methanol
Western blot wash buffer (NETT, 10x)	1,5 M NaCl 0,05 M EDTA pH 8,0 0,5 M Tris-HCl pH 7,5 0,5 % Triton X-100
Northern blot (TBE)	89 mM Tris base 89 mM Boric acid 2 mM EDTA (Free acid) pH 8.3

Buffers and solution non-stated in methods.

## 2.1.20 CHEMICALS AND MATERIALS

**Table 10.** Chemical and materials

Compound	Provider Information	Compound	Provider Information
0,45 µm filter	Sigma Aldrich, MO, USA	Hydrochloric acid(37%)	Carl Roth, Germany
1,4-Dithiothreitol	Carl Roth, Germany	Hydrogenperoxid Urea	Merck, Germany
24-well plate	Greiner, Germany	Imidazol	Sigma Aldrich, MO, USA
2-propanol	Merck-Millipore, Germany	IPTG	Carl Roth, Germany
6-well plate	Greiner, Germany	Isopropanol	Carl Roth, Germany
96-well plate	Greiner, Germany	Kanamycin	Applichem, MO, USA
Acetic acid	Carl Roth, Germany	L-Ascorbic acid	Sigma Aldrich, MO, USA
Aceton	Carl Roth, Germany	Kanamycin	Applichem, MO, USA
Acrylamid (40%) Bis solution	Biorad Ca, USA	L-Ascorbic acid	Sigma Aldrich, MO, USA
Acrylamid solution	Carl Roth, Germany	Loading Dye (DNA)	Thermo Fischer, MA, USA
Agar-Agar	Carl Roth, Germany	MagicMark™ XP Protein Marker	Invitrogen, CA, USA
Agarose	Biozym, Austria	Methanol	Carl Roth, Germany
Agarose (LE Agarose)	Carl Roth, Germany	Methylenblue	Carl Roth, Germany
Ammonium persulfate (APS)	Sigma Aldrich, MO, USA	MOPS	Carl Roth, Germany
Ammoniumchloride	Carl Roth, Germany	Nitrocellulose membrane	Macherey-Nagel, Germany
Ammoniumperoxodisulfat (APS)	Carl Roth, Germany	Paraformaldehyde	Carl Roth, Germany
Ammoniumsulfate	Carl Roth, Germany	PBS-Dulbecco	PAA, MA, USA
Ampicillin	Carl Roth, Germany	Penicillin/streptomycin	PAA, MA, USA
β-Mercaptoethanol	Carl Roth, Germany	Petri dishes	Greiner Bio-one, Austria
Biotherm Taq	Genecraft, Germany	PhenolRed	Carl Roth, Germany
Biotherm Taq 10x buffer	Genecraft, Germany	Pipette tips 10µL	Eppendorf, Germany
Bordeaux Red	Sigma Aldrich, MO, USA	Pipette tips 1000 µL	Eppendorf, Germany
Boric acid	Sigma Aldrich, MO, USA	Pipette tips 200 µL	Eppendorf, Germany
Bromphenol Sodium salt	Carl Roth, Germany	Ponceau Red	Sigma Aldrich, MO, USA
Calciumchloride (Cellpure)	Carl Roth, Germany	Potassium hydroxyide	Carl Roth, Germany
Chloroform	Merck Millipore, Germany	Potassium sodium tartrate tetrahydrate	Sigma Aldrich, MO, USA
		Potassiumchloride	Carl Roth, Germany

<b>Complete Mini</b>	Roche, Switzerland	<b>Potassium-di-Hydrogenphosphate</b>	Carl Roth, Germany
<b>Cryokonservation tubes</b>	Sarstedt, Germany	<b>Reaction tube 1,5 mL</b>	Eppendorf, Germany
<b>di Sodiumhydrogenphosphate</b>	Carl Roth, Germany	<b>Reaction tube 2 mL</b>	Eppendorf, Germany
<b>Diethyl pyrocarbonate</b>	Sigma Aldrich, MO, USA	<b>RNA later solution</b>	Ambion, MA, USA
<b>Dimethylsulfoxid</b>	Roth, Germany	<b>RNase</b>	Sigma Aldrich, MO, USA
<b>diPotassiumhydrogenphosphate</b>	Carl Roth, Germany	<b>SDS-Pellets</b>	Carl Roth, Germany
<b>D-L-Lysine</b>	Sigma Aldrich, MO, USA	<b>Serological pipette 10 mL</b>	Sarstedt, Germany
<b>diPotassiumhydrogenphosphate</b>	Carl Roth, Germany	<b>Serological pipette 25 mL</b>	Sarstedt, Germany
<b>dNTPs</b>	Sigma Aldrich, MO, USA	<b>Serological pipette 5 mL</b>	Sarstedt, Germany
<b>Ethanol</b>	Merck-Millipore, Germany	<b>Tetracyclin Hydrochloride</b>	Carl Roth, Germany
<b>DTT</b>	Carl Roth, Germany	<b>Tetramethylethylenediamine (TEMED)</b>	Sigma Aldrich, MO, USA
<b>EDTA</b>	Applichem, MO, USA	<b>Tris acetate</b>	Carl Roth, Germany
<b>EGTA</b>	Sigma Aldrich, MO, USA	<b>TRIS base</b>	Sigma Aldrich, MO, USA
<b>Ethanol</b>	Merck-Millipore, Germany	<b>Tris-HCl</b>	Sigma Aldrich, MO, USA
<b>Ethidiumbromide</b>	Carl Roth, Germany	<b>tri-Sodiumcitrate diHydrate</b>	Carl Roth, Germany
<b>Falcontube 15 mL</b>	Greiner Bio-One, Austria	<b>Triton X-100</b>	Sigma Aldrich, MO, USA
<b>Falcontube 50 mL</b>	Greiner Bio-One, Austria	<b>Trypanblue</b>	PAA, MA, USA
<b>Fetal calf serum</b>	Gibco, MA, USA	<b>Trypsin-EDTA</b>	PAN, MA, USA
<b>Glycerol</b>	Sigma Aldrich, MO, USA	<b>Tween 20</b>	Carl Roth, Germany
<b>HCL solution</b>	Merck-Millipore, Germany	<b>Whatman Paper</b>	Carl Roth, Germany
<b>HEPES</b>	Carl Roth, Germany	<b>X-Gal</b>	Carl Roth, Germany
<b>HPLC-Wasser</b>	Millipore-Merck, Germany	<b>Yeast</b>	Carl Roth, Germany

### 2.1.21 BACTERIAL STRAINS INFORMATION

**Table 11.** Bacterial strains

<b>Strain</b>	<b>Genotype</b>	<b>Source</b>
One Shot® Stb13™ Chemically Competent <i>E.coli</i>	F-mcrB mrrhsdS20 (rB-, mB-) recA13 supE44 ara-14 galK2 lacY1 proA2 rpsL20 (StrR) xyl-5 λ-leumtl-1	Thermo Fisher Scientific, MA, USA C737303
One Shot® TOP10 Chemically Competent <i>E.coli</i>	F- <i>mcrA</i> Δ ( <i>mrr-hsdRMS-mcrBC</i> ) Φ80/ <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ ( <i>araleu</i> )7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (StrR) <i>endA1</i> <i>nupG</i>	Thermo Fisher Scientific, MA, USA C404010

### 2.1.22 CELL LINES

**Table 12.** Mammalian cell lines

Cell line	Description	Distributor information
HEK293	Human embryonic kidney	(ATCC® CRL-1573™)
HeLa	Human cervix carcinoma	(ATCC® CCL-2™)

### 2.1.23 CELL CULTURE MEDIA AND ADDITIVES

**Table 13.** Mammalian Cell culture Media

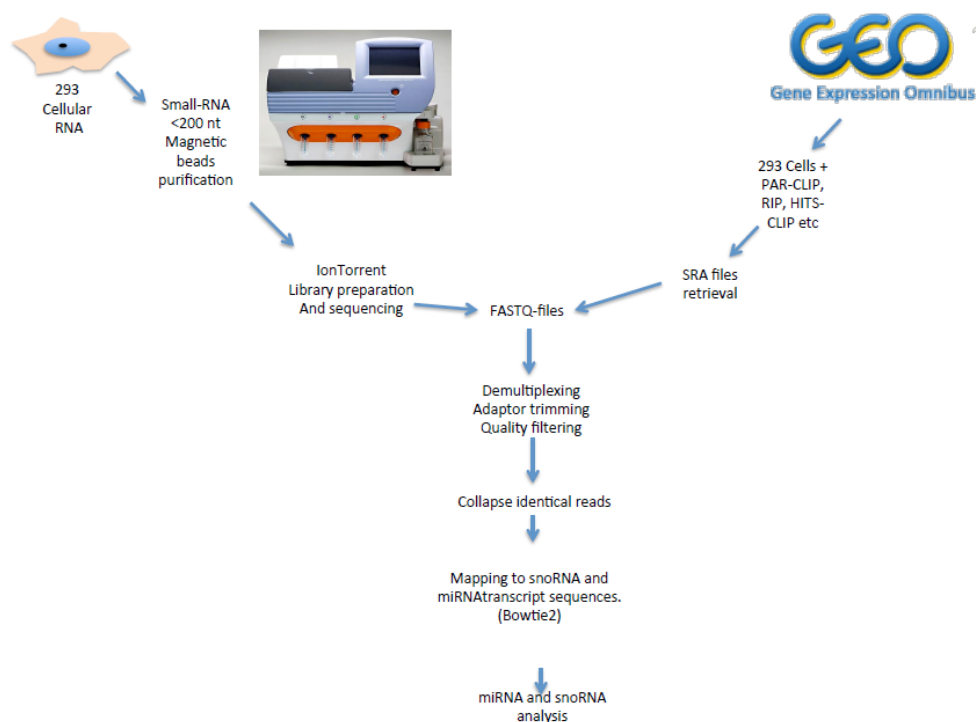
Medium	Additives	Company
DMEM	+ L-Glutamine, + phenol red, high glucose	Gibco
Optimem	+ L-Glutamine, + phenol red	Gibco
DPBS	No Ca, no Mg, no phenol red	PAN
Fetal calf serum	Heat inactivated	Gibco
RPMI	+ L-Glutamine, + phenol red	Gibco
Trypsin	0,05 % in PBS, no Mg <sup>2+</sup> , no Ca <sup>2+</sup>	PAN

## RESULTS

### NGS CHARACTERIZATION OF SNORNA-DERIVED MIRNAS

To functionally characterize C/D snoRNA-derived miRNAs (C/D-miRNAs) I used two experimental approaches: first using high-throughput derived data I described miRNAs structure, distribution, and interaction with RNA binding proteins. Secondly, I created a high-resolution assay to test miRNA functionality and applied it to supposed C/D-miRNAs.

Figure 7 outlines the experimental procedure I used to generate NGS raw data. For structure and expression, I sequenced small RNAs from Hek293 using the Ion-Torrent platform, while for RNPs interaction I retrieved raw data from CLIP and RIP-seq experiments available in the Gene Expression Omnibus<sup>110</sup>. Later, I filtered, trimmed and demultiplexed the FASTQ files and reduced PCR duplicates by collapsing identical reads.



**Figure 7. NGS Characterization of snoRNA-derived miRNAs.**

Experimental outline of small RNA characterization: Hek293 cells were used for Ion-Torrent libraries to evaluate cellular RNA content, while RAW data from GEO database was retrieved using the terms Hek293 cells and RNA immunoprecipitation. Then FASTQ files were preprocessed, collapsed and mapped to snoRNA and miRNA transcripts data base to further analysis.

#### 2.1.24 SNORNA-DERIVED MIRNA CHARACTERIZATION FROM SMALL RNA LIBRARIES.



### 2.1.24.1 snoRNA-derived miRNA characterization from Ion-Torrent NGS platform.

Since I am interested in snoRNAs-miRNAs and in reducing computing time, I aligned the collapsed FASTA files to a miRNA-snoRNA databases rather than the whole genome<sup>\*1</sup>, preferring an analysis of individual candidates analogously to Granneman reports of U3 in CRAC analysis<sup>111</sup>.

To select the C/D-miRNA candidates, I chose the top five highest expressed snoRNAs with coherent processing; for that reason, I sequenced three independent Hek293 cultures. To assure that I operate with reduced degradation bias, I just used high-quality RNA (RIN > 9) to prepare and to sequence libraries using the Ion-Torrent platform. (Supplementary Figure 1).

The three sequenced libraries show a high correlation between expressed small-RNAs (Figure 8 A) and the top 30 identified miRNAs cluster together - being L001 and L004 closer (Figure 8 C) and this is consistent with their Pearson correlation coefficient. Their read length distributions show a high peak at 21 to 22 nt with a mild peak around 60 to 70 nt (Figure 8 B), these two peaks correspond to miRNAs and snoRNAs respectively, these two peaks match to miRNAs and snoRNAs respectively, as shown in the fraction of recovered small RNA. (Figure 8 D)

Additionally the small-RNA seq yield similar patterns as other platforms (For details look next section).

To define C/D-miRNAs, I subsetted snoRNA reads by length, and classified snoRNAs in two: *mature*, with reads longer than 40 nt (Figure 9 A), and *C/D-miRNA* as sequences below 40 (Figure 9 B); the top expressed genes show dissimilar distribution between long and small snoRNAs.

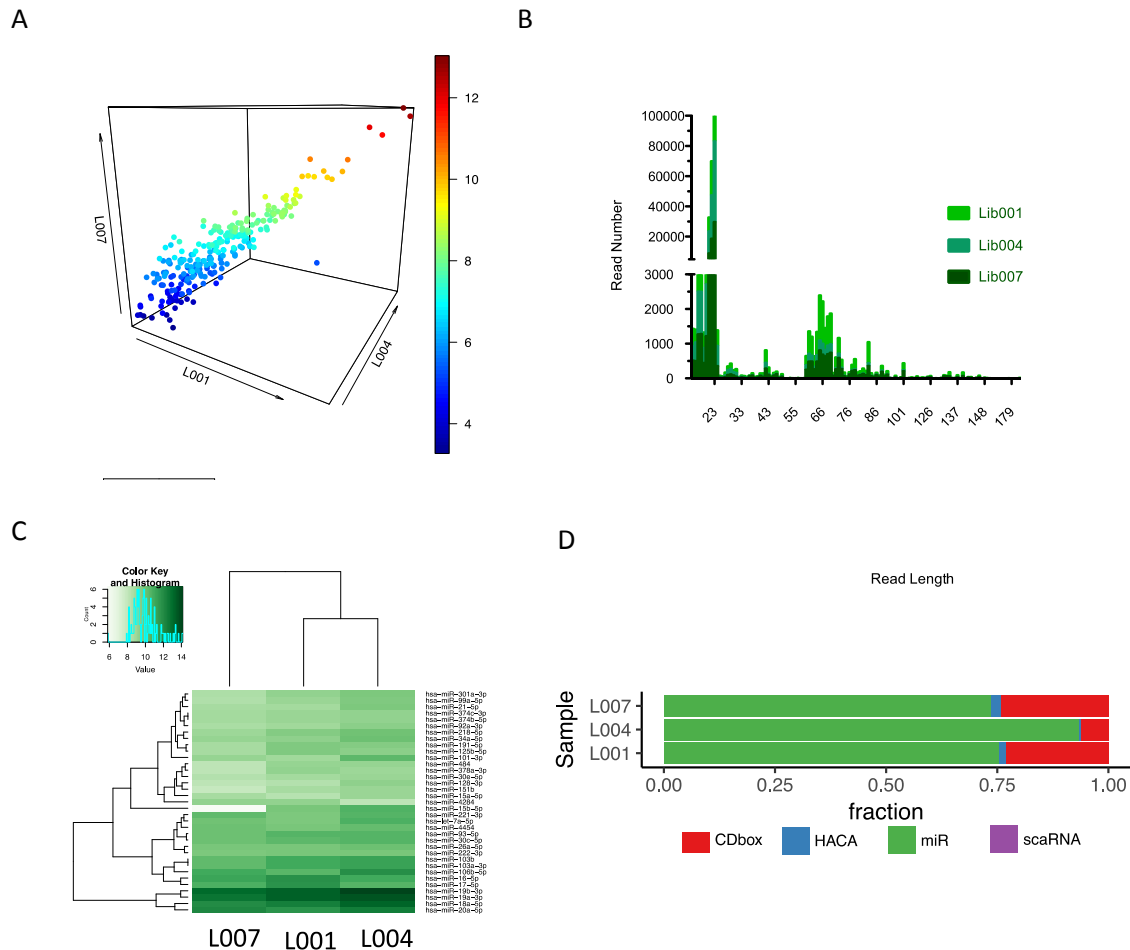
To rule out that the C/D snoRNAs are degradation remnants rather than stable miRNAs; called here coherently processed<sup>112</sup>. I checked snoRNAs read distribution on genomic tracks (Supplementary Figure 2 A), but this require complicated and time consuming manual inspection of genome aligned reads.

To simplify and automatize this procedure, I employed hierarchical clustering analysis to classify snoRNA processing using binary matrix. Concisely I calculated the binary matrix using the transcript length as columns, and the reads as rows. Since every read has defined start and end,

---

\*1 Genomic alignments were performed and used to analyzed processing patterns, see hierarchal clustering below.

each position in the transcript has a value 1, while 0 designates no read at that particular site; creating the binary matrix.



**Figure 8. small RNA seq libraries revealed expected distribution for snoRNA and miRNAs.**

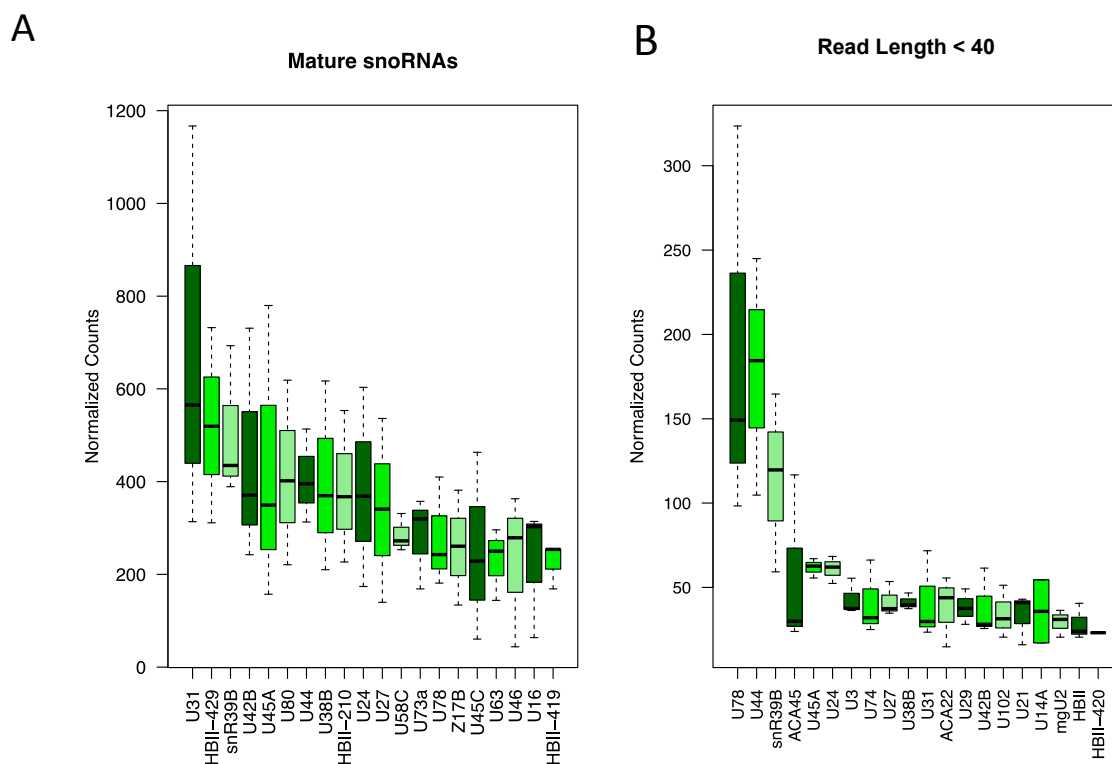
Small RNA from three independent Hek293 cultures were purified, subject to library preparation and sequenced. Analysis of small RNA content from the three samples: (A) small RNA content comparison, each dot in the scatter plot represents log<sub>10</sub> transformed reads - Pearson Coefficients L001 vs. L004 R= 0.979, L001 vs. L007 R=0.969. (B - C) Read size distribution and hierarchal clustering from normalized read counts. (D) miRNA and snoRNA (C/D snoRNA, scaRNA, and H/ACA) fraction.

To simplify and automatize this procedure, I employed hierarchical clustering analysis to classify snoRNA processing using binary matrix. Concisely I calculated the binary matrix using the transcript length as columns, and the reads as rows. Since every read has defined start and end, each position in the transcript has a value 1, while 0 designates no read at that particular site; creating the binary matrix.

Therefore, a binary matrix has columns as nucleotides in the transcript and number of rows equal to reads aligned to it. For example, a U78 array will have 65 columns and around 200 rows for the Ion-Torrent libraries (Figure 9).

Hierarchical clustering analysis (HCA) arranges similar observations sequentially, grouping small groups until it reaches a single batch<sup>113</sup>.

I assumed that coherent processing should create clear clusters, and define the patterns of processing or RNP protection. Supplementary Figure 2 shows reads before and after HCA. The HCA of the top expressed snoRNA shows three clusters: the mature snoRNAs occupying the



**Figure 9 snoRNA-derived miRNAs expression differs from mature counterparts**

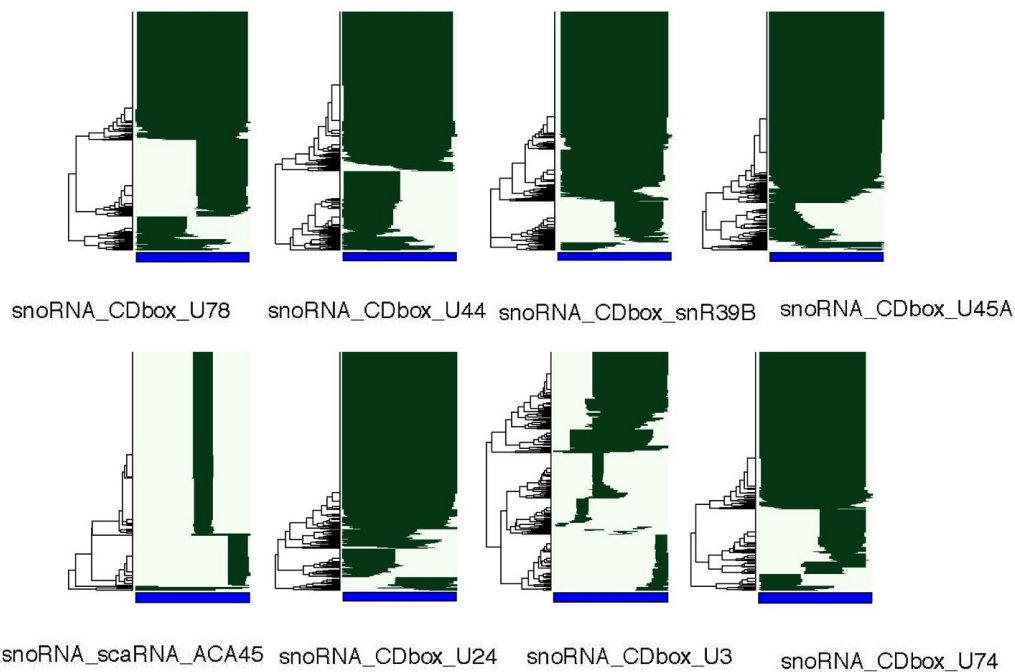
snoRNAs reads were divided into groups: Mature snoRNAs (length >40) and snoRNA-derived miRNA RNAs (length >40) and descending organized. Boxplot of the 20 most expressed (A) mature snoRNAs and (B) derived miRNA normalized reads plotted without transformation.

whole transcript and the derived miRNAs (“passenger” and “guide” strands<sup>74</sup>). Showing that HCA defines quickly the processing of snoRNAs.<sup>\*2</sup>

<sup>\*2</sup> Candidate selection (See Figure 10 and Figure 11) was also performed manually; using genomic tracks inspection IGV genome browser with an average time of 20 to 30 min per snoRNA. Binary matrices, hierarchical clustering and graphical output (in R functions) requires an average of 1 minutes for all the snoRNA of Ion-Torrent libraries.

C/D-miRNA clusters drop into two distinct groups: 1. *processed*, which have short leaves and clusters with defined long branches, and 2. *degraded*, holding long leaves and don't coerce in a specified node, compare cluster trees of U78 vs. U45A. I decided then, to further analyze just the processed C/D miRNAs.

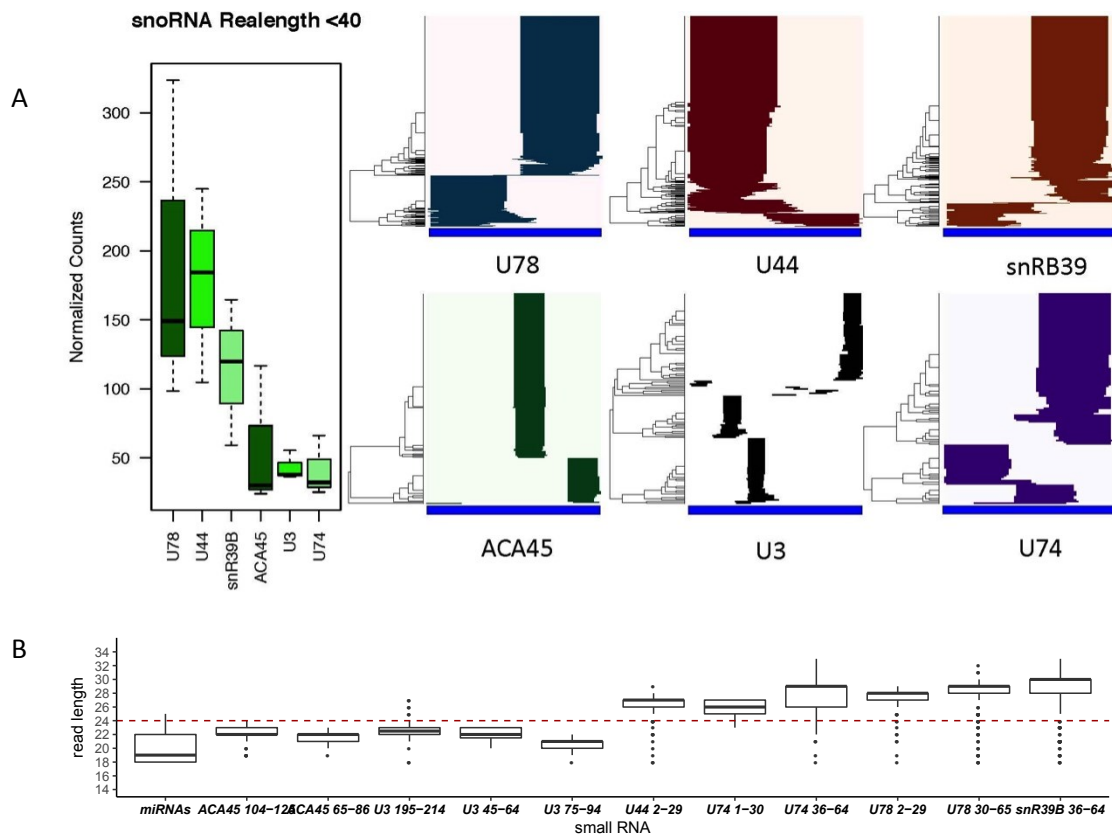
Small RNA profiling recover defined patterns of miRNAs that match the guide or the passenger strand, with the first being mostly recovered. For the candidates that filled the processed group, I observed a similar trend (Figure 10) and because the H/ACA derived miRNAs ACA45 passed the initial examination, and it is the first group VI well-characterized miRNA<sup>4</sup>, I integrated it as snoRNA-derived miRNA control.



**Figure 10 Reads hierarchical clustering analysis sorts snoRNA-processing patterns**

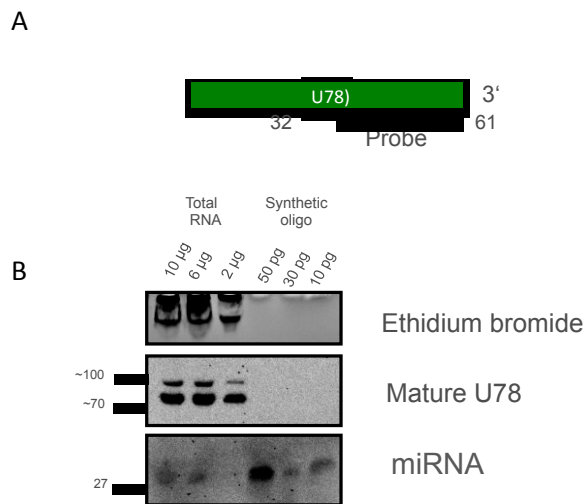
To contrast with genomic tracks analysis (Supplementary Figure 2), snoRNA reads transformed into binary matrices were subjected to hierarchical clustering analysis, cluster tree and reads over the snoRNA transcripts (blue bars) plots from the most expressed snoRNA- miRNAs.

Moreover, I compared the length distribution for the C/D-miRNAs with canonical miRNAs; I found that H/ACA-45 and the miRNAs range are below 25 nt, with a median of 22 and 19 respectively. For the C/D derived miRNAs, U78, U44, U74, and snRN39 ranges are longer than 25, while U3 derived miRNA drop into controls range (Figure 11).



Finally, I wanted to see the C/D-miRNAs without the library preparation bias; so I performed small RNA northern blot for the most expressed C/D miRNA U78 (Figure 11 A). To optimize the experimental condition that could detect RNA, corresponding to 200 reads on Ion-Torrent (approximately one pg per microgram of total RNA)<sup>\*3</sup>, I used a synthetic RNA as control Supplementary Figure 4. I ran total RNA and observed a signal in the same region as the synthesized control Figure 12.

<sup>\*3</sup> Rough calculation from total aligned reads (~ 350.000) and small RNA input (~ 100 ng), assuming linear relation 200 reads corresponds to 57 pg.



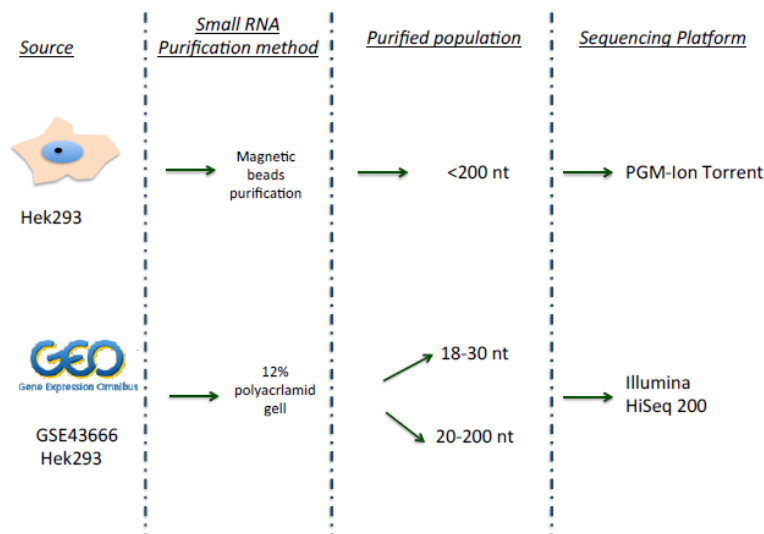
**Figure 12. U78 derived miRNA detection by northern blot.**

Decreasing concentrations of Hek293's total RNA were separated in a 15% UREA-PAGE, transferred to a nylon membrane and incubated at with the probe at 45 C for 16 h. A synthetic RNA, which contains the most expressed snoRNA-miRNA sequence, was used as reference. (A) Graphical representation of U78 C/D snoRNA and the used probe (B) Ethidium bromide stain and northern blot results for mature and snoRNA-derived miRNA.

#### 2.1.24.2 Comparison between small RNA purification methods: small RNA libraries yield of mature snoRNAs.

C/D snoRNA miRNAs are unlikely to be library preparation artifacts, and probably produced independently from the mature forms ( Figure 9 ). However, I presumed that Ion-Torrent profiles might be biased since the purification protocol is optimized for miRNAs, the read-length distribution shows more miRNAs than snoRNAs (Figure 8 A), which contradicts the assumption that for ribosome biogenesis snoRNAs should be abundant.

To compare methods and rule this bias out, I searched for Hek293 data sets in the GEO repository and I found two Illumina libraries produced with distinct purification methods (GSE43666). The original experimenters generated two small RNA pools: one for miRNAs (18-30) and others for small RNAs (20-200). Figure 13 contrasts the performed procedures.



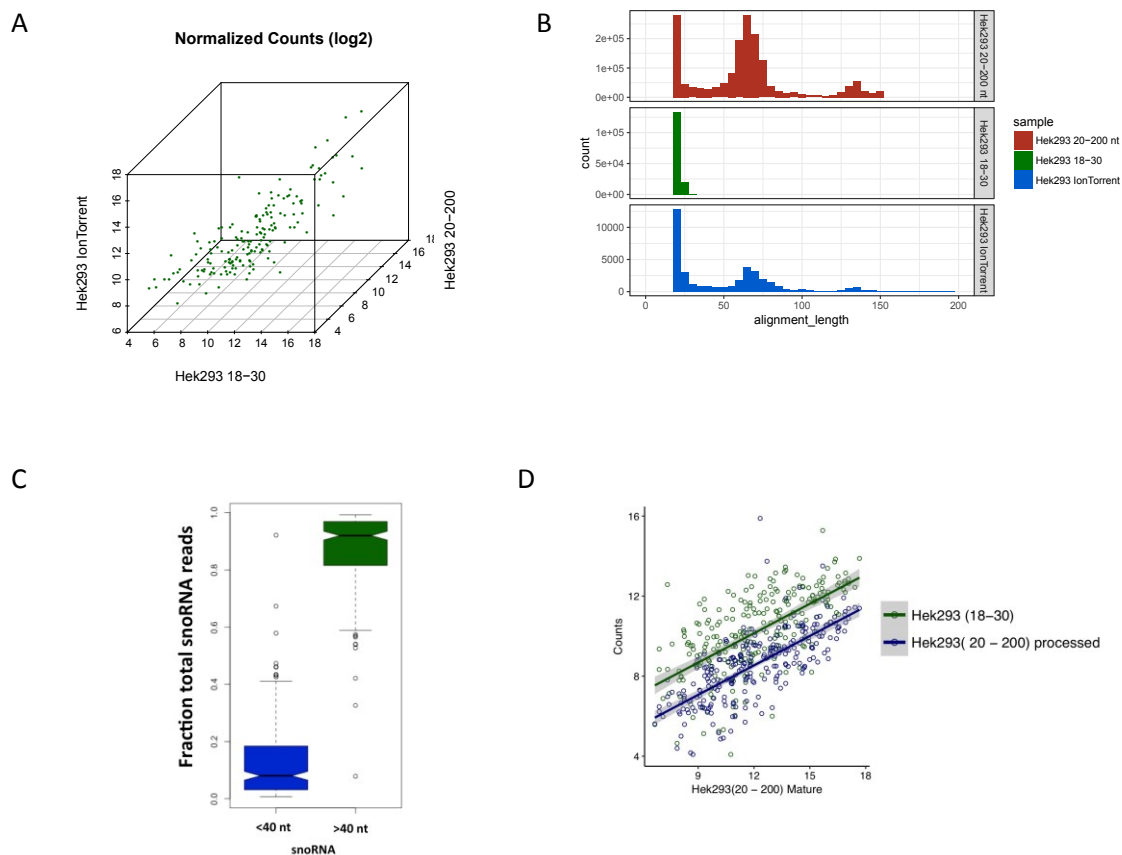
**Figure 13. Outline of raw data recovery for size separated RNA.**

Comparison of ethanol and PAGE based RNA separation for small RNA libraries for Ion Torrent and Illumina NGS sequencing, the cartoon depicts the source from cell culture and GEO raw data for Hek293 cells, the method of purification, the expected size, and the sequencing platform.

I processed the FASTQ files, aligned them to the snoRNA-miRNA database, and compared them with my Ion-Torrent libraries; I observed a positive correlation between the three data sets with a higher Pearson coefficient ( $R=0.68$ ) between the Ion-Torrent and the 18-30 nt Illumina libraries (Figure 14 A). The small-RNAs comparison relates the read length distribution and the purification method, whereas the 20-200 nt shows peaks around 21 and 70 the Ion Torrent has a smaller one around 70 (Figure 14 B); implying that Ion-Torrent profiles do not characterize properly mature snoRNAs.

Since the libraries correlate, and the 20-200 produced library has the proper small RNA distribution, I looked at the relationship between mature snoRNAs and the C/D-miRNAs. Calculating the fraction (mature-miRNA) for each C/D snoRNA, i.e., if a snoRNA has 10 miRNAs and 90 long reads, the fraction will be 0.1 and 0.9 respectively.

The whole C/D snoRNA population shows that the C/D miRNAs (processed) are below 0.2 and that recovered reads are mature forms (Figure 14 C). The expression relationship, between long and short snoRNAs, show a positive correlation for both, 18-30 and 20-200 nt, implying that the mature snoRNA are the source of miRNAs.



**Figure 14. snoRNA-derived miRNAs content is lower than mature snoRNAs and their expression positively correlates.**

Expression comparison of small RNAs libraries: Ion Torrent (generated here) and retrieved Illumina (GSE43666 18-30 and 20-200) raw data analysis. (A) Scatter plot of log transformed and normalized reads (Pearson Coefficients Ion-Torrent vs. 18-30  $R=0.68$  and Ion-Torrent vs. 20-200  $R=0.49$ ) and (B) Read length distribution between libraries. Expression of snoRNA-derived miRNAs compared with mature snoRNAs. (C) Boxplots depict mature vs micro snoRNAs fraction from each C/D snoRNA in the 20-200 Illumina library. (D) scatter plot 20-200 mature vs. snoRNA-derived miRNAs library (18-30 green) and a subset ( $< 40$  read length blue) 20-200 library.

Overall, the purification methods yield different small RNA subpopulations, and interpretations about mature snoRNA from miRNA-focused analysis require caution. But to falsify the claim “C/D miRNA works as canonical miRNAs”, any library either Illumina or Ion-Torrent for miRNAs is appropriate for interpretation.

### 2.1.24.3 snoRNA-derived miRNAs subcellular location.

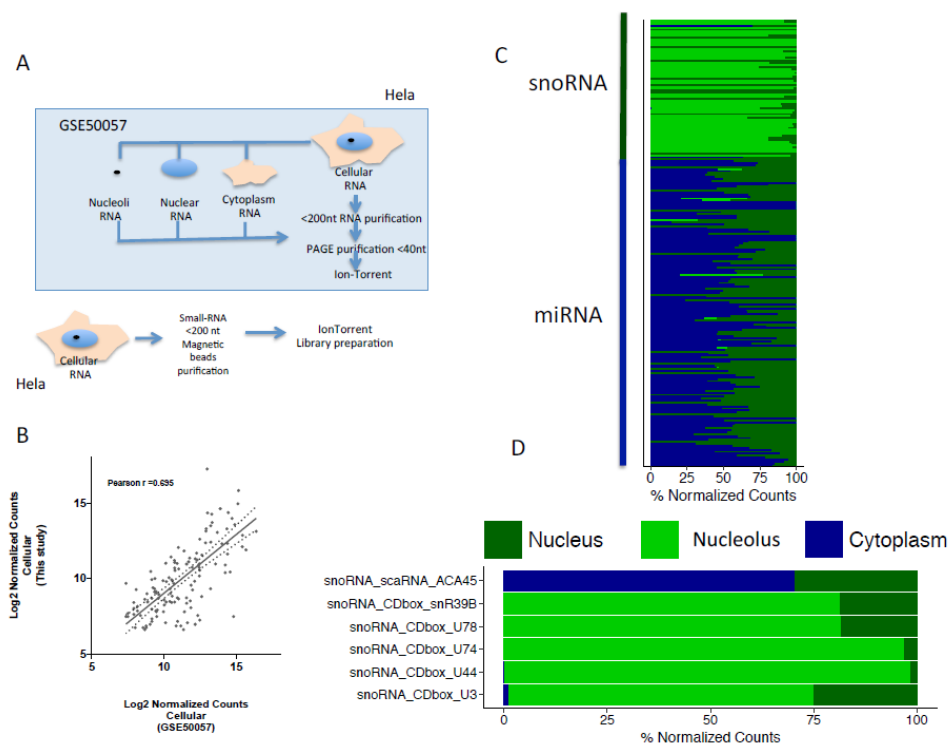
To evaluate the cellular fraction of C/D miRNAs, I found at the GEO repository a data set for cellular fractions produced from Hela cells (GSE50057). To validate this data, together with Dr. Böker we sequenced and analyzed small RNA libraries from total RNA in Hela cells (Figure 15 A),



using the same RNA purification and library standards applied to Hek293 cells (Supplementary Figure 5).

Our HeLa data is equal to the retrieved GEO libraries since all the libraries positively correlate (**Figure 15 A** and Supplementary Figure 5), and the processing clusters for ACA45, U3, U78, U44, U74, and snRN39 resembled the Hek293 (data not shown). I analyzed their subcellular percentage for the whole small RNA population and selected candidates (**Figure 15**)

The location percentages show that miRNAs locate at the cyto- and nucleoplasm, while the C/D snoRNAs mainly at nuclear compartment, which is right for U78, U44, U74, and snRN39. Interestingly, I found ACA45 and U3 miRNAs in the cytoplasmic fractions, while ACA45 shows the higher percentage, U3 is lower than its nuclear and nucleolar location ( **Figure 15 D**).



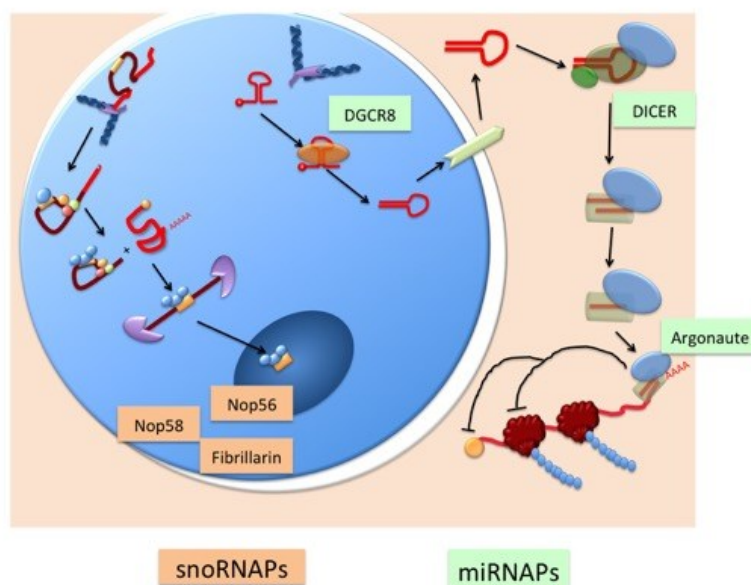
### Figure 15. snoRNA-derived miRNAs locate in the nucleolus

Raw data -from cellular fractionations (GSE50057) and Ion-torrent data from HeLa cells (kindly provided by Dr. Böker, Medical RNA Biology) were analyzed for snoRNA-derived miRNAs. (A) Outline of the experimental approach for fractionation. (B) Scatter plot compares whole cellular content libraries Böker vs. GSE50057 (Pearson Correlation R). The fractions of small RNA between cytoplasm, nucleus, and nucleolus. (C) miRNA and snoRNA-derived miRNAs depicted as a heat map and (D) for selected candidates.

Until now, I pointed out that:

- C/D snoRNAs loci produce RNA fragments of approximately 25 nt.
- C/D small fragments depict guide-passenger relationship in their locus in sRNA seq.
- C/D miRNAs are lower than their canonical counterparts (mature snoRNAs)
- C/D miRNAs locate in the nuclear and nucleolar fractions.

#### 2.1.24.4 snoRNA-derived miRNAs characterization from RIP and CLIP seq.



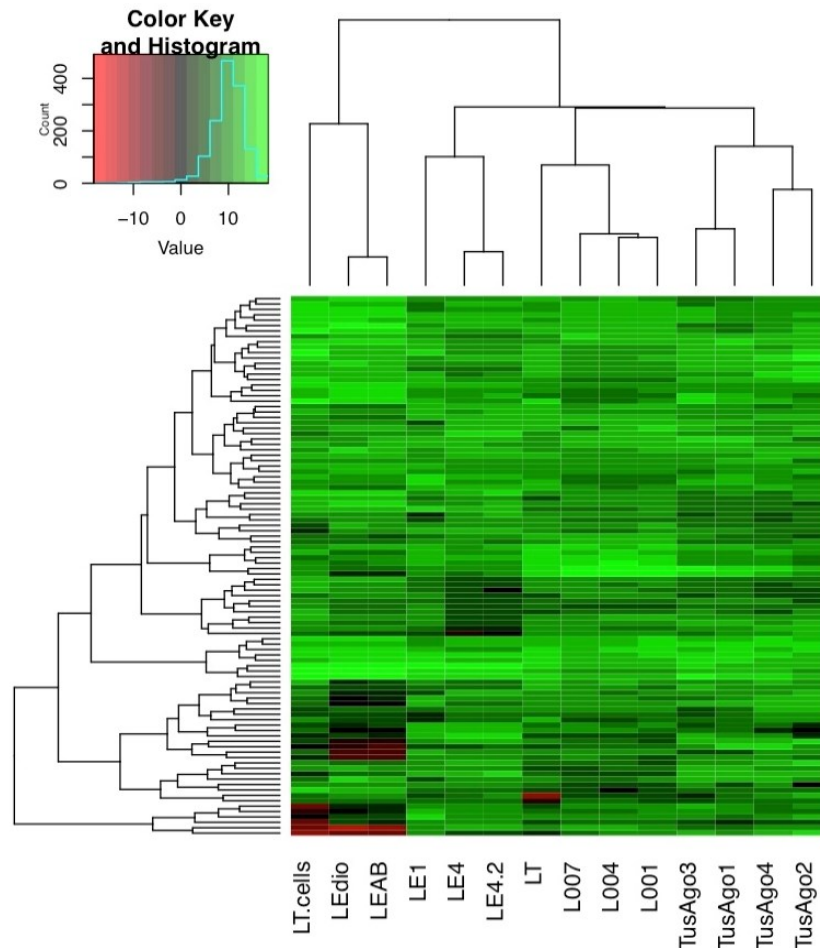
**Figure 16. CLIP and RIP-seq analyzed data.**

To check snoRNA-derived miRNAs interaction with ribonucleoproteins, GEO raw NGS data for Hek293 was retrieved. Cartoon depicts analyzed selected proteins, snoRNPs-Nop58, Nop56, and Fibrillarin (GSE43666) and miRNPs, -Dcrg8 (GSE39086), Dicer (GSE55333) and Argonaute (CLASH-GSE50452, PAR-CLIP GSE21918, and RIP GSE58127).

After I had evaluated the cellular C/D miRNAs, I decided to check the interaction with ribonucleoproteins, using again predictive the predictive power of scientific paradigms. We can define the following conditional predictive statement: *if C/D-miRNAs functional activity mirrors canonical miRNA post-transcriptional regulation, they should be pulled down concomitantly with Argonaute proteins and interact with the miRNAs maturation machinery.*

For that reason, I retrieved GEO repository data from RIP and CLIP-seq raw data for Hek293 cells and miRNAs, after specific preprocessing and PCR duplicates removing (due to different library preparation), I aligned it to snoRNA and miRNA databases and examined individually for U78, U44, U74, and snRN39. For purposes explained below (2.1.24.7), I also recovered data for C/D snoRNPs. **Figure 16** presents the obtained data and contains the respective accession numbers.

### 2.1.24.5 Argonaute PAR and CLIP seq recover U3 and ACA45.



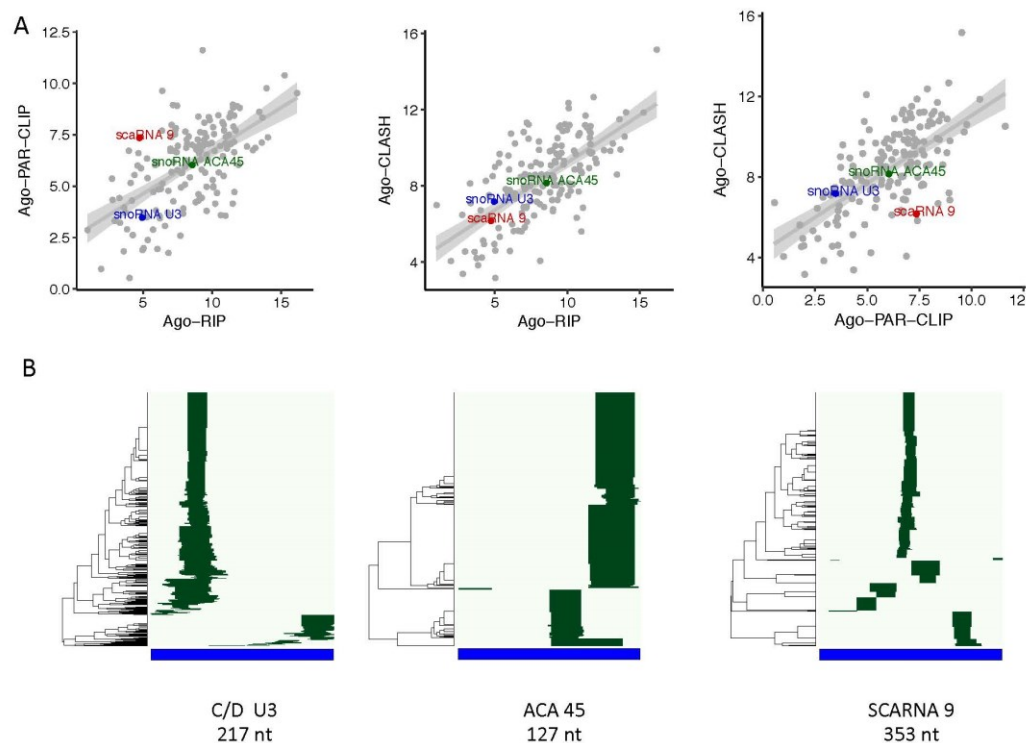
**Figure 17. Hek293 small RNA, RIP and CLIP libraries recover comparable miRNA profiles.**

Aligned reads to snoRNA and miRNA database were normalized using DESEQ2 package. Heatmap depicts the 100 most expressed genes from the analyzed libraries, including Ion-Torrent small RNA libraries sequenced in this study (L001, L004, L007). For retrieved raw data, it depicts, RIP-seq (Total small RNA -LT cells and two different Ago antibodies - LEdio for diogenade and LEAB for Abcam), CLASH (LE1, LE4 and LE4.2), PAR-CLIP (TusAGO 1-4) and Illumina small-RNA for 18-30 (LT).

To select good quality data and to take out any introduced bias by different sequencing runs and experimenters. I reasoned that the miRNA content from Ion-Torrent profiling (see 2.1.24.1) should highly correlate (Pearson Correlation  $R > 0.5$ ) with Argonaute RIP- CLIP data and possess a weak variation in their expression.

Before preprocessing, I ran FASTQC in all the raw FASTQ files and selected just the data that gave overall good quality and that the overrepresented sequences fit the reported adapters and multiplexing; surprisingly, few-data sets full filled this criterion (data not shown).

Six data sets from Hek293 cells were consistent with FASTQC reports and published details: Three RIP-seq and small RNA Illumina libraries (GSE58127) and three AGO1 CLASH data (GSE 50452), the FASTQC results from PAR-CLIP GSE21918 failed the quality test. However, I filtered all the bad quality reads, because it has the CLIP data from the four AGO isoforms.



**Figure 18. Argonaute CLIP and RIP identifies 3 bound snoRNA-derived miRNAs**

Argonaute immunoprecipitated data comparison.(A) Scatter plot showing read normalized transformed reads and the mean of biological replicas for each dataset. Colored dots show the snoRNAs bound in all the Argonaute: ACA45, U3 and scaRNA 9. Hierarchical clustering for recovered snoRNAs, (B) Cluster tree and reads (green) over the snoRNA transcripts.

The Argonaute pulldowns, Ion torrent, and small RNA sequencing positively correlate (Supplementary Figure 7), and the top 100 miRNAs expression is constant among the libraries. Unsupervised analysis clusters the samples according to their source (Figure 17); the group

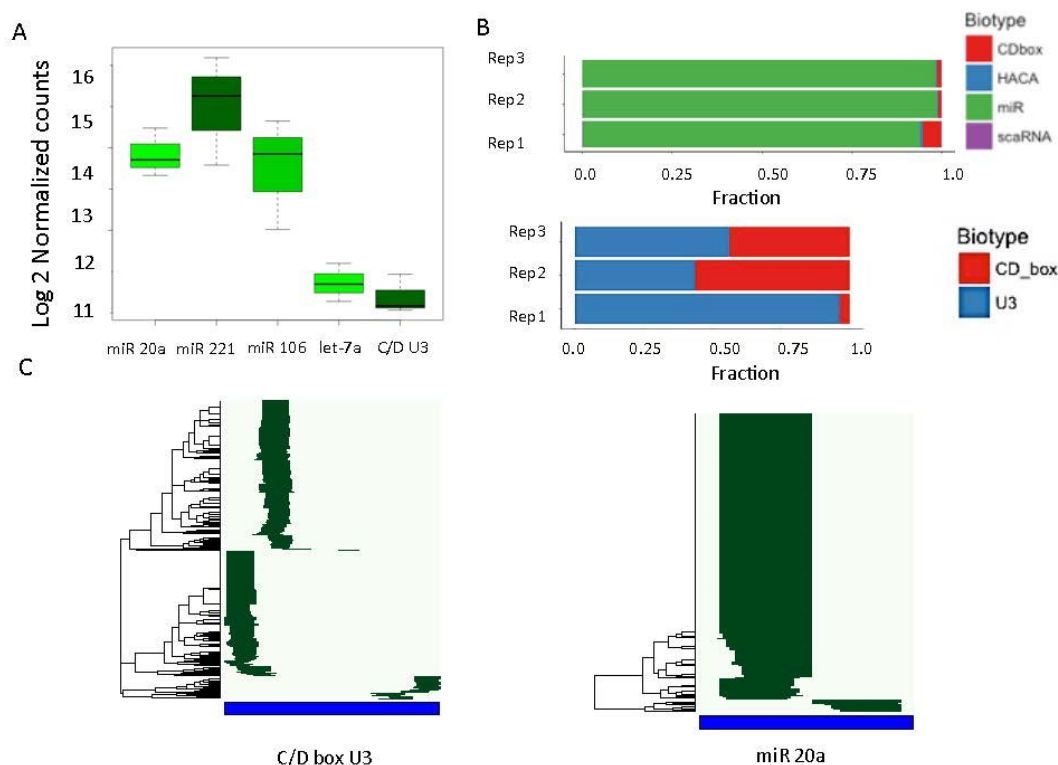
between LT (Illumina 18-30) and Ion-Torrent supports the high correlation observed before (Figure 14).

In total, I analyzed 9 Argonaute pulldowns and decided to be extremely stringent sorting out C/D miRNA-Ago interactions. So, I considered valid RNA-Ago only the RNAs recovered in the nine libraries, and noted only three snoRNAs U3, ACA45, and SCARNA9 (Figure 18 A).

To compare the small RNA distribution with the cellular libraries I ran hierarchal clustering of Ago binary matrices and observed similar processing clusters (Fig 12 B) for U3 and ACA45; I neglected SCARNA9 for further analysis because it wasn't recovered before (Section 2.1.24.1).

### 2.1.24.6 Dicer and DGCR CLIP seq recover C/D snoRNAs.

For Dicer analysis, I found PAR-CLIP data in Hek293 cells that passed FASTQC quality filter and I analyzed three biological replicates. Interestingly, U3 appeared within the top 5 and is the most recovered C/D snoRNA (Figure 19A and B), while ACA45 appeared just in one biological replicate (~20 reads).



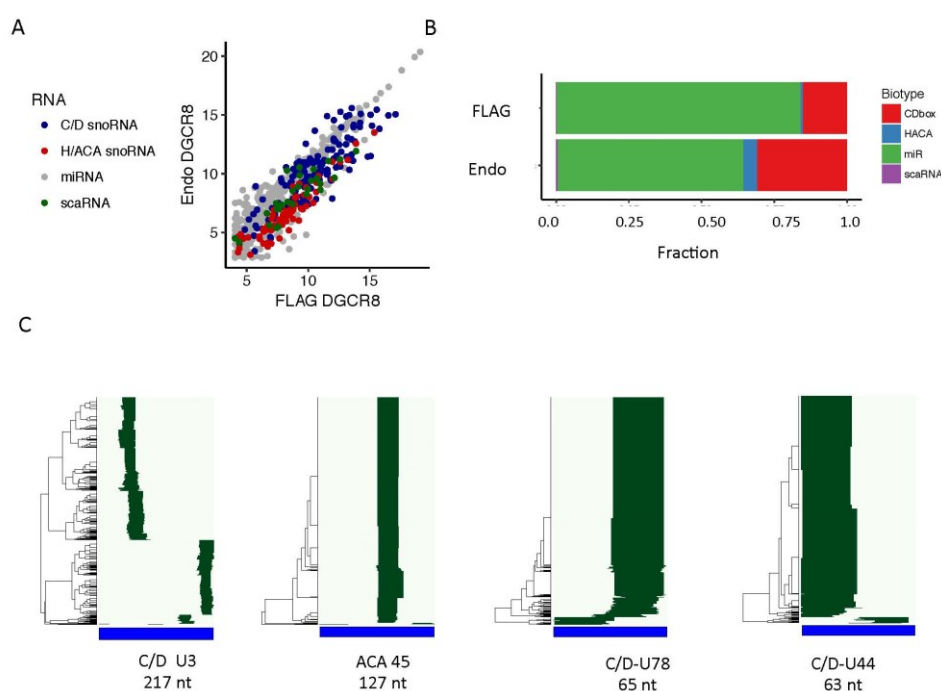
**Figure 19. Dicer PAR-CLIP recovers U3 snoRNAs fragments.**

Hek293 small-RNA recovered from Dicer PAR-CLIP comparison of three biological replicas. (A) 5 Top retrieved small-RNAs normalized and transformed reads. (B) Small RNA Fractions for the three replicates and a subset for C/D snoRNAs. (C) U3 and miR 20a Hierarchal clustering,

cluster tree and reads (green) over the RNA transcripts.

Processing clusters for U3 showed two main groups: one located as in Ago pull downs (position 45 to 67)(Figure 19 C and Figure 18 B) and the other upstream with few connecting sequences, similarly to the distribution observed in miRNAs<sup>112</sup> and (Figure 19 C).

Finally, I retrieved data from the microprocessor component DGCR8 HITS-CLIP, for endogenous and overexpressed protein and I found all C/D miRNA candidates (U78, U74, ACA45 and U3), and a relatively big number of snoRNAs (Figure 20 A and B).



**Figure 20. DGCR8 HITS-CLIP recovers snoRNA and miRNAs.**

Endogenous and FLAG expressed DGCR8 recovered small RNAs, (A) scatter plot of normalized-transformed reads, colored dots show snoRNA subpopulations. (B) The fraction of small RNA for endogenous and FLAG-immunoprecipitated RNA, (C) Hierarchical clustering, cluster tree and reads (green) over the RNA transcripts of top 5 expressed snoRNA-derived miRNAs selected before (Figure 11).

DGCR8 processing clusters resemble C/D-miRNA from cellular small-RNA profiling (Figure 11); U78, U44, and ACA45 retrieved groups are similar to the putative guide strands, while U3 clusters show similar distribution between the 5' and 3' C/D miRNAs (Figure 20 C).

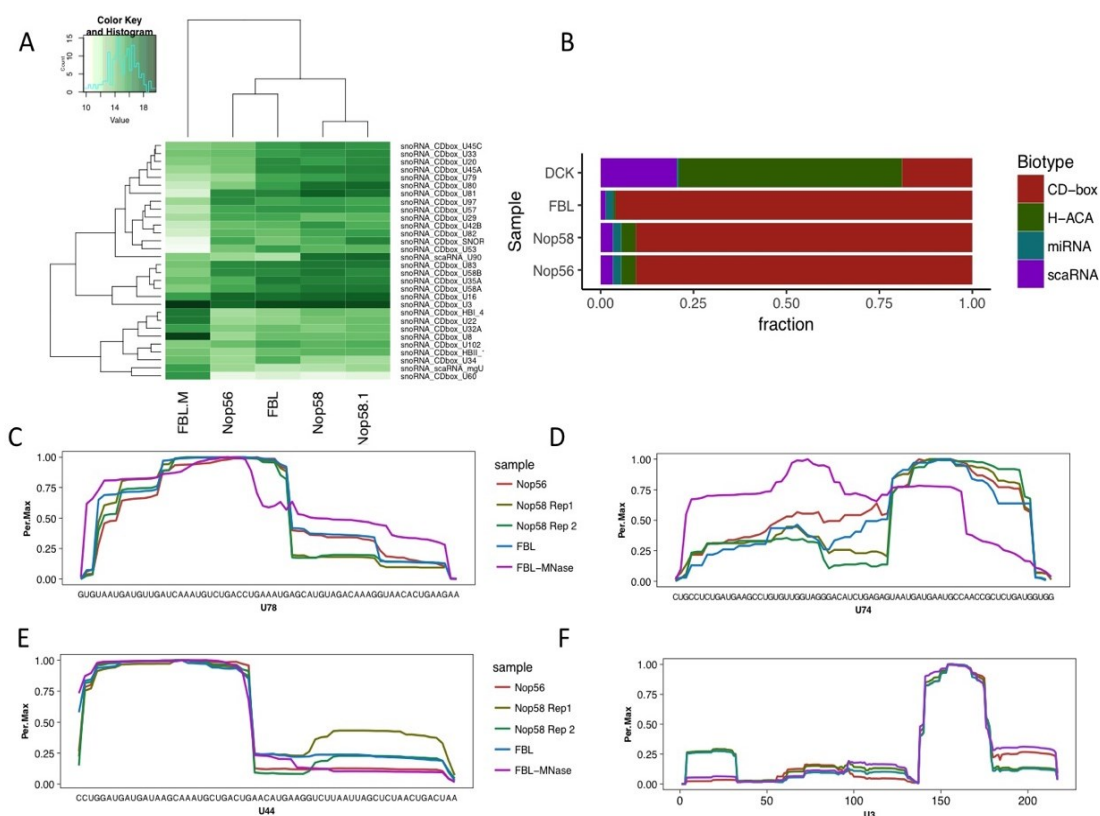
I noticed that the C/D snoRNA DGCR8 resembles the C/D miRNA-processing clusters, and since small RNA recovered from total RNA reflects RNA-RNP protection from nucleases either by AGO or Dicer<sup>114</sup>, I asked whether the small RNA profiling show RNP protection of snoRNAs by RNPs different from miRNPs.

### 2.1.24.7 snoRNP CLIP seq recover similar patterns C/D snoRNA-derived miRNA clusters.

I questioned how does sRNA-seq recover these processing patterns?

So if Argonaute pulldowns do not enrich intronic snoRNAs and that coherent processing retrieved in small RNAs profiling mirrors RNPs interaction (typically for miRNAs), then other RNP interaction should explain the coherent process of C/D derived miRNAs. Thus, I retrieved and analyzed PAR-CLIP data of snoRNPs

A significant number of recovered RNAs are C/D snoRNAs; their expression is even between replicates and snoRNP, confirming that the data is technically and biologically reliable (**Figure 21 A and B**).



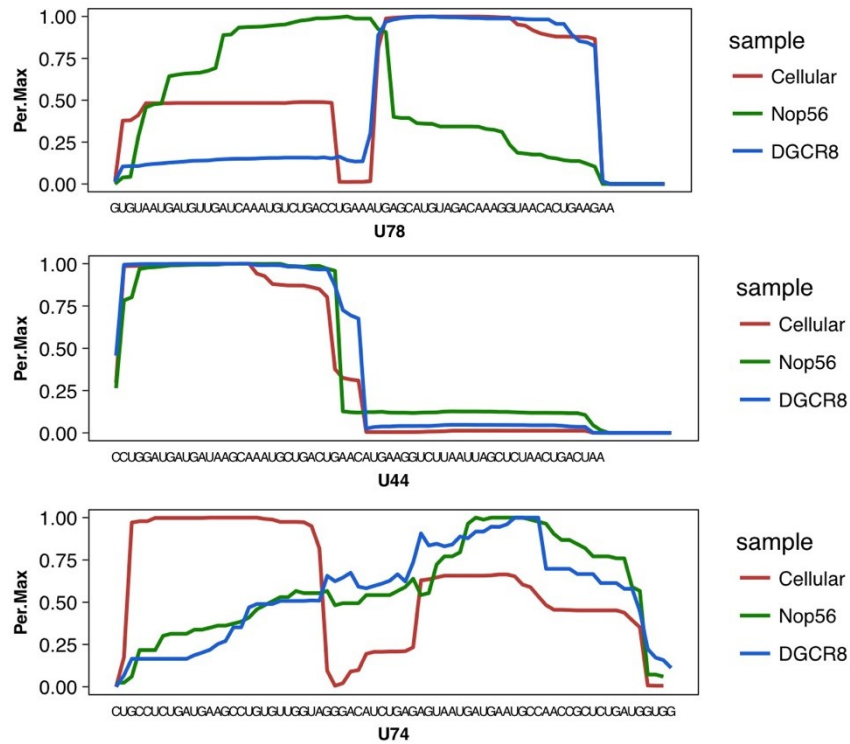
**Figure 21. snoRNPs PAR-CLIP recover known snoRNAs.**

Profiles of snoRNAs recovered from PAR-CLIP of snoRNA-derived miRNAs (A) Heatmap and (B) small RNA fractions from the PAR-CLIP seq libraries. Density graphs show reads position as a function of percentage of maximum (Read count/Maximum read in the transcript); U78, U44, U74 and U3 are plotted.



In detail, C/D snoRNPs (Nop56, Nop58, and Fibrillarin) pull together C/D snoRNAs, while Diskeryn does it for H/ACA and scaRNAs (Figure 21 B). To display the similarity between the samples, I replicated Granneman CRAC data display<sup>111</sup>, generating density plots that have the nucleotide position as a function of normalized reads (using percentage of maximum). The C/D snoRNAs distributions are similarly within snoRNPs, and the read length recovery resembles processing clusters of Ion Torrent profiles, including canonical C and D boxes, but also complete snoRNAs (Figure 21 C-F).

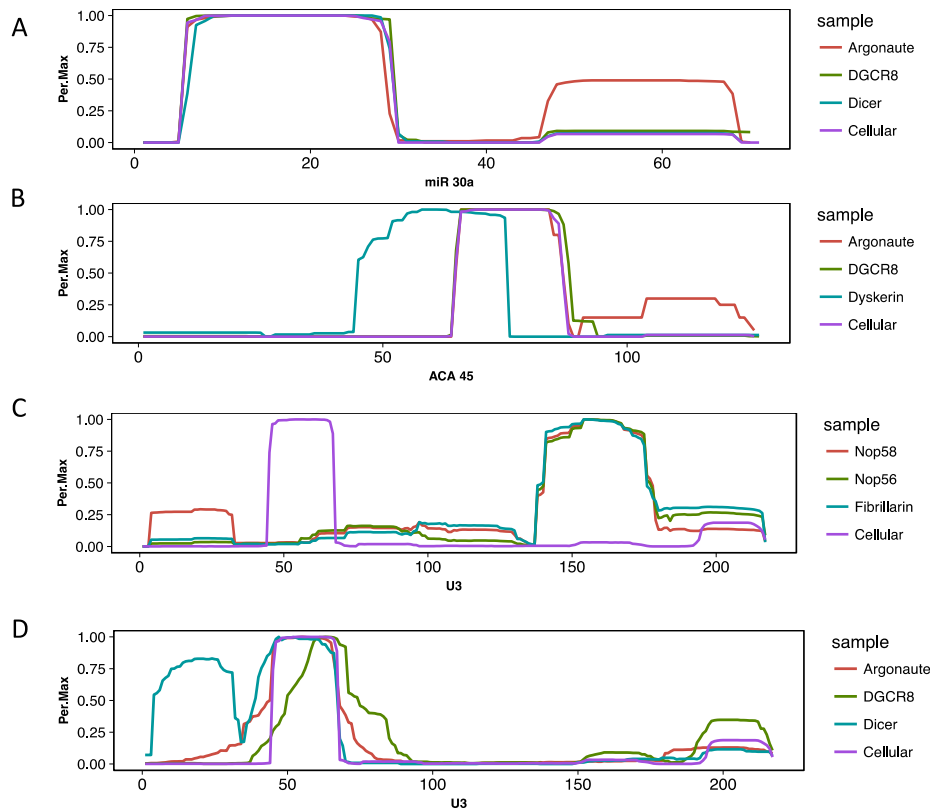
Finally, to test whether the processing clusters in Ion-Torrent profiles reflect *in vivo* RNP protection, I compared C/D-miRNA from cellular RNAs with the one snoRNP and DGCR8 -known to interact with snoRNAs and to work as an adaptor for the nuclear exosome.



**Figure 22. Nop56 and DGCR8 snoRNA recovery resembles snoRNA-derived miRNA clusters.**

Selected candidates for intronic C/D snoRNAs U74, U44, and U74, identified for snoRNA-derived miRNAs recovery. Density plots depict the distribution for Nop56, DGCR8, and small cellular RNA. Graphs show Reads position in function of the Fraction maximum, (Read count/Maximum read in the transcript).

The density plot from RNP and cellular reads overlapped, for U44 the overlay is complete, for U78 the overlap is perfect for DGCR8 and partial for Nop56 and U74 recovery has no obvious overlap between RNPs and the Ion-Torrent Profile (Cellular)(**Figure 22**).



**Figure 23. snoRNP and miRNPs CLIP-seq recover U3 and ACA45.**

Reads distribution for snoRNPs and miRNPs CLIP-seq. (A) Canonical distribution of miRNPs and cellular recovered on miR-30a transcript, (B) ACA45 distribution of Dyskerin and miRNPs, and (C) U3 identified reads from miRNPs and snoRNPs, Density graphs show reads position in function of the Fraction maximum, (Read count/Maximum read in the transcript)

So far, I described that the snoRNA loci produced miRNA size fragments, with a structure that resembles the guide and strand guide recovery of miRNAs; pointing out that their central location is nucleolar and that they are not bound to Ago but with the microprocessor unit DGCR8. However, I found two candidates that do not fit those characteristics, one that serves as control ACA45, reported and well proved H/ACA snoRNA-derived miRNA, and U3, that is a C/D snoRNA but evolved with distinct function and expression.

However, I found two candidates that do not fit those characteristics, one that serves as control ACA45, reported and well proved H/ACA snoRNA-derived miRNA<sup>4</sup>, and U3, that is a C/D snoRNA but evolved with distinct function and expression.

U3 is not intronic and larger than the average C/D snoRNA (217 vs. 70 nt), interact with the classic snoRNPs and with the particular U3 protein (U3-55K)<sup>115,116</sup>. It also has an intricate protein network and location during its biogenesis and is necessary for the early cleavage of the 5' ETS (see 1.1.3.2).

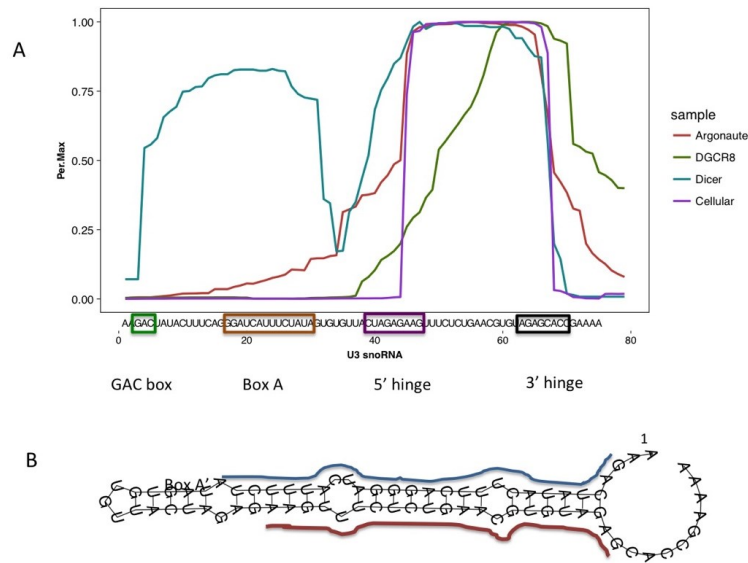
CLIP data showed that ACA45 interacts with Argonaute, DGCR8 and poorly with Dicer (~20 reads in one replicate), while its interaction with Dyskerin recovered a different pattern; resembling canonical miRNAs (**Figure 23 A and B**).

Retrieved and analyzed U3 CLIP data display two distinct densities. miRNPs (Ago, Dicer, and DGCR8) recovered the 5' domain; this region that interacts with the 18S and the 5' ETS of the pre-rRNA (**Figure 23 D** Meanwhile, the snoRNPs densities enhanced the 3' domain that holds the C', D, C and B boxes (**Figure 23 C**).

In summary, all this evidence exposed my inability to deny U3 snoRNA as a source of canonical miRNAs using NGS derived data. Consequently, I concentrated on U3 5' domain to disprove its functions as miRNAs using other experimental approaches and predictions inferred from the canonical miRNA paradigm.

#### **2.1.24.8 snoRNP recovers ACA45 and U3 derived miRNAs**

Surprisingly, zooming into the U3 5' domain, the sequences recovery from Dicer follows the typical guide and passenger strand structure of miRNAs, including a drop in the density that could be the loop in a hairpin substrate. Furthermore, the RNA folding analysis yielded a hairpin structure that matched Dicer U3 5' domain PAR-CLIP density (**Figure 24 A and B**) and the guide strand predicted by a trained shRNA expert (Dr. Gruber personal communication), fit with the recovered by Argonaute CLIP data.



**Figure 24. U3 snoRNA 5'-domain interacts with miRNPs**

Read distribution in U3 5' domain from miRNPs CLIP-seq and cellular data, (A) Read position as a function of the fraction maximum, U3 5' sequence motifs are depicted in colors. (B) Details of the folded 5' domain, guide, and passenger strand are drawn according to Dicer and Ago recovered profiles.

Briefly, I could not rule U3 out as a source of miRNAs, while I could show it here for the selected intronic C/D snoRNA. For that reason, a high-resolution tool is necessary to detect U3 endogenous miRNA activity, and to deny its posttranscriptional repression *in vivo*.

To this point I reported for C/D snoRNAs that:

- The snoRNA loci produced miRNA size fragments, with a structure that resembles the guide and strand guide recovery of miRNAs.
- C/D-miRNAs are nuclear and nucleolar.
- C/D-miRNAs are not bound to Ago but with the microprocessor unit DGCR8.
- U3 C/D snoRNA-derived miRNAs are bound to Ago, Dicer and DGR8.
- U3 C/D snoRNA-derived miRNAs are also found in the cytoplasm.

## SCRUTINIZING MIRNAS REPRESSION ACTIVITY AT THE SINGLE CELL LEVEL

miRNAs studies have increased in resolution thank to NGS, computational analysis, *in vitro* and kinetics studies (See 1.1.2.2 Introduction). This understanding allows system biologist to apply modeling and explain the intricated miRNA networks.

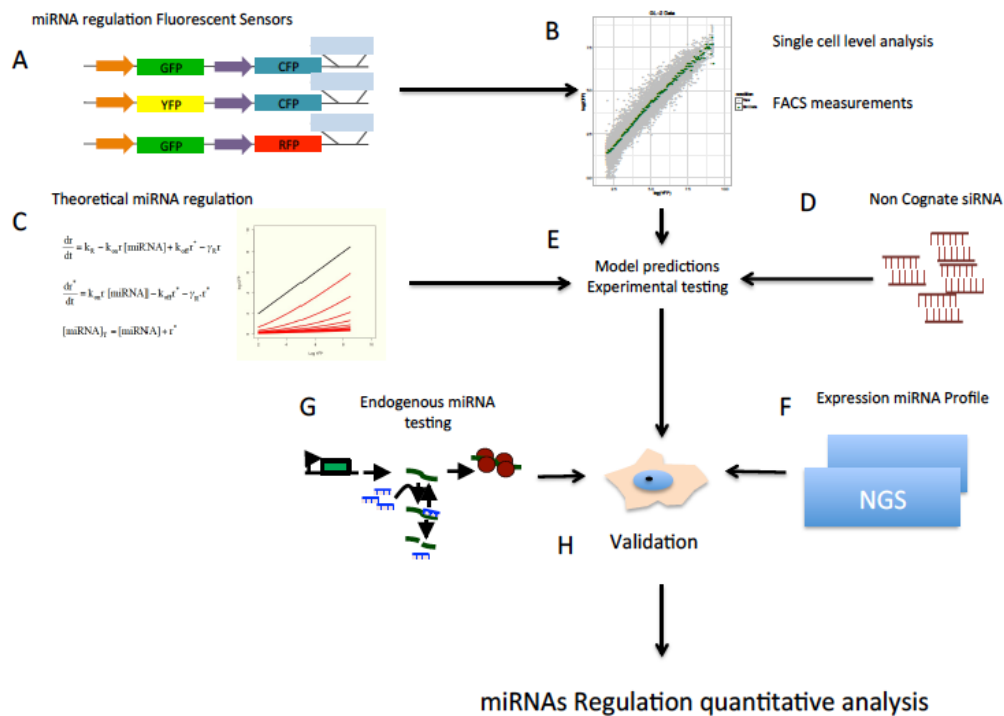
Despite these advances, the field lack of functional analysis with higher resolution. For instance, the broadly used luciferase assays give a small hint on miRNAs functionality because it provides an average functionality for the whole miRNA population. Moreover, luciferase-based assays have an expression-threshold that yields discernable differences for high expressed miRNAs, leaving out the vast majority expressed miRNAs and disregarding the fine-tuning action of miRNAs.

In the first section, I described that C/D-miRNAs concentration is low, making luciferase assay or quantifying RNA targets (mRNA) unsuitable. To test C/D-miRNAs endogenous activity, I decided to design my functional standard.

I reasoned that if a dual reporter sensor -like the luciferase, could be quantified at single cell level, It should increase resolution allowing to test C/D-miRNAs activity. I thought that FACS analysis would allow individual cells and specific fluorescence analysis. Thus, I interchanged the luciferase genes for fluorescent proteins.

I held that the new technique required a systematic development to give reliable results; therefore I generated the constructs and proofed its concept; evaluated and validated it for canonical miRNAs, and apply it for C/D miRNAs. **Figure 25** describes each performed step.

First I generated fluorescent plasmid and compared it with luciferase, and then I set up an R environment to use FACS outputs. Using a mathematical model I predicted different miRNA behaviors and I evaluated them with a siRNA as control of active RISC because siRNA use the same functional pathway, and finally, I validated the model for miRNA using the expression from Ion-Torrent profiles.



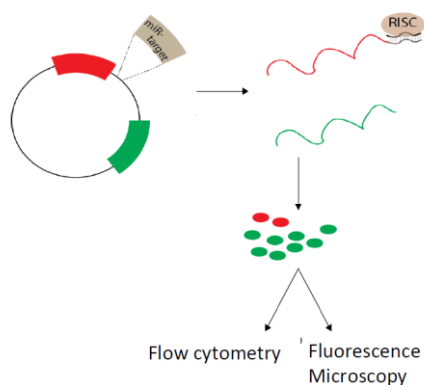
**Figure 25. Advanced dual fluorescent reporter for the comprehensive, high-resolution analysis of miRNA based regulation.**

Generation, evaluation, and validation of a fluorescent reporter for miRNAs functionality. Outline describes three steps: 1. Biologics (A) Generation of constructs with distinctive fluorescent proteins and (B) data acquisition and analytical processing, - analytical function generation. 2. Data Analysis: (C) Mathematical description of miRNAs regulation regarding  $K_{on}$ ,  $K_{off}$ , and miRNA concentration, (D-E) *in-vivo* predictions test with non-human targeted siRNA. 3. Validation: (F) Expression levels from small RNA NGS libraries and (G) analytical function test for endogenous miRNAs using the fluorescent reporter, (H) quantitative analysis of miRNA regulation *in vivo*.

#### 2.1.25 GENERATION OF DUAL FLUORESCENCE PLASMIDS: PROOF OF PRINCIPLE.

##### 2.1.25.1 Whole population analysis and comparison with luciferase based assays

The simple principle of the system is to use a single plasmid with two fluorescent proteins; it uses one as a sensor and one as a reference, both will be transcribed, while the ribosome translates the patron, the miRNA machinery inhibits the translation. In the end, the difference sensor-reference would serve as miRNA functional quantification (Figure 26).



**Figure 26 Untranscribed trans assay: theoretical principle**

The reporter system is based on two fluorescent proteins with different excitation and emission spectrum; one is used as transfection reference and the other as miRNA sensor. Upon transfection mRNAs is transcribed, translated and the relation between the two reporters describes the activity of endogenous miRNAs.

This principle is analogous to the F/R from luciferase assays, so I took its backbone and replaced it with different fluorescent proteins, I generated three constructs with CFP/GFP, CFP/YFP, and RFP/GFP combination. Table 1 summarizes the generation strategy.

As proof of principle, a transfected the plasmids, analyzed them by microscopy and flow cytometry, and compared it with the luciferase assays. For the sensor proteins, I used two psi-Check constructs (miRNAs miR-23a-3p and miR-27a-3p, produced by Henrike Höke), and I introduced the exact sequences on CFP or RFP.

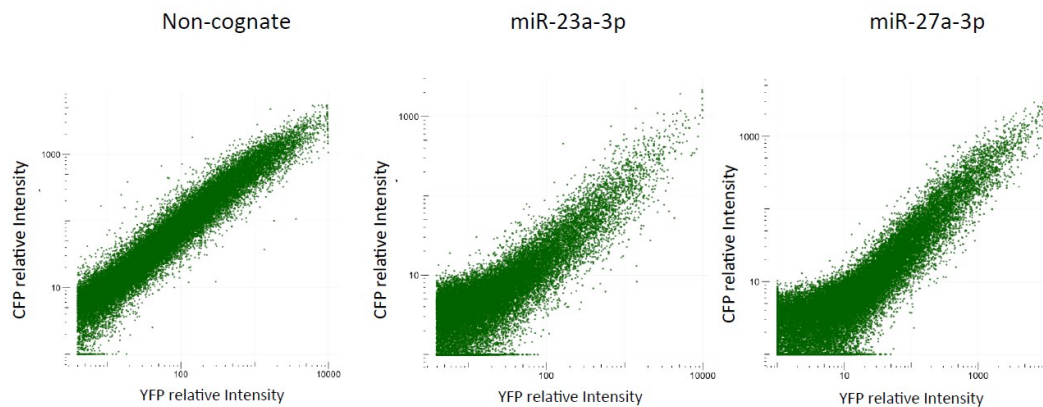
The microscopy results showed that the miRNA insertion on RFP reduced its expression comparable to non-cognate insert, while GFP expression is similar on the three of them. Relative intensities were calculated using the ImageJ software, and the ratio of both proteins calculated ( Supplementary Figure 8). Boxplots show that the miRNA insertions have a lower RFP/GFP ratio than the non-cognate control.

FACS analysis of the CFP-YFP construct show similar pattern, the scatter plot indicate that the CFP and YFP are proportional in all the constructs, while the non-cognate control displays a linear trend, the miRNA samples resembles a quadratic equation (Figure 27).

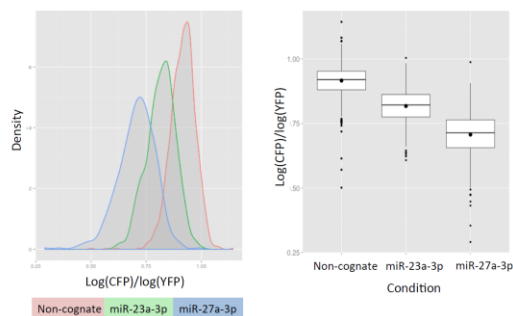
Density plots of CFP/ YFP ratio show a normal distribution, while the lowest average was for miR-27a-3p, it has the higher standard deviation (Figure 22 B); furthermore, the whole population analysis for fluorescence displayed similar trend to the luciferase R/F assays.



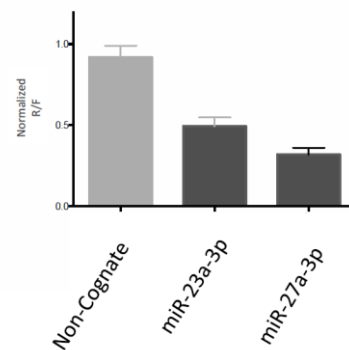
A



B



C



### Figure 27 Dual Fluorescence Reporter characterized in-vivo miRNA activity

The reporter uses two independent fluorescent proteins expressed individually from two different promoters. CFP contained a target region for miRNAs within the 3' UTR, while YFP is unaffected. Human Hek293 cells were transfected with three constructs (miR-23a-3p, miR-27a-3p and Non-cognate as control) and evaluated 72 h after. (A) FACS Scatter plots measurements. (B) Bulk quantitative analysis using CFP/YFP ratios, as density and boxplot, were calculated using custom R scripts. (E) Standard luciferase assays from Hek293, psi-check constructs were transfected with miRNA 23a-3p and miRNA 27a-3p targets located within the 3'UTR of Renilla luciferase and evaluated after 72h, bar plots show the ratio between Renilla and Firefly luciferase <sup>4</sup>\*

#### 2.1.26 SINGLE CELL LEVEL ANALYSIS: MOLECULAR TITRATION MODEL FOR MIRNA INDUCED POSTTRANSCRIPTIONAL REGULATION

##### 2.1.26.1 Mathematical models for sRNA regulation

<sup>4</sup> \*Data analysis performed together with Michael Mitter during his Lab Rotation Molecular Biology program under Lemus Supervision).

Looking at the scatter plots from the fluorescent reporters, I noticed three characteristics:

- The transient transfection creates a broad range of intensities for YFP and CFP.
- The non-cognate control shows a linear trend, where CFP increases similarly to YFP.
- miRNA-target insertion kept the both proteins proportional, but I recognized a deviation from linearity.

I thought that the FACS outputs could be understood using mathematical expressions. Initially, I tried with a simple nonlinear fit, but parameters outputs didn't give any intuitive meaning (data not shown). For that reason, I looked for distinct mathematical models and found that regulatory networks and molecular titration models offered a suitable alternative to interpret raw FACS data.

Molecular titration in regulatory networks<sup>117</sup> offered a good alternative, titration occurs when an active molecule (X) is sequestered into a non-active complex by a titrating molecule (Y). When Y concentration is larger than the  $K_d$  of the complex, it works as a buffer for the accumulation of X, so when X equals Y, overcomes the Y buffer activity and increases rapidly creating a threshold response.

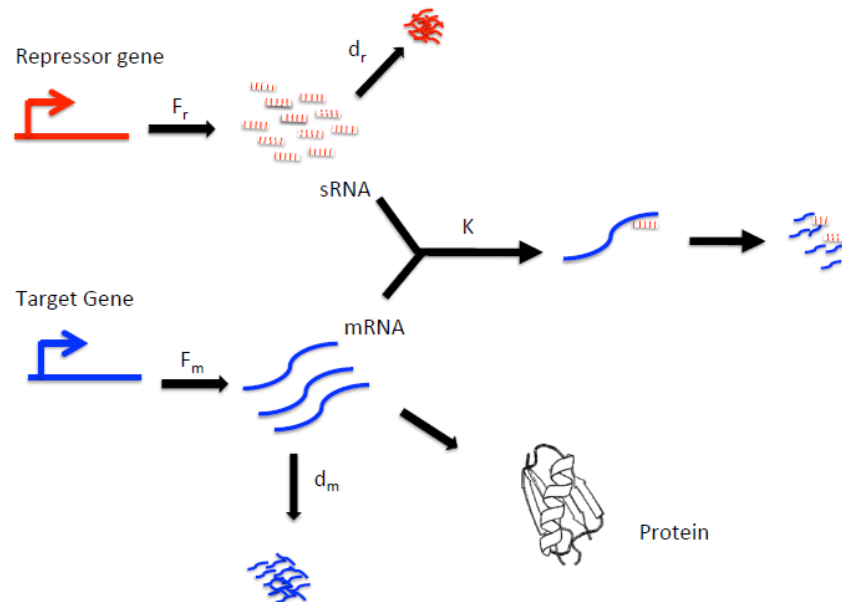
Briefly, the threshold response means (analogous to a buffer in pH titrating) that at a concentration below the threshold X changes are buffered (sequestered) by Y, and miRNAs certainly, behave as sequesters of mRNAs. I decided to implement this molecular titration into my reporter system and found that similar system has been used to define switch-like and tune activity for miRNAs and sRNA<sup>37,44,117,118</sup>.

For the sake of simplicity, I will describe in this section a simplified model based on bacterial sRNA regulation and simulate data, explaining the concept of threshold-linear behavior. Then I will describe and simulate a molecular titration model for miRNAs, defining structured predictions. In the next section, I will experimentally test these predictions and validate them using endogenous miRNAs.

To keep coherence I will follow the structure of scientific papers using mathematical models, for it, I will describe in results the model, the set of ODE and the results of simulations.

#### **2.1.26.1.1 sRNA regulation in Bacteria: Repression Threshold mathematical definition.**

I wanted to apply mathematics and apply it to my single cell analysis; not as system biology study but a tool to functionally characterize endogenous miRNAs, for its simplicity and to define threshold-linear behavior I will explain the model from <sup>118</sup> (Figure 28).



**Figure 28 Small RNAs gene regulation: the simple picture.**

Simplify model of sRNA regulation. The cell transcribes mRNAs and sRNAs at  $F_m$  and  $F_r$  rates respectively; free mRNA could be translated, sRNA repressed or decayed at  $d_m$  rate, the relation between  $F_r$  and  $F_m$  depends on the binding constant between sRNA and mRNA, showed here as  $K$ .

The mathematical model defines the following sets of ODE:

**Equation 1**

$$\frac{dr}{dt} = F_r - d_r r - k m r$$

**Equation 2**

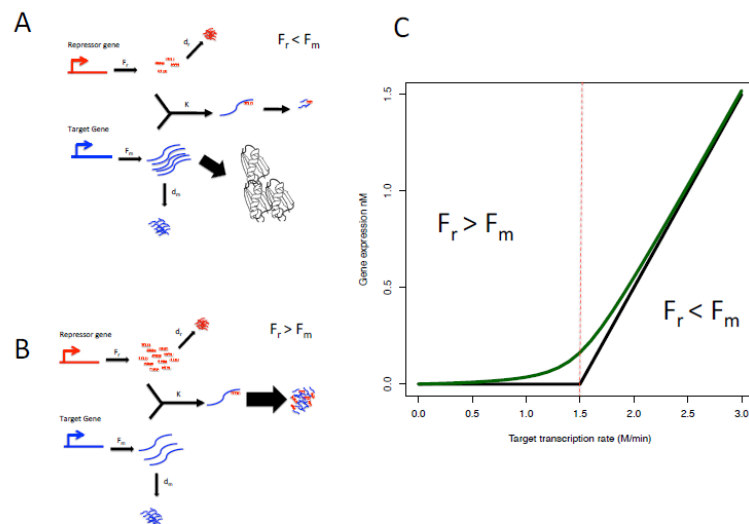
$$\frac{dm}{dt} = F_m - d_m m - k m r$$

**Equation 3**

$$m = \frac{1}{2d_m} \left( F_m - F_r - \phi + \sqrt{(F_m - F_r - \phi)^2 + 4F_m\phi} \right) \quad \text{where } \phi = d_r d_m / k$$

Equation 1 Describes the change in time for of the mRNA; it is transcribed at a rate  $F_m$  and degraded at  $d_m$ , while the second equation defines the expression of the repressor sRNA, with a similar set of rates  $F_r$  and  $D_r$ . The complex sRNA-mRNA has a second order kinetic constant  $K$  (Figure 28).

Solving equations 1 and 2 for steady state, drops Equation 3. Using mathematical simulations for several transcription rate  $F_r$  (M/m) and gene expression  $m$  (nM of protein), assuming that all the mRNA is translated. We can observe two scenarios in the sRNA induced "buffering" (Figure 29. When the mRNA concentration is higher than sRNA ( $F_r < F_m$ )(Figure 29 A) we get a linear increase of protein (Figure 29 C); whereas, in the case ( $F_r > F_m$ ), no protein is detected (Figure 29 B and C). This behavior is what I will call repression-threshold.



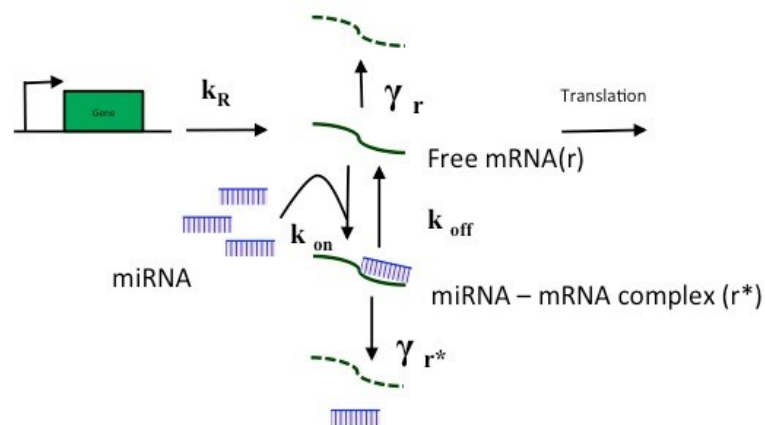
**Figure 29 Threshold linear response of a target mRNA**

Small-RNA's regulation idealized model describes the relationship between the transcription rate ( $F_m$ ) and protein concentration through an sRNA response function. (A) If  $F_r$  (sRNA) is lower than  $F_m$  (mRNA), the free mRNA is translated. While, (B) If  $F_r$  is higher than  $F_m$ , the mRNA is degraded reducing protein yield. (C) Simulation plot depicts mRNA transcription rate (M/min) versus gene expression (nM). If the sRNA is higher than the mRNA, it represses protein expression, while if the mRNA concentration is over the sRNAs, the relation between the transcription rate and the gene expression will be proportional. The start of linear behavior defines a repression threshold, at idealized conditions is equal to the sRNA concentration (Black line), while in a biological system, it has a gradual onset of linearity and the association constant ( $K$ ) governs the curve shape (green line).

In idealized conditions all the sRNA degrades the mRNA, the linear function intercept will be equal to the sRNA concentration and the repression-threshold. However, a biological system is not ideal, so the linearity on is gradually governed by K.

### 2.1.26.2 Molecular titration model for miRNAs posttranscriptional regulation

After defining repression threshold and describing the threshold-linear behavior with a simple sRNA model, I used a more complex set of ODE for miRNAs, described by Mukherji <sup>44</sup>; the model is depicted in Figure 30



**Figure 30. Titration model of miRNAs induced regulation**

The model for miRNA's regulation describes the free mRNAs available for translation. In the absence of miRNA, the cell transcribes and degrades mRNA at rates  $k_R$  and  $\gamma_r$  respectively. In presences of miRNAs, the free mRNA available for translation depends on the miRNA-mRNA ( $r^*$ ) complex, its formation depends on on-rate ( $k_{on}$ ) and miRNA concentration. Then two mutually exclusive paths occur. Either it induces mRNA degradation (at  $\gamma_{r^*}$  rate), or disassembles into its components (at  $k_{off}$  rate). The first order equations that describe the model are in the text.

The model describes the titration of miRNAs of mRNA, the complex formation  $r^*$ , depends on the  $K_{on}$  and  $K_{off}$  of the complex, that depends on the miRNA complementarity and targets in the mRNA. The following equations described the change of the mRNA free to translate and the miRNA complex.

**Equation 4**

$$\frac{dr}{dt} = k_r - K_{on}r[miRNA] + K_{off}r^* - \gamma_R r$$

**Equation 5**

$$\frac{dr}{dt} = K_{on}r[miRNA] - K_{off}r^* - \gamma_R r^*$$

In this context, total miRNA concentration is considered as the free miRNAs plus the miRNAs included in the complex ( $r^*$ ):

**Equation 6**

$$[miRNA]_T = [miRNA] + r^*$$

It is assumed that no translation is happening from miRNA bounded RNA, and only from the free RNA ( $r$ ). And solving for the steady-state levels of  $r$  as before <sup>44</sup>:

**Equation 7**

$$r = \frac{1}{2} \left( r_0 - \lambda - \theta + \sqrt{(r_0 - \lambda - \theta)^2 + 4\lambda r_0} \right)$$

Where =

$$r_0 = \frac{K_R}{\gamma_R} \quad \lambda = \frac{\gamma_{R^*} + K_{off}}{K_{on}} \quad \theta = \frac{\gamma_{R^*}}{\gamma_R} [miRNA]_{total}$$

So, for the sake of simplicity, I will use the term miRNA binding for  $\lambda$ , even though it could have more sense to use as synonym of Kd. However, lambda has more elements and probably can confuse the educated if I use  $\lambda$  as Kd.

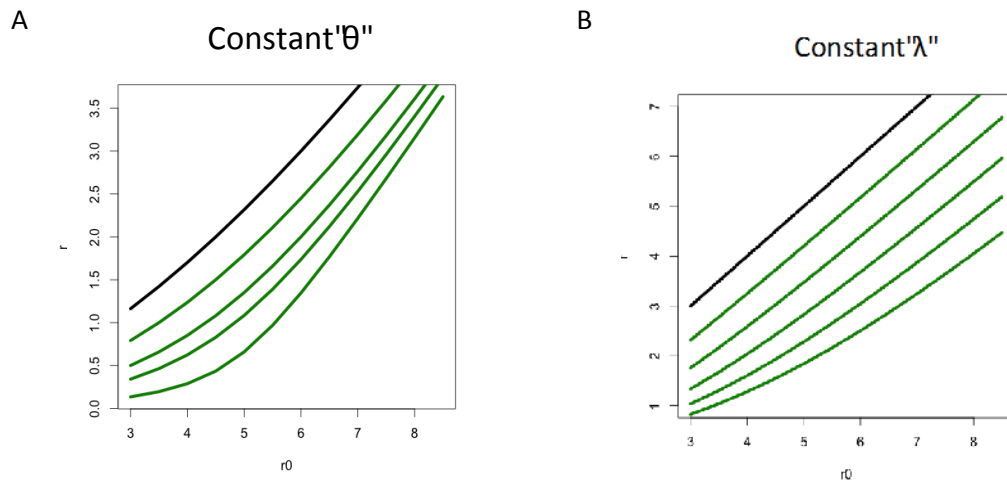
**2.1.26.3 Molecular titration model simulations**

Since the threshold-repression model matches qualitatively the fluorescent reporter scatter plots (**Figure 27**), I thought to create a set of predictions to test how well it describes the *in vivo* FACS outputs.

I simulated values of  $r$  (mRNA free for translation) using the equation (Equation 7), with increasing values of  $R_0$  (mRNA without miRNA regulation), and varying miRNA concentration ( $\theta$ ) and binding ( $\lambda$  ( $\sim 1/K_{on}$ )).

Results of two scenarios resemble well the scatter plots results when miRNA is kept constant,  $1/\lambda$  increasing values introduce a sharp in the repression threshold, defining that the binding of the miRNAs  $\lambda$  controls the threshold-linear onset. And the miRNA levels control the regulation intensity, so when threshold-linear onset is determined, the increasing miRNA concentration  $\theta$  will shift the curve (Figure 31).

Admitting that the analogies have limits and the weakness, we can say that  $\lambda$  is like the slope and  $\theta$  the intercept in linear regression, however from the simulation it is clear how each parameter influences the curve shape.



**Figure 31 Numerical simulations describe miRNA-induced thresholding**

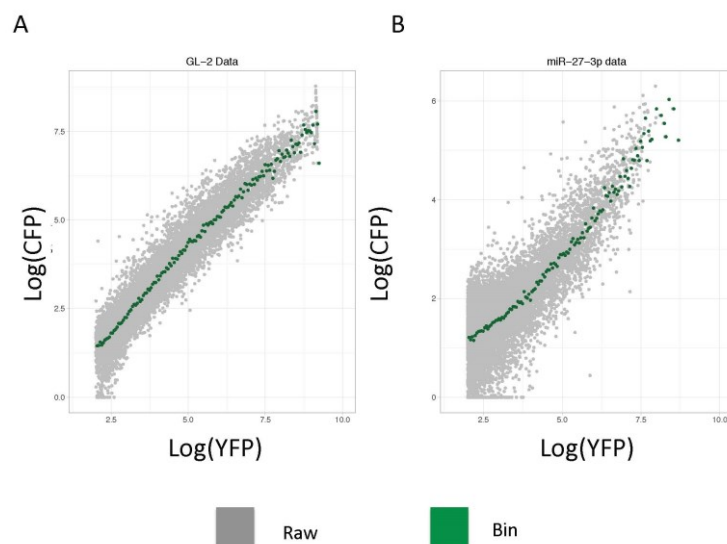
Steady state solution for free mRNA ( $r$ ) as a function of initial mRNA transcription ( $r_0$ ). The solutions yield to lumped parameters (Equation 4):  $\lambda$  inverse proportional to  $K_{on}$ , and  $\theta$  equivalent to the miRNAs concentration. Numerical simulation of steady state solutions for (A) several  $\lambda$  ( $1/K_{on}$ ) and constant  $\theta$  (miRNAs), and (B) for increasing amounts of miRNA ( $\theta$ ) and constant  $\lambda$  ( $1/K_{on}$ )<sup>4\*</sup>.

#### 2.1.26.4 Molecular titration model test experiments

I wanted to simplify the output from the raw data from single cells, therefore I generated a processing data and visualization tool in R, and called the result analytical function (**Figure 32**)

The R environment creates the analytical function by transforming the raw data into a simple line, for that, it divides the YFP positive cells in small bins of 0.05, and calculates the descriptive statistics per bin. Then it uses this value to create a curve. Briefly, each dot in the analytical function is the average value CFP value for a range of 0.05. The transfection standardization and negative controls showed that 72 h and just one microgram yield reliable analytical functions (Supplementary Figure 9 and Supplementary Figure 10).

<sup>4\*</sup> Simulation can be reproduced at <https://medicalrnabiology.shinyapps.io/pUTApp/>



**Figure 32 Dual fluorescent reporter analytical flow cytometry: Function generation from single cell measurements**

Analytical processing of transfected cells starts selecting YFP expressing cells (Grey dots), only events greater than two Standard deviations from the negative control (non-transfected cells) are considered. The function (Green dots) represents the mean from binned intensities -1 to 8 at 0.05 intervals. As an example, Hek293 cells were transfected with two reporter plasmids, and FACS analyzed after 72 h. Analytical functions depicted for (A) Non cognate and (B) miR-27-3p targeted reporters.

#### 2.1.26.4.1 *Lambda and theta experiments with GL-2*

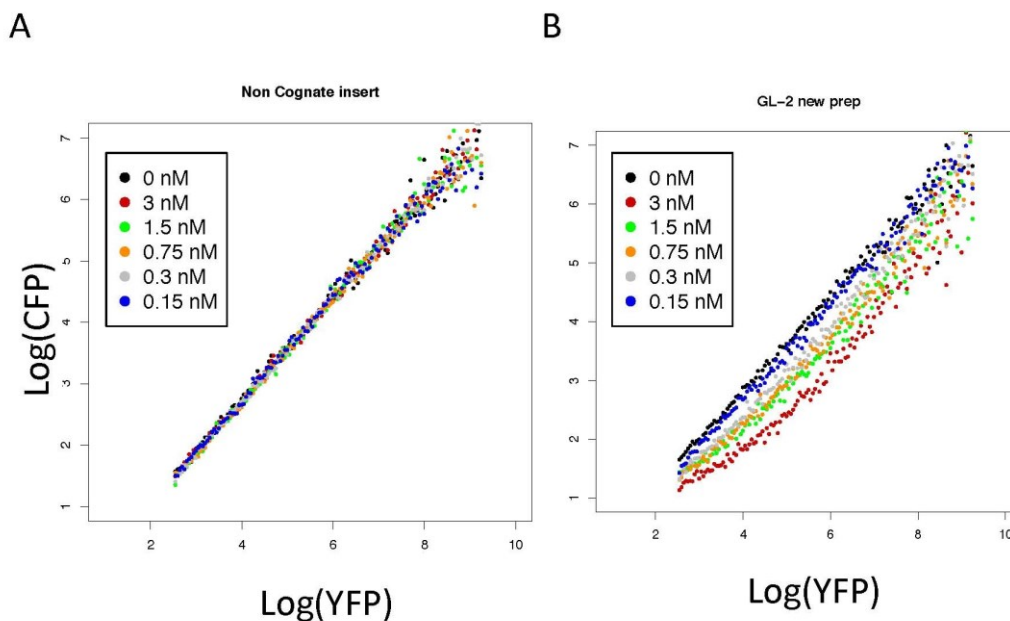
In the previous section, I described that the parameters  $\theta$  and  $\lambda$  influences the shape of miRNA titration curves;  $\theta$  represents miRNA concentration and shifts the titration line down, while modulating  $\lambda$  ( $1/K_{on}$ ) sharpens the threshold-linear onset, so that lower values give a distinctive transition to linearity.

I reasoned that if the model applies to the fluorescent system it should be tested it empirically, so using a siRNA (GL-2 that targets luciferase gene) at increasing concentration I tested whether siRNA increments behaves as the predictions for  $\theta$ .

To inquire  $\lambda$  predictions, I modified the target sequence in CFP 3'-UTR, to increase the miRNA-mRNA complex formation I inserted one and three perfect copies of GL-2 target sequences, and to reduce miRNA-mRNA interaction, I added incomplete GL-2 sequences and test it with siRNA transfection.

Co-transfection experiments (siRNA-fluorescent sensor) show that upon siRNA increasing concentrations the analytical functions shift down for the GL-2 targeted CFP (Figure 33 B and Figure 34 B), without changes for the non-targeted control (Figure 33 A and Figure 34 A). I observed similar behavior for three and incomplete sequence (Figure 34)





**Figure 33 Exogenous expression of siRNA shift analytical functions *in vivo*.**

The miRNA-regulation model describes  $\theta$  as proportional to the total miRNA levels and the numerical simulations (Figure 31) show function shift upon increasing  $\theta$  values. Increasing concentrations of a non-cognate siRNA (GL-2 Control) were co-transfected in Hek293 with two dual fluorescent reporters, (A) negative control and (B) GL-2 targeted, analytical-functions derived from FCAS were produced after 72 h post transfection. Scatter plots show representative data from at least three independent experiments (Representative of at least 3 experiments).

Remarkably, modulating the target sharpens the analytical function, the co-transfection of modified GL-2 sites (bulged) shows a small shift in the analytical curve, and three copies have a clearer threshold-linear onset than one or bulged sensors.

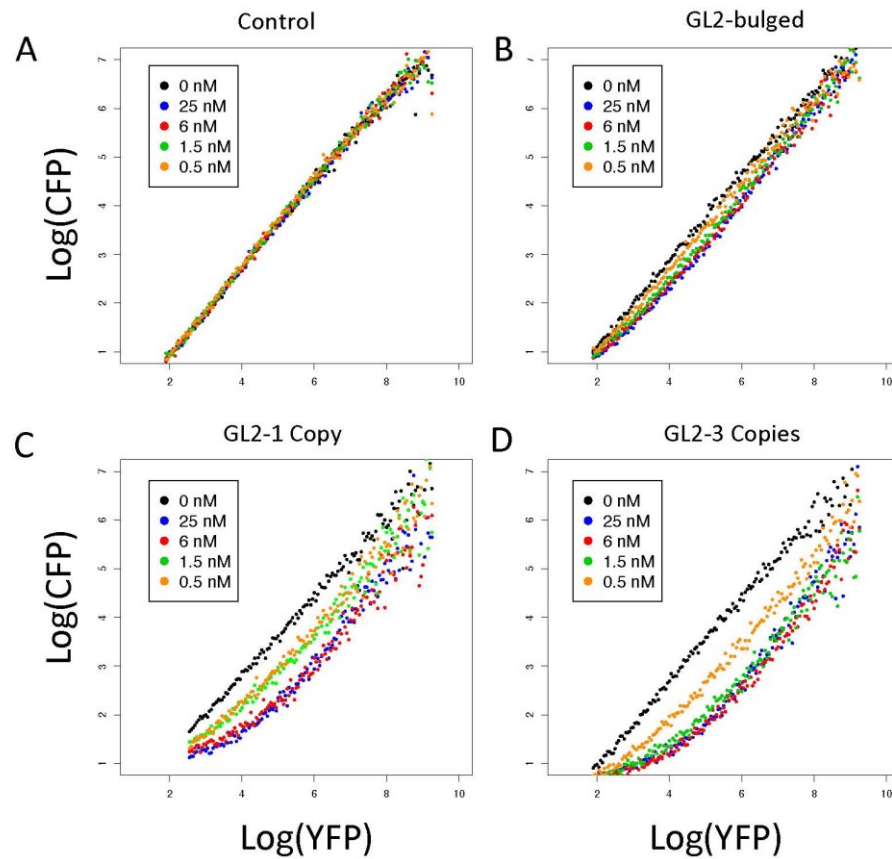
Briefly, I considered that I could not deny that the molecular titration model can be used for the FACS analysis; so, we decided to call the complete analysis (FACS and R environment), as analytical flow cytometry for miRNA function *in vivo*. Contemporary to this thesis Bartel and Stoffel used also the term analytical flow cytometry<sup>119</sup>.

### 2.1.26.5 Insight on miRNAs posttranscriptional regulation

#### 2.1.26.5.1 siRNA experiments yield three functional groups.

The analytical curves from GL-2 experiments resemble the titration model predictions for increasing miRNA concentration and binding capability. So, I decided to apply it to miRNAs, but with a systematic perspective derived from the model.

Since expression is the often-analyzed parameter from miRNAs and defines the "repression" ability, I observed all the analytical results and established three categorical variables. To clarify miRNA functionality, applying the effects of  $\theta$  and  $\lambda$ .



**Figure 34. Increasing reporter binding sites sharpens siRNA induced threshold *in vivo*.**

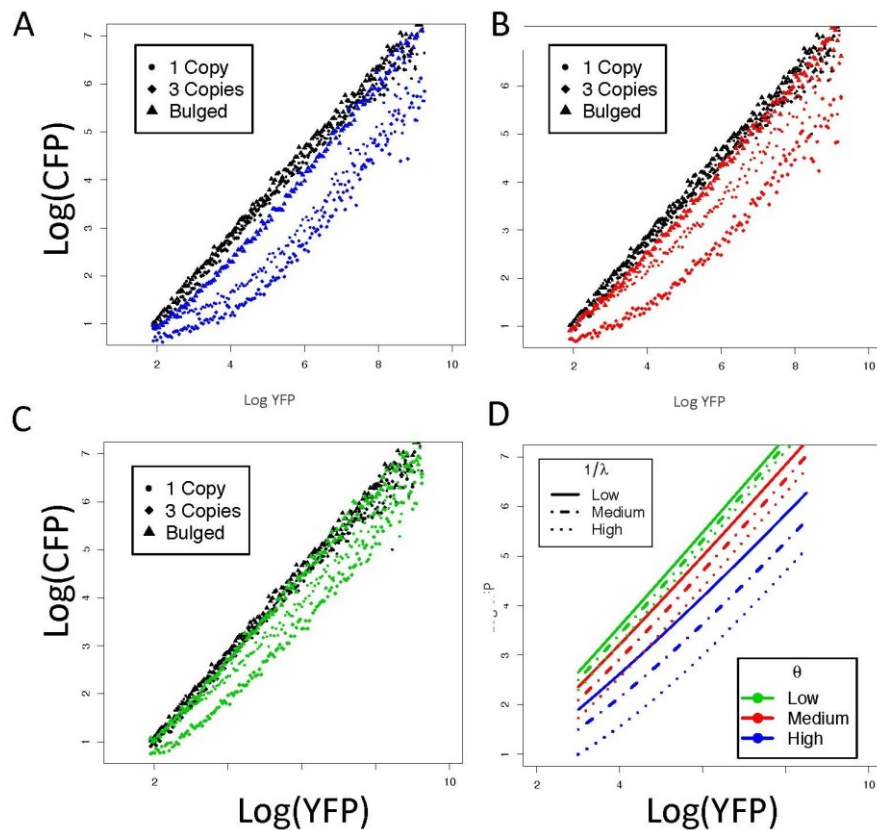
The miRNA-regulation model describes  $\lambda$  as inversely proportional to  $K_{on}$ , and numerical simulations (Figure 24) show threshold sharpening after decreasing  $\lambda$  values. To artificially increased or decreased binding events, several non-cognate (GL2) targeted sites were inserted into the dual fluorescent reporter. Constructs were co-transfected with increasing siRNA-GL2 concentrations and analyzed by FACS; Scatter plots depict an analytical function for (A) Negative control (B) Bulged GL-2 (18 complementary plus three unpaired bases, and for (C) single and (D) three perfect GL-2 complementarity copies. (Representative of at least 3 experiments).

These categories (Qualitative variables), which are defined by the curve shape and position, are high, mid and low functional. To simplify the description (Experimental and *in silico*), I will describe the curve shape-position, parameters estimation, and simulation, and conclude with an integrated definition for the groups.

These categories (Qualitative variables), which are defined by the curve shape and position, are high, mid and low functional. To simplify the description (experimental and *in silico*), I will

describe the curve shape-position, the parameters estimation, and conclude with an integrated definition for the groups using an empirically based simulation.

For high siRNA concentration (blue), all " $\lambda$ s" reduce CFP expression being three and one closer in their repression curve. For a mid siRNA concentration (red) one and bulged copies repress CFP similarly, and for small siRNA concentration (green) the three positions are closer to the controls (**Figure 35**).



**Figure 35. *In vivo*-derived analytical functions describe three functional groups**

*In vivo* analytical functions from GL-2 siRNA experiments define three categorical variables based on siRNA concentration and binding sites: low, mid and high functional. The scatter plots depict three siRNA concentrations - (A) 20 nM (B) 1.5 nM and (C) 0.5 nM and three GL-2 targeted constructs - Bulged, one and three copies. To define  $\theta$  and  $\lambda$  ranges we fitted the analytical functions from GL-2 (siRNA) using non-linear regression, (D) numeric simulations from *in vivo* derived parameters; the plot shows the steady state solutions for nine combinations of  $\theta$  (Colors) and  $1/\lambda$  (lines).

For the curve shape, high siRNA concentrations depict a threshold-linear onset and keep it for three copies at 1.5 nM of siRNA, while no thresholds appear at low siRNA concentration (Figure 35 A-C).

Using the intensity ranges for YFP, I used nonlinear least squares to estimate  $\theta$  and  $\lambda$  using the analytical cytometry data and defined the limit, which can be used for estimation in experimental data, to simulate curves based on the actual experimental window.

The simulation shows a trend that justifies the categorical classification (Figure 35 D), showing that the bulged-siRNAs functions have no observable threshold and depict a lateral shifting. Perfect one and three copies exhibit threshold differences, while for the high functional group the curves are nonlinear, the middle group curves display an intermediate shape.

In model terminology, the low functional miRNAs own a small mRNA repression within several  $1/\lambda$  at low  $\theta$ , showing no threshold and shifting the intercept of the linear behavior.

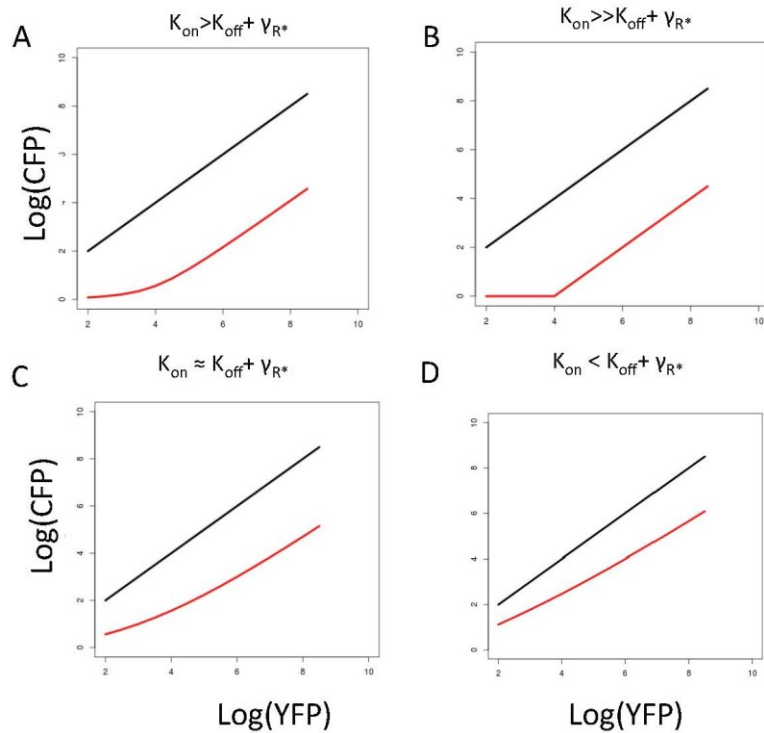
For the high and mid functional groups, the curve shape depends on  $1/\lambda$  and  $\theta$  cooperation. Meaning that high values for  $1/\lambda$  drag medium-expressed miRNAs into the high functional group (blue), while, highly expressed and low  $1/\lambda$ , bring the function close to the mid (red) group (Figure 35 D).

#### **2.1.26.5.2 Insight on the miRNA titration model**

Till now, I described and simulated a miRNAs molecular titration model, I tested its prediction experimentally and proposed three functional groups and re-ran the simulation with estimated factors. I wanted to apply this classification and modeling to endogenous miRNAs to analyze functionality low expressed miRNAs, but also to distinguish in detail responses of high expressed miRNAs.

Furthermore, to define and understand any behavior from endogenous miRNA, I run several simulations to clarify distinct scenarios, creating a deeper understanding *OF THE MIRNA TITRATION MODEL*, using first situations with different  $K_{on}$  and  $K_{off}$ , changing the concentration afterward.

The numeric simulation reveals the  $K_{on}$  influence upon the threshold on set (Figure 36 A - B). For  $K_{on}$  equal or lower than the  $K_{off}$ , the curves lose an identifiable thresholding (Figure 36 C - D), supporting that the miRNA-mRNA complex formation determines the threshold-linear transition.

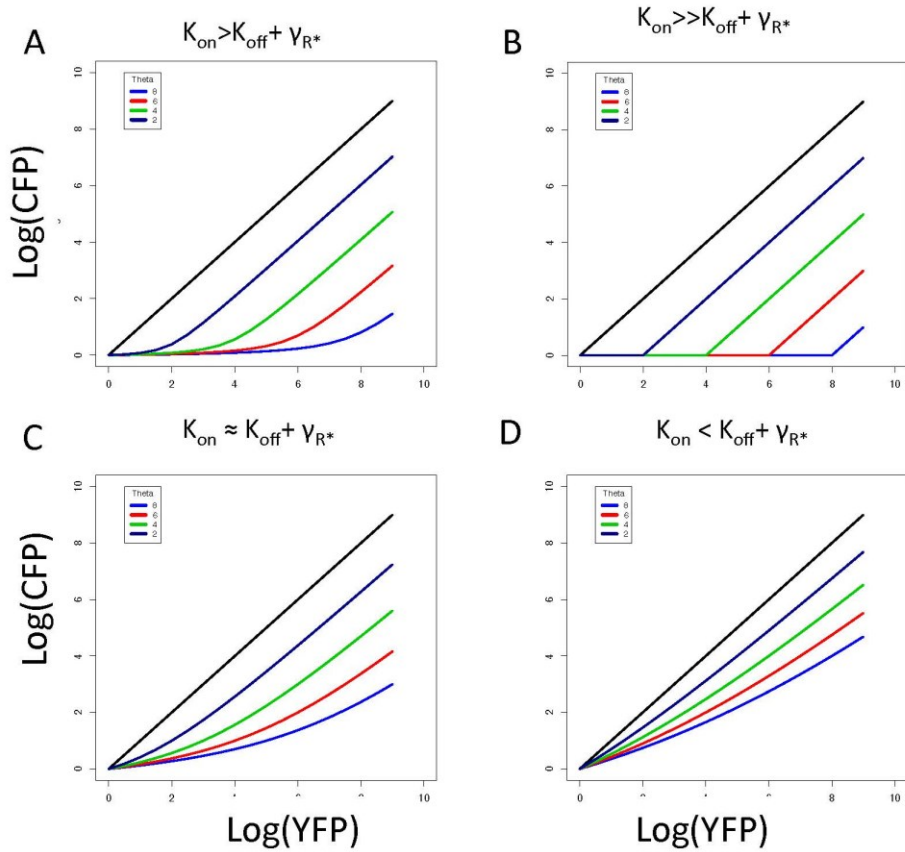


**Figure 36 .Threshold transition to linearity depend on miRNA-mRNA complex formation**

Plots depict numeric simulations performed using the *in vivo* derived  $\lambda$  range, four scenarios defined by  $K_{on}$  and its relation with  $K_{off}$  and  $\gamma^*r$ . In A the plot shows the threshold transition at high  $K_{on}$  values, and B depicts a situation with an ideal miRNA, where the  $K_{on}$  extremely high, and the dissociation ( $K_{off}$ ) is close to zero. C describes the function shape in conditions where  $K_{on}$  is similar to  $K_{off}$  and D depicts the function form when  $K_{on}$  is lower than  $K_{off}$ . Simulation can be reproduced at <https://medicalrbiology.shinyapps.io/pUTApp/>

The increasing concentrations ( $\theta$ ) show that high  $K_{on}$  sets the thresholds closer to the miRNA expression, indicating that just in ideal conditions the expression equals repression (threshold) (Figure 37 A-B), but in biological systems, the threshold-linear onset behaves dynamically.

Increasing miRNA concentrations of  $K_{on} \approx K_{off}$  show experimentally undetectable YFP threshold-linear –Compare YFP axis of simulations and analytical functions. Finally, the linear onset for  $K_{on} < K_{off}$  happens at small YFP intensities, anticipating that empirically the analytical cytometry would just depict the linear section of the threshold-linear behavior.



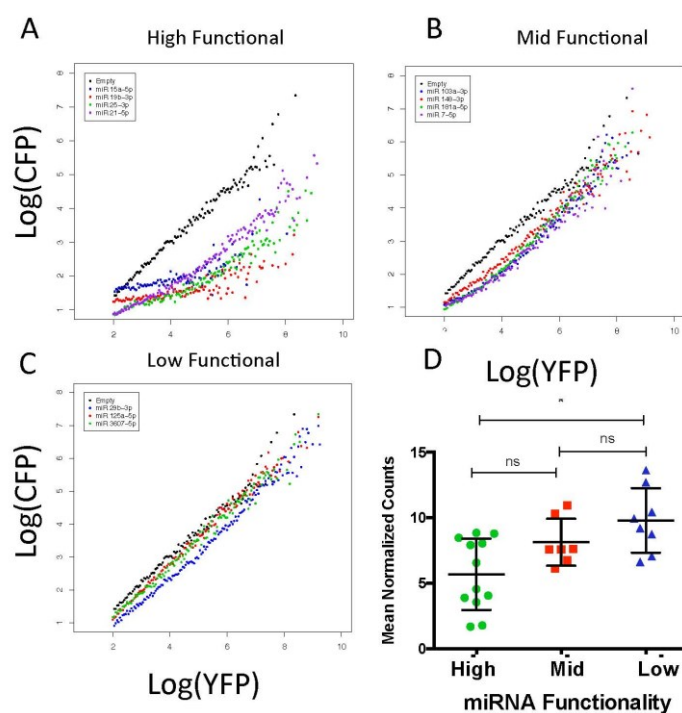
**Figure 37 Threshold limit is controlled by miRNA expression.**

Plots depict numeric simulations performed using the *in vivo* derived  $\lambda$  and  $\theta$  range, four scenarios defined by  $K_{on}$  and its relation with  $K_{off}$  and  $V_{R^*}$ , all the graphs have four  $\theta$  values. In A the plot shows smooth threshold transition for high  $K_{on}$  values, while B depicts ideal miRNA, where the  $K_{on}$  is extremely high, and the change to linearity is abrupt. C describes the function shape in conditions where  $K_{on}$  is similar to  $K_{off}$  and presents a mild threshold transition. Finally, D depicts the function form when  $K_{on}$  is lower than  $K_{off}$ ; the threshold is smooth and at low YFP values. Simulation can be reproduced at <https://medicalrnbiology.shinyapps.io/pUTApp/>

## 2.1.27 MEASURING ENDOGENOUS MIRNA ACTIVITY AT THE SINGLE CELL LEVEL

### 2.1.27.1 Functional Test using analytical flow cytometry

I demonstrated that molecular titration simplifies and estimates factors out of fluorescent miRNA sensor; I described three categorical variables, which represent miRNA functionality (low, mid and high). Next, I applied it to selected miRNAs to examine endogenous miRNAs activity; using the small RNA Ion- Torrent. I chose several miRNA candidates within a broad expression range and tested them using analytical flow cytometry (**Figure 38**).



**Figure 38 miRNA Analytical functions uncouple expression from functionality**

Perfect complementary sequences for thirty selected miRNAs (with distinct expression patterns Figure 8) were inserted into CFP 3'UTR, transfected into Hek293 and evaluated after 72 h. Scatter plots depict analytical functions for three groups, according to the shape of their curve. (B) High functional that have a threshold up to high YFP intensities, (C) mid with a threshold at moderate YFP intensities, and (D) low functional with no observable threshold (Compare with **Figure 35**). (E) Functional groups were used as ordinal variables to compare the mean normalized counts between groups ( $P < 0.01$  Kruskal-Wallis test, Dunns multiple comparisons test  $* < 0.05$ ).

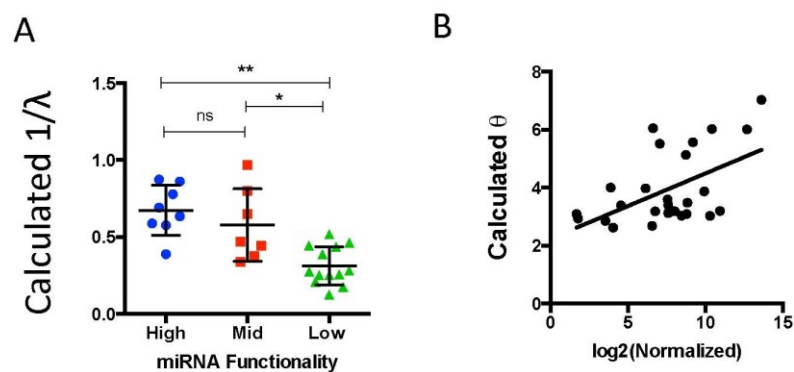
Initially, I tested candidates from all the quartiles (transformed reads) evaluating outliers and average miRNAs; noticeably I found (as the model predicted) an inconsistency between

expression (relative miRNA concentration) and the function shape. So I increased the measurements evaluating 30 different miRNAs.

The miRNA's analytical curves parallel the categorical classification derived from siRNA experiments; consequently, high functional miRNAs display a clear threshold, mid produce a gradual unsharp linearity onset; and the last group (low) exhibits a little shift resembling linearity.

Comparing the expression within each class (As anticipated by the titration model **Figure 35**), I observed no discrepancy between adjacent groups and significant difference between high and low functional groups, showing that expression begets no through repression for endogenous miRNAs. To explain it, I used nonlinear least squares and estimated  $\theta$  and  $\lambda$ ; to do so, I reduced the bin size (0.01) to increase data points securing accuracy in the estimates, then I compared the estimated factor with expression (Ion-Torrent profiles) and functionality (high, low and mid).

Calculated  $\lambda$  discriminates high and mid from low functional miRNAs, while  $\theta$  is proportional to Ion-Torrent expression profiles (**Figure 38** and **Figure 39**), indicating that differences between mid and high functional miRNAs, could be influenced by separate factors exceeding miRNA expression.



**Figure 39 miRNA titration model integrates expression and binding *in vivo***

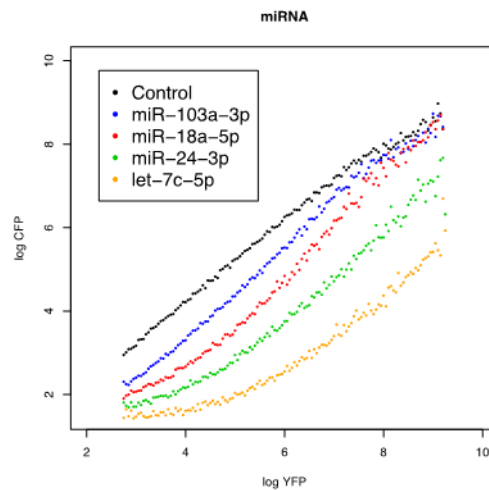
Produced analytical functions were fitted (non-linear least squares regression), to estimate  $\lambda$  and  $\theta$ . (A) Calculated  $1/\lambda$  ( $\sim$ Kon) plotted vs. different functional groups (P < 0.05 Kruskal-Wallis test, \* < 0.05 \*\* < 0.01). (B) Expression as mean normalized counts (Figure 8) vs. calculated  $\theta$  ( $\sim$ miRNA).

I reasoned that factors like ratio miRNA-targets or miRNAs subcellular location could reveal how expression deviates from functionality. Using subcellular RNA profiling of HeLa cells

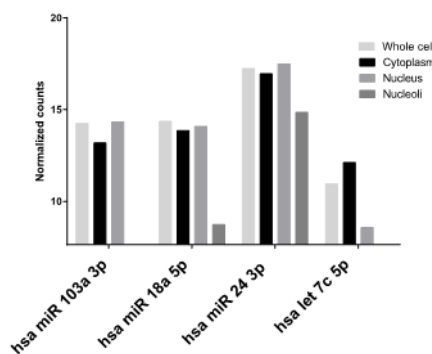


(3.1.1.3), and after observing that the analytical cytometry yield the same shapes and groups for HeLa (Supplementary Figure 11), I examined whether miRNA subcellular location influences the FACS functional outputs ( **Figure 40**).

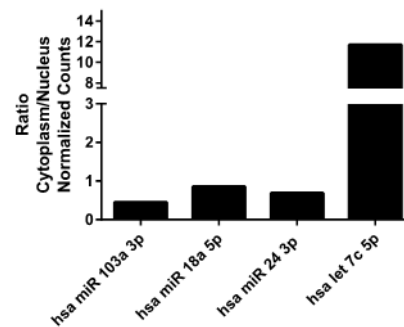
A



B



C



### Figure 40 miRNAs Subcellular localization influences their analytical output

Four individual miRNAs were chosen according to their cellular expression, 3 miRNAs on low and mid functional groups with expression higher than the high functional median (miR-103a-3p, miR-18a-5p, miR-24-3p) and one high functional miRNA lower than the mid functional median of normalized cellular read counts. **(D)** Analytical transfer functions **(E)** Bar plots for cellular compartments of small RNA libraries and **(F)** Cytoplasmic/Nuclear ratio for selected candidates.

Transfection figure A performed together with Maximilian Arlt during his Bachelor thesis under Lemus Supervision).

I selected two outliers: the first (miR-103a-3p) from the low functional miRNAs, which is over the percentile 75% of its cluster, and the second (let-7c-5p) from the lower expressed inside high functional miRNAs; at the same time, I assessed the intersection between mid-high miRNAs selecting miR-18a-5p and miR-24-3p.

The combination of analytical cytometry and subcellular profiling indicate that the miR-103-3p (highly expressed but limited functionality) holds a small cytoplasmic-nuclear ratio. Additionally, the analysis shows that miR-18a-5p and miR-24-3p own small cytoplasmic-nuclear proportions. Finally, the fourth candidate let-7c that performs as a high functional miRNA but mildly expressed is located solely in the cytoplasm.

In summary, these findings underline that analytical flow cytometry applies to other cells types, it can define alternative hypothesis based on its outputs, like that miRNAs subcellular location determines miRISC functionality, suggesting that small RNA profiles without systematic and detailed functional outputs could be misleading

### **2.1.27.2 Threshold modulation of endogenous miRNAs**

I showed that the miRNA titration model, together with the analytical outputs, resolves better the repression of endogenous miRNA, defining precise curves and allowing the analysis of new hypothesis when integrated with expression profiles.

As with C/D-miRNA experiments, I wanted to refute that *endogenous miRNA follow the same trend as the exogenous siRNAs experiments* (2.1.26.4), in other words, to inquire whether  $\theta$  and  $\lambda$  parameters apply to endogenous miRNAs. I used several target sequences for miR-27a-3p, introducing different copies (1 and 4 bulged -incomplete miRNA target and 1 to 3 perfect complementary) into the fluorescent sensor and examining them in Hek293 cells; to estimate parameters ( $\theta$  and  $\lambda$ ) using equation 7.

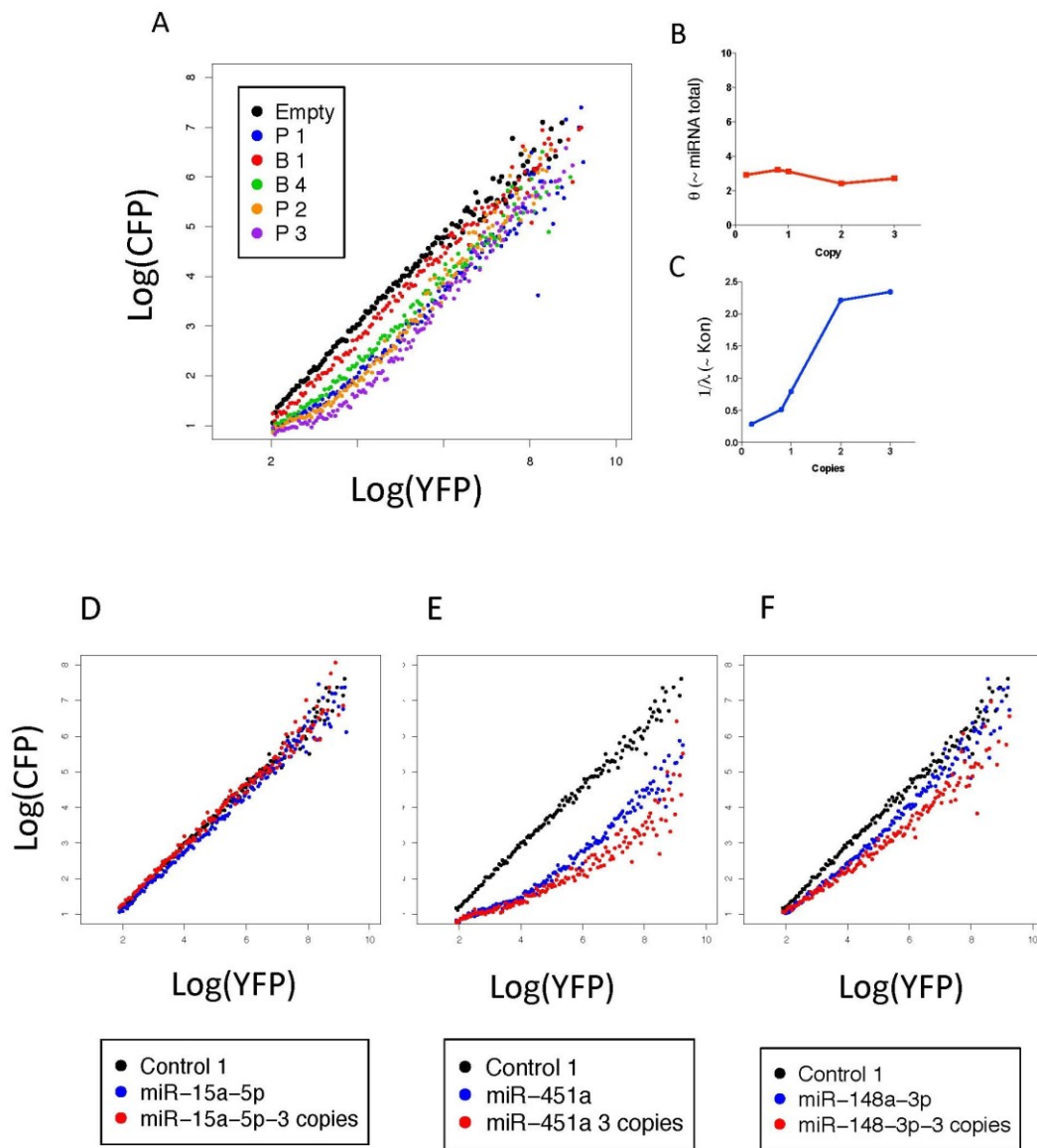
Adjusting miR-27a-3p target in the reporter increases the threshold sharpness (Figure 41 A), linking it to  $1/\lambda$  values, keeping  $\theta$  estimates without changes for all copies (Figure 41 B - C), this trend also applies to other miRNAs (Figure 41 D-E).

To examine  $\theta$  without the external use of miRNAs (exogenous expression of miRNA), I thought to change the active miRNA concentration inside the cell, using siRNAs (with and w/o target) to out-compete endogenous miRNAs shifting up the analytical function. In simpler terms, overloading the cells with sRNAs shall reduce the miRNAs working concentration.

Remarkably, co-transfection with siRNA shifted the analytical curve for three selected miRNA; reducing its effective miRNA amount, independently of siRNA target (GL-2 and LMNA) and working for all the groups (low, mid and high).

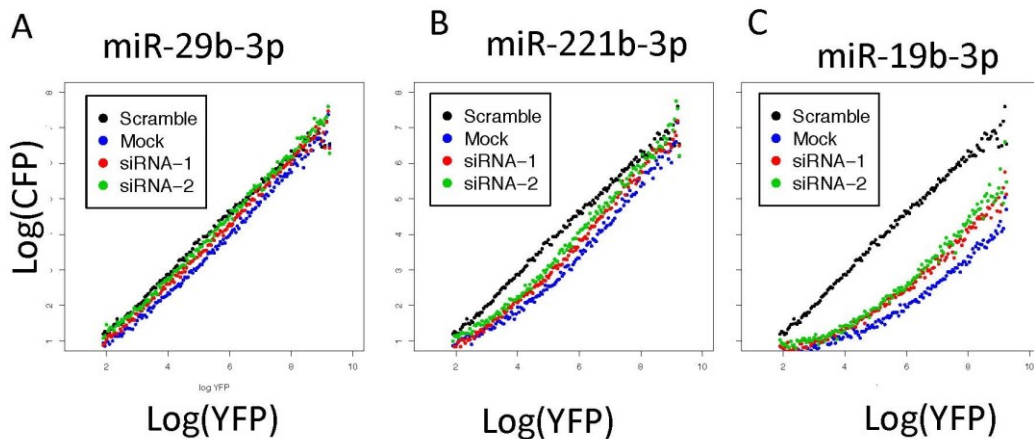
Remarkably, co-transfection with siRNA shifted the analytical curve for three selected miRNA; reducing its effective miRNA amount, independently of siRNA target (GL-2 and LMNA) and working for all the groups (low, mid and high).

Considering that human miRNA targets are not entirely complementary and defining whether the analytical cytometry is also valid for endogenous targets. I inserted the seed, seed plus motif and the complete targets for miR-92a (derived from CLASH analysis), to point out that the system is useful for endogenous 3' UTRs. For open-source sharing, the plasmids are located at Addgene.



### Figure 41. Threshold sharpening for endogenous miRNAs

To sharp threshold of miR-27a-3p several target sites, bulged and perfect, were inserted at 3'UTR of CFP, Hek293 cells were transfected and analyzed by FACS after 72 h. (A) Analytical functions for perfect binding sites ("P"- P1-P3) with one, two and three copies, and one and four copies of a bulged site ("B" -B1 & B4) -three unmatched bases-. Analytical functions were fitted using non-linear least squares to estimate  $\theta$  and  $\lambda$  values. Plots show miRNAs copy number vs. (B) estimated  $\theta$  (~miRNA) and (C)  $1/\lambda$  (~Kon); values for bulged were arbitrary set to 0.2. Three miRNA candidates for each functional group were selected and tested to sharp their threshold; plots show (D) low, (E) mid and (F) high functionality functions for one and three perfect copies.



**Figure 42. Threshold shift by outcompeting miRNA with siRNA *in vivo*.**

To modify  $\theta$  ( $\sim$ miRNA), but without changing endogenous miRNA expression, siRNAs were co-transfected with the reporters to outcompete miRNA, reducing the functional miRNA concentration. Three different miRNAs were selected, co-transfected with two different siRNA and evaluated after 72 h. Plots depict (A) low, (B) mid and (C) high functional miRNAs.

#### ENDOGENOUS ACTIVITY OF C/D SNORNA-DERIVED MIRNAS

Summarily I pointed that C/D snoRNA locus can produce RNA in the range of miRNAs, they're located mainly in the nucleus, and low expressed compared to canonical miRNAs. Furthermore, they are undetectable in Argonaute and Dicer RNA pull-downs but interact with DGCR8, suggesting that C/D miRNAs are not working as classic miRNAs.

However, 5' domain of the C/D snoRNAs U3 produces RNAs around 21 nt, it is found in the cytoplasm and interacts with all the miRNPs, but it is scarce expressed. Since no technique can quantify endogenous miRNA function *in-vivo* at high resolution, and the C/D-miRNAs are in few number, I developed an analytical assay that characterize low expressed miRNAs activity *in vivo*.

#### 2.1.28 ANALYTICAL FLOW CYTOMETRY OF C/D SNORNA-DERIVED MIRNAS THRESHOLD MODULATION.

To perform another disproval attempt for the C/D-miRNA, I inserted the complete complementary sequences for the C/D miRNA (U78, U44, and U3) in the fluorescent reporter.

As a control, I used the H/ACA miRNA ACA45 and an extremely low expressed miRNA-451a (Not found in Ion Torrent, and few in the GEO data sets).

The miRNA titration model predicts for low expressed miRNAs that increasing  $1/\lambda$  shift minimally the analytical curve, while the line displacement is greater for mid and high functional miRNAs respectively (Figure 43A). Experimentally miR451-3p, ACA45, and U78 functions show a small shift, while U44 completely overlap the controls (Figure 43 B and C).

For U3 derived miRNAs, located at 5' and 3' domains Figure 44 A, the analytical functions depict a slight displacement for the U3 (45-64) and U3 (195-214) (Figure 44 B and D), while U3 (75-94) keeps its lines overlying the control.

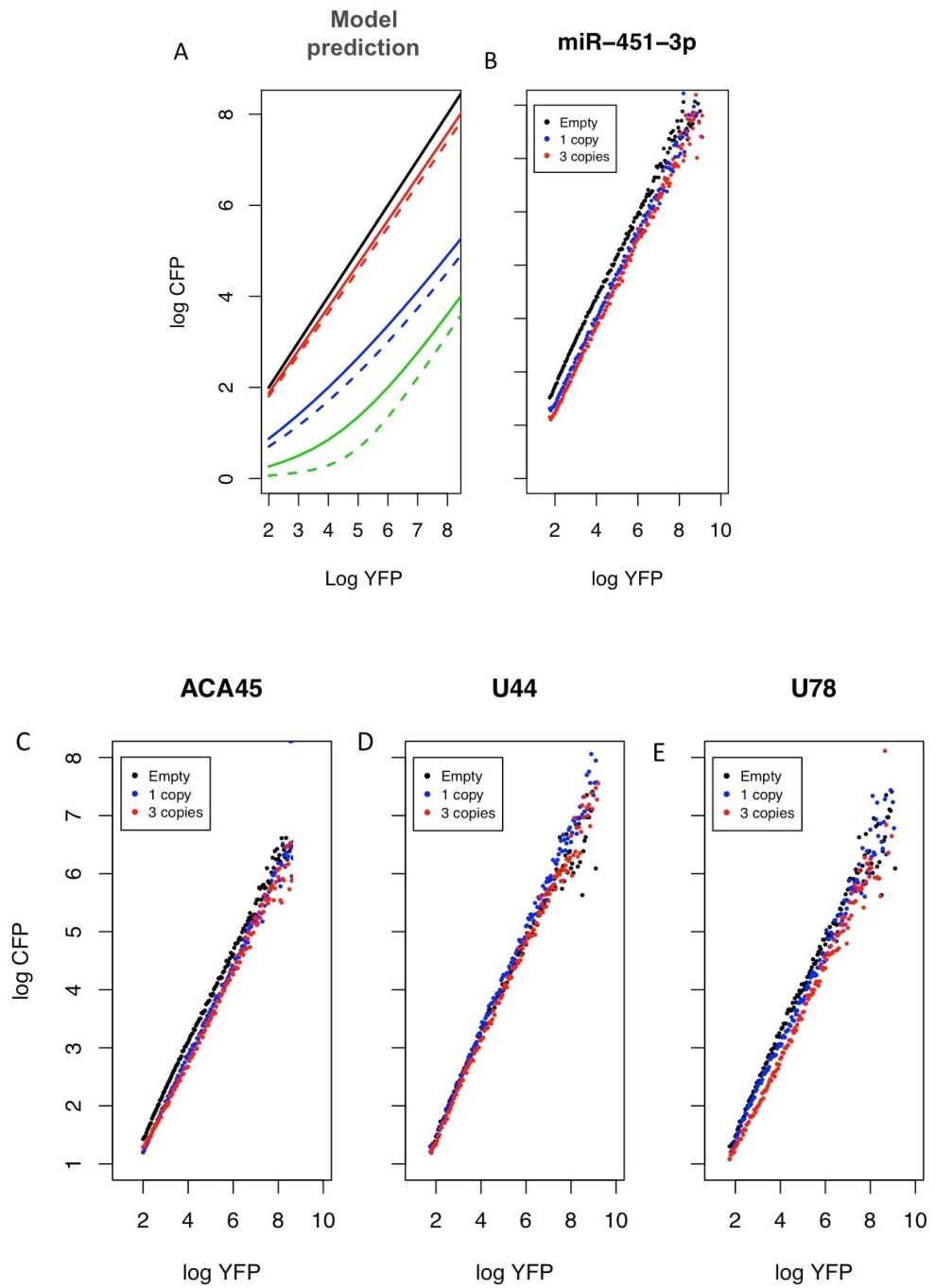
From the miRNA titration model and the analytical flow cytometry, I defined two steps to modulate endogenous miRNA activity *in vivo* (Figure 41 and Figure 42): improving binding by increasing target copies in the sensor (CFP 3' UTR), and out-competing miRNAs by an external siRNA overload.

#### 2.1.29 ANALYTICAL FLOW CYTOMETRY OF C/D SNORNAS: OUTCOMPETING MIRNAS.

From analytical flow cytometry, I established two steps, which use the titration model, to examine miRNAs activity *in vivo* by increasing binding (Figure 44) and overloading the cells with siRNAs.

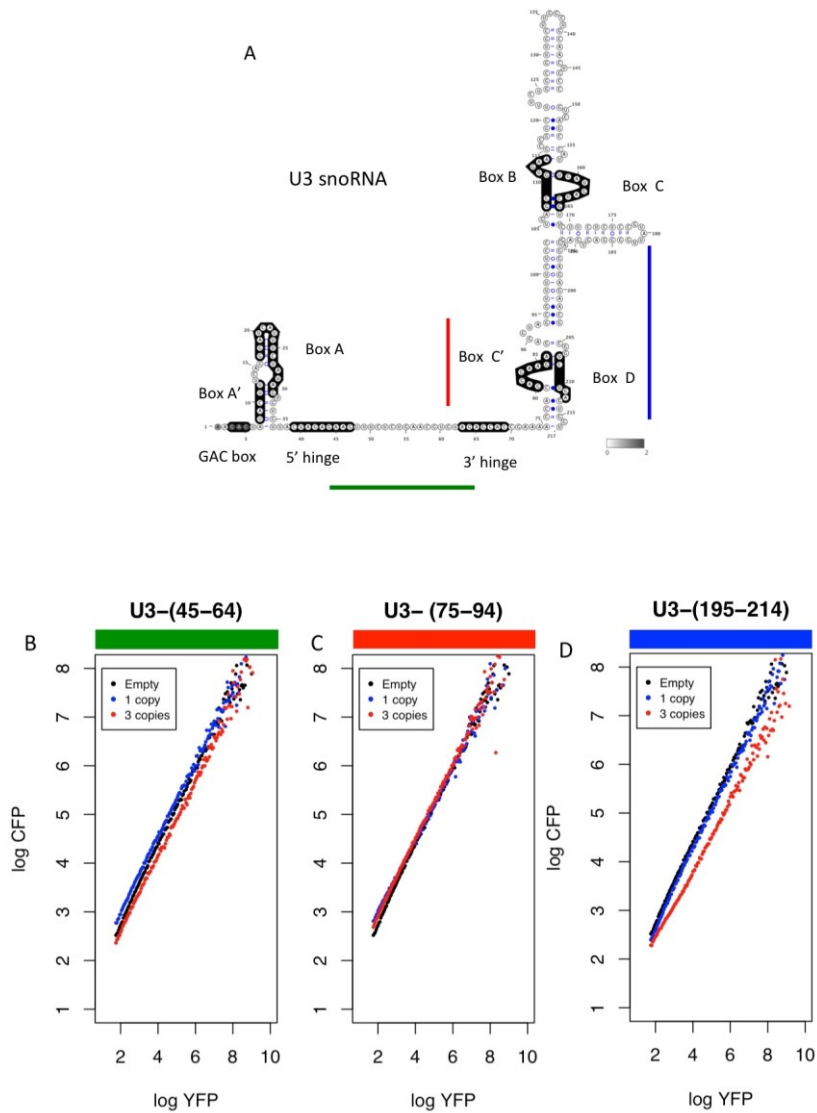
The out-competing experiments for miRNAs show a shift again up of miRNAs curves, in these experiments, I used miR-451-3p as lowly expressed miRNA, miR 221-3p as mid functional and miR-92a high functional miRNA (Figure 45 A - D and Section 2.1.26.5). Overloading the cells with siRNAs upshift the analytical curves for U3 (45-64) and ACA45 (Figure 45 E and H), while it does not affect U3 (195-214) and U78 (Figure 45 F and G).

Consequently, the analytical cytometry outlook determines that all the selected C/D-miRNAs contradict the miRNA model, except for the U3 5' domain located miRNA and H/ACA miRNA ACA45, suggesting that the intronic snoRNA yield no canonical miRNAs.



**Figure 43. snoRNA-derived miRNAs analytical resemble low functional miRNAs.**

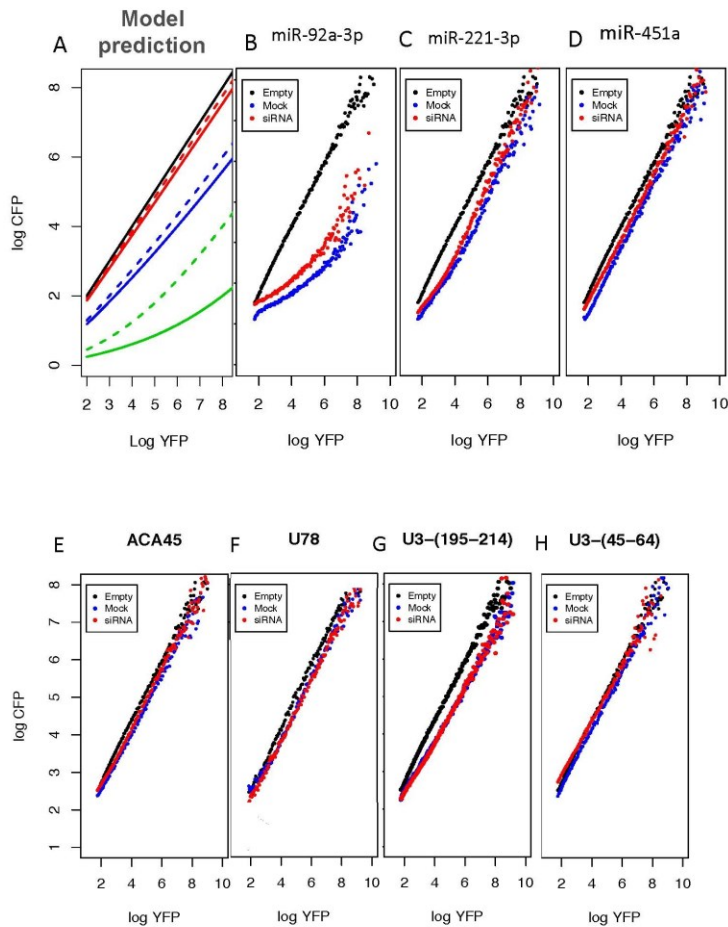
Dual fluorescent constructs containing full complementary sequences of snoRNA-derived miRNA were transfected in Hek293 cells and evaluated after 72 h. (A) Depict model prediction for three miRNAs with low, mid and high functionality to contrast curve shift upon  $1/\lambda$  increments (full vs. dashed line). Scatter plots show analytical functions for (B) Positive control miR-451a, H/ACA snoRNA-derived miRNA (C) ACA45, (D) U44, and (E) U78.



**Figure 44. U3 derived miRNA function resembles low functional miRNAs.**

Dual fluorescent constructs containing full complementary sequences of snoRNA-derived miRNA were transfected in Hek293 cells and evaluated after 72 h. (A) The cartoon depicts U3 secondary structure, in bold shows functional elements and it highlights the small-RNA found in NGS libraries. Full complementary regions inserted at the CFP 3'UTR were transfected in Hek293 cells and evaluated 72 h after. Scatter plots show analytical functions for (B) U3 (45-63), (C) U3 (74-94) and U3 (195-214), colors are paired between the cartoon and the scatter plots.





**Figure 45. Outcompeting miRNA-derived snoRNA shift the ACA45 and 5'-U3 domain analytical function.**

Constructs containing full complementary sequences of snoRNA-derived miRNA and siRNA were co-transfected in Hek293 and evaluated after 72 h. (A) Model predictions for three different functional miRNAs and analytical functions for (B)High, (C) mid and (D) low functional RNA, red dots depict the outcompeting siRNA resultant function. snoRNA-derived miRNA analytical curves for (E) H/ACA box ACA45, (F) C/D snoRNA U78 and (G) 3' and (H) 5' domains U3 derived miRNAs. Simulation can be reproduced at <https://medicalrnbiology.shinyapps.io/pUTApp/>.

### 2.1.30 CLASH CHIMERAS FOR U3 5' DOMAIN DERIVED MIRNAS.

As last attempt to disprove that U3 5' domain works as a miRNA, I used the Hyb pipeline in CLASH data sets and subset specifically for snoRNA-mRNA chimeras, because this technique recovers miRNA targets accurately, U3 5' domain shall show hybrid formation as canonical miRNAs.

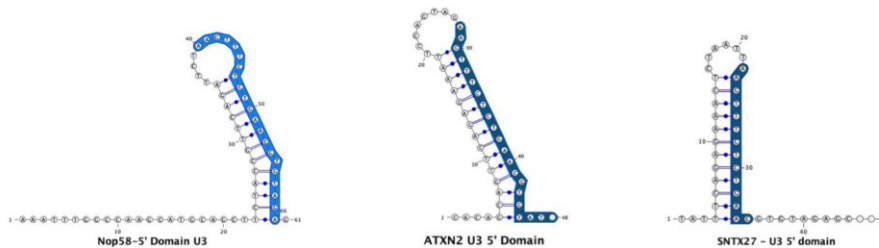
Calling chimeras for snoRNA in CLASH data<sup>30,91</sup>, recover three chimeric reads for U3 (45-64), and their base pairing resemble non-canonical miRNA motif (Figure 46).

Transcript	Number quimeras	Region
SNX27		3'3'UTR
LMTK2		2 CDS
NOP58		1 CDS
ATXN2		1 CDS

```

AAATTTGGCCAAGCATGCAGCTTCTACCGTTCAGATTCT-----
-----AAGTTTCTCTGAACGTGTAGAG
.....((((((((((((.....)))))))))).)))).
|
CACAGCAGTTCAGAGAAATTCAGT-----
-----AGAAGTTTCTCTGAACGTGTAT
....((((((((((((.....)))))))))).)))).)....
|
TATTTAGAGAAACTCTAATT-----
-----AAGTTTCTCTGAACGTGTAGAGC
...((((((((((((.....)))))))))).)))).).....
NOP58_mRNA
U3_snoRNA
|
ATXN2_mRNA
U3_snoRNA
(-17.8)
|
SNX27_mRNA
U3_snoRNA

```



**Figure 46. 5'-U3 domain derived miRNA's chimeras identify mRNA targets**

Ago1 raw data from crosslinking and analysis of hybrids (CLASH), was analyzed with Hyb pipeline subsetting for snoRNA. (A) Table summarized called chimeras from U3 5' domain, and dot-brackets notation from identified chimeras. (B) Cartoons depict secondary structure of chimeric read.

#### INSIGHT INTO SNORNA-DERIVED MIRNAS BIOGENESIS

So far I could deny, to a substantial extent, that C/D-miRNAs are argonaute miRNAs, since the snoRNP and DGCR8 CLIP-seq results on U78, U74, and U44, depict a recovery pretty similar to the small RNA library clusters. I imagined that this coherent recovery of sRNA could represent RNA protection of snoRNPs and DGCR8 during degradation.

To sort this out, I believed that if they are degradation products, to reduce nuclear RNA degradation should decrease C/D-miRNA recovery. The nuclear exosome requires the activity of Helicases, for the human exosome the adaptor protein hMTR4 fulfill this function.

#### 2.1.31 hMTR4 KNOCK-DOWN INCREASES C/D SNORNA EXPRESSION IN HEK293 CELLS.

I decided to analyze the effect of hMTR4 on C/D-miRNA because of its helicase activity and especially due to its role as a nuclear exosome adaptor. I used siRNA knockdowns and analyzed the small RNA content by NGS.

Before sequencing, I tested the knockdown by western blot and quantified levels of known hMTR4 substrates PROMPTS (Promoter Upstream Transcripts)<sup>55,120,121</sup>, knockdowns show hMTR4 protein reduction and increase of PROMPTS (**Figure 47 A – B**), then small RNA was purified and sequenced using Ion-Torrent platform.

The small RNA purification and the library preparation showed similar distribution and quality as the control libraries (prepared before 2.1.24), The libraries were aligned to transcriptome database and normalized using Deseq package; hierarchical clustering divided the libraries between mock and siRNA treated (**Figure 47 C**).

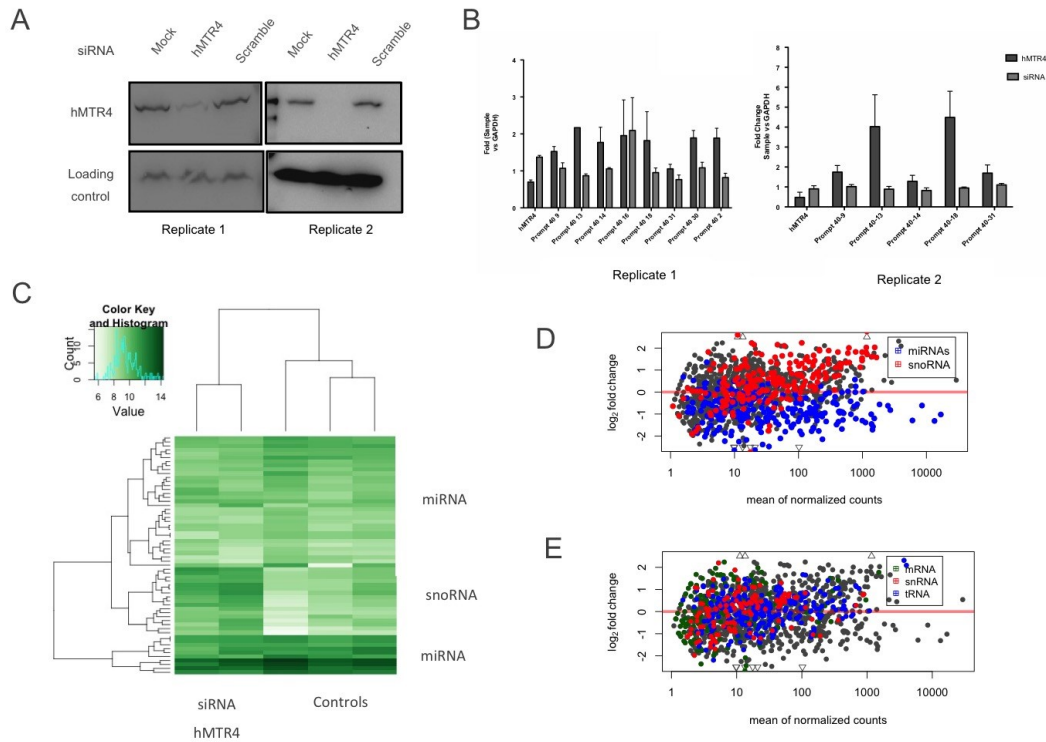
The heatmap and fold-change scatter plots show that hMTR4 reduction increases the expression of mature snoRNAs, while the C/D miRNAs are unchanged. To see whether this observation is similar for the two replicates, I analyzed the mature snoRNA population and used miRNAs as controls.

The independent replicates analysis shows that hMTR4 knockdown significantly increase mature snoRNAs, while the miRNAs are kept without changes, interestingly I noted no major differences for C/D miRNAs between knockdowns and the controls (**Figure 48**).

#### 2.1.32 EXOGENOUS GENOMIC GAS5 INCREASES U78, U44 AND U74 DERIVED MIRNAS.

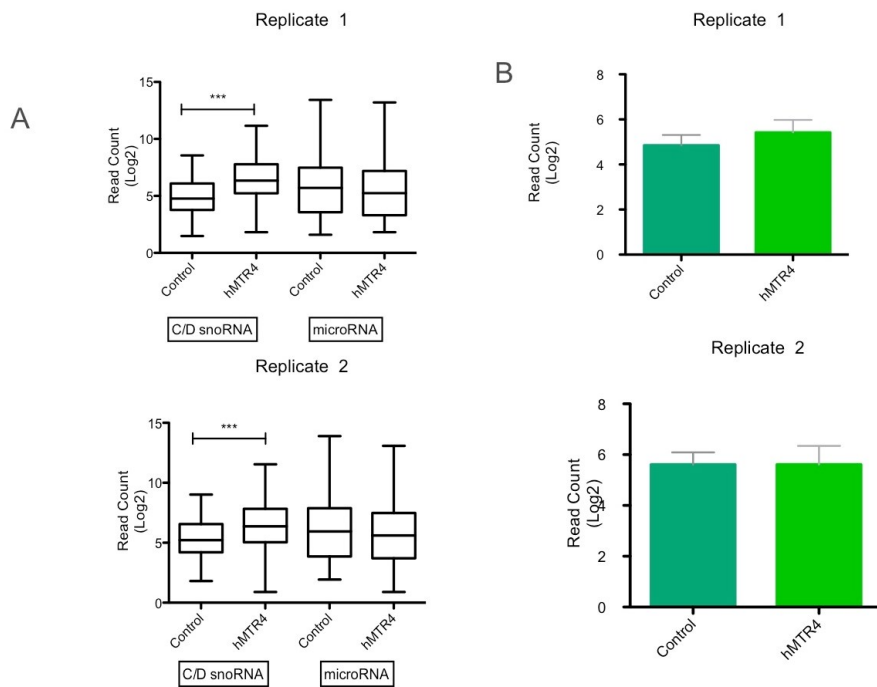
If these C/D miRNAs perform a particular cellular function in the cells, they might have a distinct production pathway, but due to low expression, a stable source is needed to evaluate biogenesis and possible functions.

In 2015 C/D miRNAs HeX proposed that GAS5 encoded the snoRNAs (U74, U75, U76, U77, U44, U78, U79, U80), which can be incorporated into Piwi proteins and regulate transcription<sup>77,78</sup>; however they showed results from exogenous expression of the "piRNA" from snoRNA expressed outside their intronic context.



**Figure 47. hMTR4 knock-down increases mature C/D snoRNA expression in Hek293 cells.**

hMTR4 small-RNA analysis. Hek293 cells were transfected with hMTR4 siRNA, smallRNA libraries were prepared and sequenced (Ion-Torrent). Three independent control (mock transfection) were used for differential expression analysis. (A) hMTR4 knockdown characterization by western blot (B). hMTR4 and PROMPT RNA relative quantification by real time RT-PCR. Notice the differences on the Y-axis from the replicates. (B) Heatmap 50 more expressed genes from the five analyzed libraries. (C) miRNA, snoRNA (D) mRNA, snRNA and tRNAs fold change between hMTR4 knockdown and control.

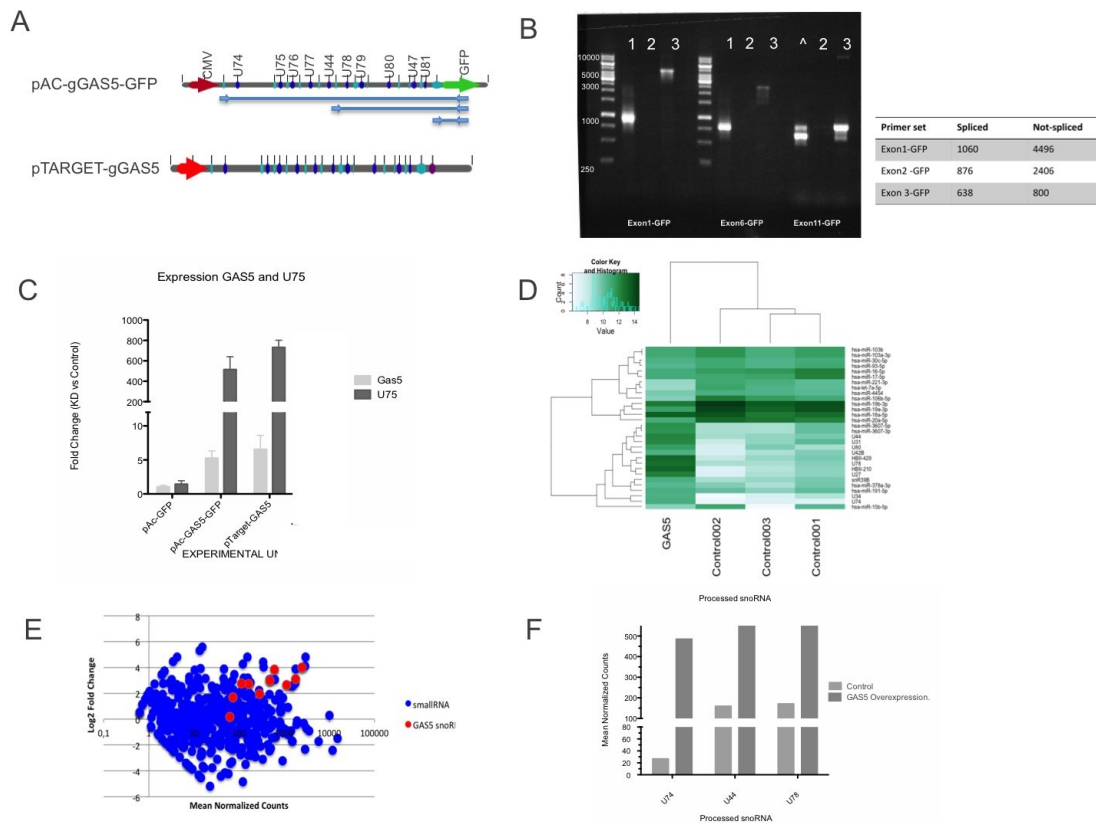


**Figure 48. snoRNA-derived miRNAs are unaffected by hMTR4 knock-down C/D snoRNA expression in Hek293 cells.**

Hek293 cells were transfected once and twice with hMTR4 siRNA, and smallRNA libraries were prepared and sequenced (Ion-Torrent), three independent control (mock transfection) were used for differential expression analysis. (A) Boxplots depict Log<sub>2</sub> normalized counts for mature C/D snoRNAs and miRNAs and (B) bar plots compare expression of snoRNA-derived miRNAs between controls and siRNA transfections, in the two experimental replicates.

I thought that an intronic snoRNA overexpression should be the first step to define C/D miRNA genesis and possible functionality. For that reason I cloned Gas5 genomic region, including of the snoRNA and exons into two CMV promoter vectors, with and without GFP (Figure 49).

I tested the splicing and expression of GAS5 final mRNA and U75; I found that the cells splice the exogenous GAS5, confirmed the lncGAS5 lncRNA identity and the distinct splice variants by Sanger sequencing Table 14. Using qPCR, I observed that genomic loci transfection increases the lncRNA-GAS5 and U75. Establishing that exogenous GAS5 loci (exon-snoRNA-exon) expression follow the same path as the endogenous snoRNAs, it is spliceosome substrate, generating the spliced mRNA (lncGAS5) and individual snoRNAs. Therefore, I extracted small-RNA, sequenced Gas5 and compared with untransfected controls, checking for quality and evaluating the presence of C/D miRNAs Figure 49 A-C.



**Figure 49. Genomic GAS5 (gGAS) exogenous expression increases C/D derived miRNAs in Hek293 cells**

Hek293 cells were transfected with GAS5 genomic sequence and analyzed after 72 h.(A)The cartoon shows the GAS5 constructs used, arrows represent the forward and reverse primers used for PCR. (B) PCR gel electrophoresis cDNA from transfected cells: 1) (1) gGAS5, (2) Empty backbone and (3) GAS5 plasmid table depicts expected PCR products (representative image from three independent experiments). (C) real time RT-PCR for Gas5-lncRNA and U75 (bar plot from three replicates). (D) gGAS5 exogenous transcript small RNA analysis. Heat map for the 30 more expressed genes (E) Scatter plot show relative small-RNAs expression vs. fold change Gas5 vs. Control red dots represent (Gas5 encoded snoRNAs). (F) Bar plots show the number of Normalized reads for the GAS5 snoRNAs derived miRNAs.

Ion-Torrent profiling comparison (GAS5 vs. Mock) shows that the Gas5 encoded snoRNA pull the snoRNAs up and create a distinct cluster on unsupervised analysis. GAS5 encoded snoRNAs increase over two and four-fold. Also, GAS5 encoded C/D miRNAs increase after transfection of exogenous GAS5 expression, but the fold change is not the similar to mature snoRNAs.

Briefly, the hMTR4 knockdown enhances the recovery of mature snoRNAs as well as the exogenous genomic Gas5 expression (for its encoded snoRNA), while the hMTR4 knockdown induces no changes in the recovery of C/D miRNAs, the exogenous Gas5 expression increases the number of C/D miRNAs on its encoded snoRNAs. However, any conclusion mature snoRNAs from the Ion-Torrent profiles performed here should be taken carefully since the library preparation is not optimal for mature snoRNAs (Figure 13 and Figure 14).

---

**Table 14 Gas5 spliced variants**

---

Exon 7 incomplete (7a), Exon 8 and 10

Exons 5, 8, 9 and 10

Exon 5, 6, 7, 8, 9, 10 and 11, complete Intron 3, containing U76

Exon 7 incomplete (7a) Exon, 5, 8, 9 and 10

Exon 5, 8, 9 and 10

cDNA from transfected Hek293 cells was clone into p-GEM and sequenced by sanger. Work performed together with Jasmin Preis during her Lab rotation of the master Molecular Medicine).

---

### 3 DISCUSSION

#### C/D SNORNA-DERIVED MIRNAS-LIKE MOLECULES

snoRNA-derived miRNAs, with less than 30 nt, were initially reported by Ender *et al.*, who described that the H/ACA 45 snoRNA loci produces miRNAs. Their evidence indicated that ACA45 is bound to Ago, it is independent of the microprocessor, and requires Dicer activity; has as its putative target CDC2L6<sup>4</sup>.

Scott and Brameier independently increased the evidence for snoRNA-derived miRNAs, determining that C/D snoRNA loci produce miRNA-sized RNA and showing that their distribution of reads resembles that of miRNAs, with a guide and passenger strand. This processing is different but conserved among various cell types. Furthermore, many annotated miRNAs carry snoRNA-like sequences, while other experimental sources suggest that snoRNAs hold other functions including alternative splicing, mRNA post-transcriptional regulation and piwi-RNA like activity 71,72,74.

However, a clear consensus for these “novel” snoRNA features is still in debate, and many reviews recommend deeply exploring snoRNA-to-miRNA processing and its presumed functionality.<sup>6,7,9-11</sup>

Here, I concentrated on C/D snoRNAs attempting to disprove systematically that C/D-miRNAs have canonical miRNA activity. I characterized small RNAs from Hek293 cells using cellular RNA and RNP-associated RNA (Figure 7). To evaluate miRNAs endogenous functionality, I developed and implemented an analytical flow cytometry assay for miRISC activity *in vivo*. Using a molecular titration model and the R environment, I validated this assay for canonical miRNAs and applied to C/D miRNAs (**Figure 25**).

Based on the results, it appears intronic C/D-miRNAs are not canonical miRNAs rather degradation products. It is important to examine how the 5' domain of U3 (C/D snoRNA) produces canonical miRNAs and propose (hypothetically) that U3-derived miRNAs are in fact regulators of rRNA biogenesis.

I will also explain how analytical flow cytometry can be utilized as a high-resolution sensor of endogenous miRNA activity, including further applications and comparison with similar tools.



### 3.1.1 C/D SNORNA-DERIVED MIRNAS DIFFER FROM CANONICAL MIRNAS.

I applied a deductive approach to characterize C/D-miRNA; disproving rather than confirming a hypothesis, this empirical falsification works by establishing a set of testable predictions that would reject the hypothesis. My hypothesis claims that C/D-miRNAs functionally operate as canonical miRNAs, which reduces the scope to classic Ago-dependent regulation and discards other miRNA mechanisms. Using the miRNA paradigm as theoretical predictor I set up the following "diagnostic criteria" for miRNAs":

1. miRNAs are shorter than 25 nucleotides <sup>8,112</sup>.
2. The miRNAs guide and passenger strands can be detected in small RNA libraries <sup>114</sup>.
3. miRNA-function depends on Argonaute, which thus should interact with canonical miRNAs <sup>101</sup>.
4. Dicer processes the vast majority of pre-miRNAs and their hairpin structure can be defined <sup>100,114</sup>.
5. DGCR8 processes pri-miRNA, so miRNAs should be recovered from DGCR8-bound RNAs <sup>59,99</sup>.
6. They should act according to the molecular titration model in analytical flow cytometry <sup>119,122,123</sup>.
7. miRNA targets can be identified from Chimeric reads on immunoprecipitated NGS data.

Table 4. Summary of functional criteria for selected C/D snoRNA-derived miRNAs.

Candidate	miRNAs	ACA45	U3	U78	U44	Section in the dissertation
Length <25	+	+	+	-	-	2.1.24.1
Cytoplasmic	+	+	+	-	-	2.1.24.3
Guide strand cluster	+	+	+	+	+	2.1.24.1
Argonaute	+	+	+	-	-	
Dicer	+	-	+	-	-	2.1.24.4
DGCR8	+	+	+	+	+	
Threshold sharpening	+	+	+	+	-	2.1.28
Shift after outcompeting	+	+	+	-	-	2.1.29
Described Targets	+	+	+	-	-	2.1.30

Table 15 miRNA features evaluated in this study.

Length and strand clusters small RNA Seq.

Argonaute, Dicer, DGCR8 CLIP Seq.

Analytical flow cytometry (threshold and outcompetition)

Targets from CLASH chimeras call.

\*Described in <sup>4</sup>

First, I evaluated the small RNA content of Hek293 cells using NGS (Ion-Torrent). All libraries correlated positively with a high Pearson coefficient, hold comparable read distribution and a

large number of miRNAs were recovered (Figure 8). The correlation and sno/miRNA distribution match the data sets derived from GEO data sets.

The libraries retrieved different snoRNA classes and subclasses (mature and miRNAs), without an exact correlation between the mature snoRNA (~70 nt) and small fragments (<40 nt) (Figure 9), suggesting that C/D miRNAs might have a different source, assuming that the abundance of degradation products should be proportional to that of the mature snoRNAs.

To discern the differences between C/D miRNA processing and degradation products, I used a simple unsupervised analysis (hierarchical clustering), differentiating C/D miRNA coherent processing and degradation<sup>112</sup>. The hierarchical clustering defines a distinct beginning and end for C/D mature and derived-miRNA (Figure 10), facilitating the selection of small RNA with coherent processing.

Using hierarchical clustering and expression analyses, I selected six snoRNA candidates to evaluate as a source of miRNAs: U78, U74, U44, snR39, and U3. I chose ACA45 as a positive control for snoRNA-derived miRNAs because it appeared in the clustering analysis, and it has been described as an miRNA<sup>4</sup> (Figure 11).

Candidates' clusters (hierarchical clustering of reads) show the guide and passenger strand distribution observed for miRNAs<sup>112</sup>, but the read length differs from canonical miRNAs. The mean read length for U78, U74, U44, and snR39B is longer than 25, while the controls and U3-derived miRNAs are shorter than 25 (Figure 11). This suggests that the intronic C/D miRNAs are bigger than the canonical miRNAs and probably can be interpreted as piwi-RNAs<sup>124</sup>.

The differences in recovery of mature snoRNAs versus further processed miRNAs (< 40 nt) indicate that the latter might have a distinct production mechanism, because the highly expressed long snoRNAs have no processed counterpart.

Since there was a possibility that library preparation introduces a bias that could explain the forehead mature-miRNA/snoRNA distribution. I checked various library preparations and observed that the ethanol-based purification (used for the Ion-Torrent) poorly recovers mature snoRNAs, and PAGE-based separation (20-200) improves canonical snoRNA retrieval (Figure 14 B). The C/D miRNA number is just a small fraction of the whole snoRNA population, and their expression is proportional to mature snoRNAs (Figure 14 D), suggesting that C/D miRNAs levels are low in basal culture conditions.

Furthermore, C/D miRNAs are differentially located within cell. Canonical miRNAs are mainly cytoplasmic and the majority of the C/D snoRNAs are nucleolar and nucleoplasmic, except for ACA45 and U3, which are also found in the cytoplasmic fraction (Figure 15). A biochemical characterization observed a similar distribution of mature snoRNAs inside the nucleus, and showed no methylation modifications over mRNA. However, it is unclear whether this effect could be distinct between the short and long fragments<sup>76</sup>.

From small RNA libraries, I showed that C/D miRNAs are weakly expressed and located mainly in the nuclear compartment. Their clustering resembles the miRNA guide/passenger strand recovery, and their mean read length is approx. 25 nt, indicating a distinct role and probably putting them closer in nature to piRNAs than to canonical miRNAs (Discussed in 3.1.3). Briefly, this revealed that C/D snoRNA candidates do not follow the behavior of miRNA controls (miRNAs and ACA45-miRNA), except for U3, which is not intronic but is bigger (217 vs. 70) than the other C/D snoRNAs.

Argonaute CLIP and RIP seq analyzed data did not contain the intronic C/D snoRNA candidates (U78, U44, U74, snRN39B), but recovered ACA45 and U3. Also, Argonaute-bound miRNAs are distributed similarly to the miRNAs contained in the small RNA profiles (Figure 17 and Figure 18). I observed a lack of the intronic candidates in Dicer PAR-CLIP data (Figure 20).

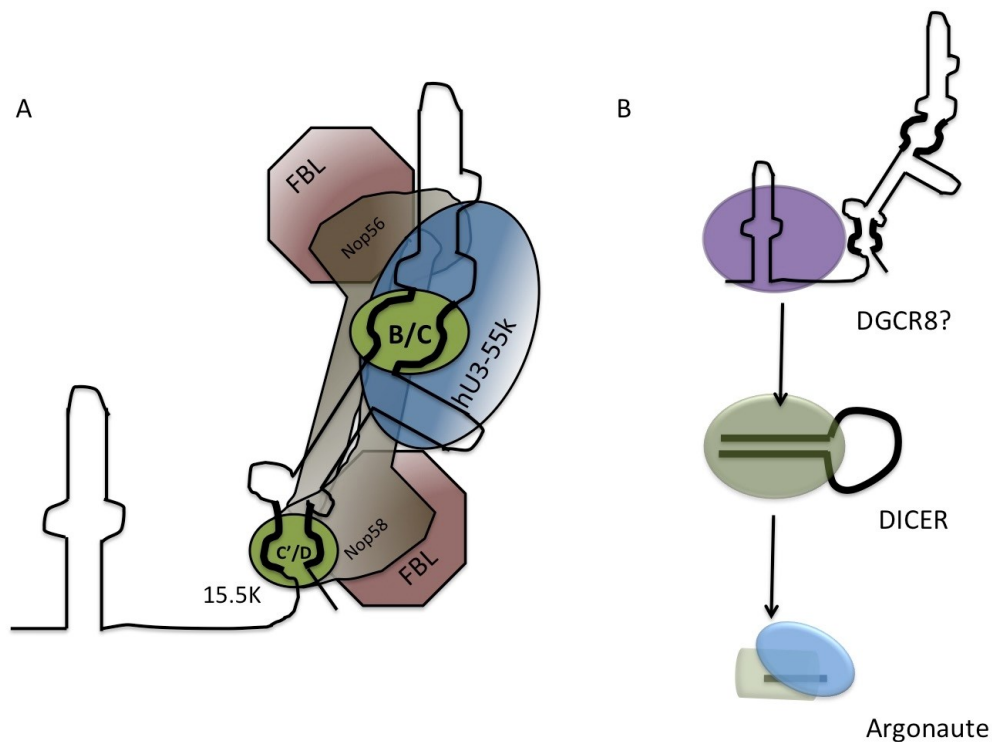
For DGCR8, I recovered C/D snoRNAs and miRNAs (Figure 21). Interestingly U78 recovered reads resemble the processing cluster of small RNA profiles (Figure 22), suggesting that the C/D miRNAs could be RNP interactions, particularly in terms of DGCR8 protection of U78 until it is degraded. (See below 3.1.3)

From analytical flow cytometry, U78 shifts the analytical function, without being outcompeted by siRNA, implying that this performance is miRNA independent (Figure 43 and Figure 45). I cannot discount the possibility that the U78 inserted site can be modified by the mature U78 and affect CFP translation, analogously to snoRNA base vectors that downregulate complementary mRNA by introducing its sequence between the C and D boxes<sup>80</sup>, or to endogenous snoRNAs that modify the steady state of mRNAs (as described from LIGR-seq)<sup>79</sup>.

If U78 is responsible for miRNA-independent CFP regulation, a CRISPR knockout of U78 should not lead to molecular repression in analytical cytometry measurements, and different experiments could elucidate putative snoRNA-mRNAs regulatory mechanism.

### 3.1.2 5' U3 DOMAIN SOURCE OF MIRNAS

Based on their length (above 25 nt), poor interaction with Argonaute and Dicer, and their endogenous functionality, there is no evidence that C/D miRNAs perform miRNA-like regulation.



**Figure 50. U3 structure and 5' domain miRNP interaction**

(A) Schematic U3 snoRNP structure, RNA is shown in black and protein organization is indicated<sup>63</sup>. (B) Summary of the model described here, notice that only the 5' U3 domain was recovered in miRNPs.

However, for the 5' domain of C/D snoRNA, which contains a GAC box, the Box A and A'; and a 5' and 3' hinge, I recovered miRNAs of 22 nt read length and found them in the cytoplasm (Figure 15 and Supplementary Figure 6). For miRNP, I found that the U3 5' domain miRNA (5' 46-64 3') bound to several Ago IP preparations (Figure 18, Figure 23 and Figure 24), and it interacted with Dicer and DGCR8.

Dicer PAR-CLIP data recover almost the complete 5' region, and Dicer read distribution shows a hairpin structure fitting with the folded 5' domain (Figure 24). Moreover, analytical flow

cytometry registered that the U3 miRNA (5' 46-64 3') behaves similarly to low expressed miRNAs like ACA45 or miR-451-3p. (Figure 44 and Figure 45)

Furthermore, the chimeric call from CLASH data recovered miRNA targets for 5' U3 domain derived-miRNAs, including Nop58 and ATXN2, hinting that this region yields functional miRNAs (Figure 46).

I posit that that U3 could be ncRNA with dual activity and since it is the center for the early steps of rRNA processing, I presume that if it is also a source of miRNAs, it might regulate proteins related to ribosome biogenesis.

This hypothesis requires experimental inquiry, so I will describe what I consider to be the pivot point between snoRNA and miRNAs, including empirical outlines that could disprove the hypothetical mechanism.

U3 snoRNA associates with the canonical snoRNP Nop56, Nop58, and Fibrillarin (Figure 21 and Figure 23), and its unique interacting factor U3-55k, which binds to the distinct U3 motif B/C 44,61. Importantly, the distribution of the mammalian snoRNP is analogous to the snoRNPs location in the yeast U3 (compare Figure 23 with CRAC data)<sup>111</sup>.

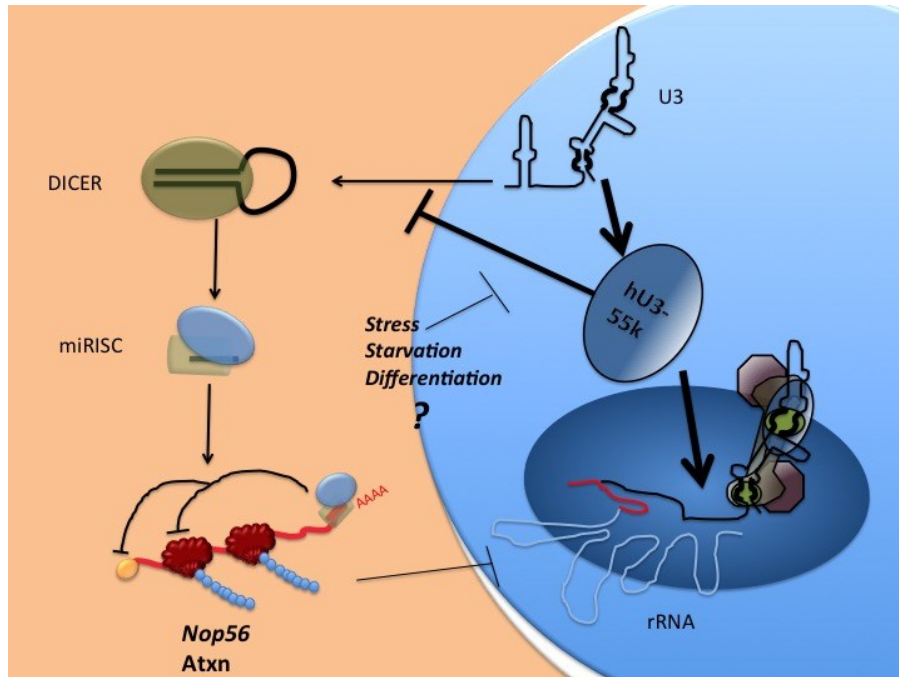
Studies in U3 biogenesis suggested that U3 does not have a cytoplasmic phase and goes through the Cajal body, even though U3 interacts with the nuclear export factors PHAX and CRM1<sup>60,125</sup>. Additionally, the interaction with U3-55K is needed for nucleolar location<sup>63,115,126</sup>, including several trans acting factors interacting during biogenesis<sup>125</sup>.

However, certain circumstances change the U3 cellular location. For example, in the absence of proper C, C' D or B boxes, the U3 is not retained in the nucleolus and could end in the cytoplasm<sup>61,69</sup>. Foremost in starvation, U3 move to the cytoplasm<sup>68</sup>. Probably U3 location might be a ribosome biogenesis pivot control. I would consider it possible that during starvation, the U3 moves into the cytoplasm and becomes a Dicer substrate, creating functional miRNAs that down regulate Nop58.

Recently, Chen *et al.* reported that hypotonic stress reduces ribosome biogenesis by U3 55K hyperacetylation, leading into a decreased interaction with U3<sup>64</sup>. Also, epithelial differentiation reduces ribosome biogenesis, and this seems to depend on U3 and U3-55K expression because during differentiation, U3 snoRNA and U3-55k levels are significantly smaller than undifferentiated controls<sup>63</sup>.

On the miRNA front, starvation increases the expression of group V miRNAs (Drosha independent and PHAX-CRM1 dependent), overcoming group I miRNAs (Drosha-Exp5) by reducing Exp5 expression. This suggest that PHAX-EXP1-dependent miRNAs are specific for stress conditions and governed at the nucleocytoplasmic transport<sup>29</sup>.

Based on this information, I speculate that this model is possible:



**Figure 51. Hypothetical model for U3 derived miRNAs**

hU3-55K controls U3 levels and regulates ribosome biogenesis<sup>63,64</sup>, while U3 steady state levels are controlled post-transcriptionally<sup>127</sup>. In cases of differentiation or stress, hU3-55k probably interacts less with U3, allowing it to enter the miRNA pathway, where it could manage stress responses by regulating Atxn or stopping ribosome biogenesis by reducing Nop56.

Stress or differentiation induce post-transductional changes a to the U3-55k protein and reduce its interaction with U3. Simultaneously, the miRNA machinery recognizes the U3-55k-free snRNA and produces miRNAs transported into the cytoplasm by the PHAX-Exp1 pathway. In the cytosol, Dicer processes the hairpin and yields double-stranded RNA included in Ago. Since the 5' contains a methyl-cap, just the 3' strand is selected and regulates stress proteins like ATXN2 or snoRNPs

The evidence in this dissertation and the literature together provide feasibility to the model, giving predictive power to test experimentally and disprove or reconsider the hypothesis.

First, U3 5' domain exogenous expression should produce functional miRNAs, either using shRNA (pol III promoter) or pol II-driven vectors. Second, decreasing U3-55K RNA protein interaction

should increase U3-derived miRNAs. Experimentally, one should use U3 exogenous expression with modified B/C and C'/D motifs or knock down U3-55k protein. Finally, stress conditions or differentiation should reduce U3-55K expression increasing U3 miRNA generation.

In conclusion, U3 5' domain is a *bona fide* miRNAs locus, and U3-55K protein probably controls U3 fate. If the cell needs ribosomes U3-55K bound to U3 allows the 80 SSU (small subunit) processome formation, while in case of differentiation or starvation, U3-55K interaction is reduced permitting U3 to enter the miRNA pathway through the PHAX and CRM1 pathway.

### 3.1.3 C/D MIRNAS DEGRADATION VS. FUNCTIONALITY.

miRNA "diagnosis criteria" do not apply to intronic C/D miRNAs (*Table 15*), insinuating the improbability that C/D miRNAs operate as canonical miRNAs. What can we say about their role? And how good is the evidence providing biological explanations?

Here are two perspectives on C/D miRNAs: one as alternative to miRNAs (piRNAs), and the second that posits they are degradation remnants.

In 2015, He *et al.* found small RNAs derived from the Gas5 locus and proposed them to be piRNAs<sup>77</sup>, similarly to my description for small RNA profiles (Figure 11), exposing differential expression between tumor and normal tissues, and presuming that these "sno-piRNAs" regulate transcription through epigenetic mechanisms.

My primary concern refers to their exogenous expression experiments, which were performed with constructs containing only the snoRNA gene without flanking exons. Also, I replicated the loop PCR they used and found that their recommended primers cannot accurately differentiate between the "piRNA" and the mature snoRNAs (data not shown). They continued describing snoRNA-derived piRNAs with a similar set of techniques<sup>78</sup>.

The differential expression of tumor-normal tissues<sup>77</sup> (or in activation of lymphocytes<sup>78</sup>) could be explained alternatively by different ribosome needs as for undifferentiated vs. differentiated<sup>63,77</sup>. Their conclusion could be exemplified as a problem of induction, where true premises (arguments, in this case, sound experiments) lead to false conclusions. To clarify: their experiments also prove other scenarios and are not restricted to "piRNAs".

piRNAs are an excellent functional alternative since they are bigger than miRNAs (Figure 11 C). However mechanistic experiments are needed to overcome the biological and technical weaknesses from<sup>77</sup>. At the moment, I know that the exogenous expression of the genomic Gas5

increases the number of C/D miRNAs and it will probably be a good starting point to develop **(Figure 49)**.

C/D miRNAs can also be discussed as degradation remnants. Assuming that sRNA-seq reads reflect RNP protection, which holds true for miRNAs in Ago and Dicer, I recognize that DGCR8 snoRNA distribution for U78 is similar to the reads recovered in small RNA libraries, including the patterns from snoRNPs PAR-CLIP data sets **(Figure 22)**.

To test whether C/D miRNAs are degradation products, I evaluated their production upon reducing the exosome adaptor hMTR4, and observed an increase in mature snoRNAs without affecting the C/D miRNAs production, challenging the ideas of degradation products **(Figure 47 and Figure 48)**. However, HITS-CLIP and mass spectrometry for DGCR8 showed that the microprocessor is an adaptor of the nuclear exosome too, explaining how hMTR4 could not affect the C/D miRNAs.

These two speculated scenarios not mutually exclusive, and require experimental evaluation using risky predictions and deductive falsification (as performed here for C/D miRNAs). This might give evidence to disprove the idea that they are piRNAs or degradation products.

#### *3.1.4 ANALYTICAL FLOW CYTOMETRY HIGH-RESOLUTION PLATFORM FOR MIRNAS FUNCTIONALITY.*

The underlying assumption in deep sequencing studies is that miRNA quantification reflects miRISC activity because small RNA libraries display Ago2 RNA stabilization,<sup>128,129</sup> and Ago2 stability depends on miRNA abundance<sup>130,131</sup>.

The dynamic function of miRNAs introduces uncertainty into small RNA profiling. For example, Ago 1–4 proteins are less abundant than miRNA molecules, so the cell deals with Ago sub-stoichiometric levels by multiple recruitments to free miRNA-mRNA complexes<sup>132</sup>, suggesting that cellular profiles contain also non-protein bound mRNA-miRNA molecules.

At the functional level, lentiviral-based high-throughput experiments showed a breach between silencing and expression<sup>41,133</sup>, because more than 50% of miRNAs hold no detectable activity, meaning that they are working or that the sensor resolution is weak. Moreover, many highly abundant Ago2-bound small RNAs showed no inhibitory activity<sup>41,102,134</sup>.

Considering that various studies tested miRNA functionality using luciferase-based assays, and those experiments had low resolution and narrow detection limit<sup>41,102,134</sup>. I reasoned that a high-



resolution assay for miRNAs functionality was required. Therefore, I generated and validated a single plasmid fluorescent reporter to use for individual cell measurements.

To establish the raw data from individual cell measurements (around 50.000 cells per miRNA), we adapted, tested and validated a molecular titration model (Figure 28 and Figure 31). This model was tested only for miR-20a and involved a TET dependent bi-directional promoter, demanding cell selection and the concomitant expression of the tetracycline repressor<sup>44</sup>, showing the need for at least two plasmids and a lot of work.

The single plasmid simplifies the experimental workflow and expands the model to other cell types. The dual fluorescence system uses standard constitutive promoters (Tk and SV40, Figure 25, Figure 26 and

Table 1), enables quantification by flow cytometry (Figure 27) and reproduces the results obtained in luciferase assays for the same miRNA candidates (Figure 27 and Supplementary Figure 8).

Regardless of the differences among the fluorescent sensor used by Mukherji et al.<sup>44</sup>, the results of the Gl-2 siRNA experiments that modulated the threshold-linear response (through different binding sites) and shifted the function (increasing siRNA concentration) agreed with the simulated model predictions (Figure 33 and Figure 34). More convincingly, this applies exceptionally to endogenous miRNAs (Figure 38 and Figure 39).

Interestingly, our sensor disconnects miRNA silencing from their expression (Figure 38 and Figure 35) and we observed a significant difference between low and high functional groups (Figure 38 D), which might associate with the detection threshold for luciferase-assays (1000 reads (~9 log<sub>2</sub> scale))<sup>134</sup>. I cannot reject that stabilizing mechanism might also change the ,into P bodies<sup>135,136</sup>, or a different ratio of miRNAs to target mRNAs<sup>41</sup>.

We considered that the analytical flow cytometry could generate hypothesis for Ago2 functionality *in vivo*. As proof of principle, we tested how the miRNAs subcellular location influences their function. Using HeLa cells, we showed that some low functional miRNAs are nuclear, while highly active miRNAs are in the cytoplasm ( Figure 40).

Several studies indicate that many small RNAs associate with Ago2<sup>4,74,134,137,138</sup>, lacking proper functional readouts. For instance, they use either luciferase assay<sup>4,74,134</sup> or artificial exogenous expression<sup>77,78,139</sup>, suggesting that the real physiological role is still undefined and contradictory in the literature<sup>10,76-78,140-142</sup>.

Furthermore, threshold modulation and out-competing miRNAs allowed the analysis of low expressed miRNA functionality. For example, U3 5' domain and ACA45 replicate the behavior of miR-451a-3p, while U78 shows regulation independent of Ago (Figure 43, Figure 44 and Figure 45). I consider that molecular titration models in combination with FACS analysis could evaluate expanded processes, like protein or nucleic acid degradation, producing *in vivo* quantitative outputs that might increase the analytical toolbox for cell biologists.

Finally, our reporter system (in combination with its data analysis) quantitatively measures miRNA activity, defining parameters equivalent to dissociation constants and concentrations, thus empowering high-resolution quantitative assay for miRNA functionality, bringing a better and sounder picture of miRNA-induced regulation *in vivo*.



## 4 SUMMARY.

U3 is the exception for snoRNA's canonical functionality. It is transcribed independently, guides rRNA early cleavage, and requires a unique set of interacting proteins. Structurally, U3 has two functional domains, the 5' domain, which has the A box and two hinge regions and interacts with the rRNA 5'-ETS, and the 3' domain that interacts with the classic snoRNPs -Nop56, Nop58, 15.5K and Fibrillarin- and its unique protein U3-55K. This interaction requires the sequence motif B/C, C'/D and a secondary structure that allows a correct cellular location.

U3 controls ribosome biogenesis and its activity evolved as an intricate regulation. For example, cells with low de novo ribosome synthesis, such as final differentiated cells, reduce U3 synthesis by controlling U3-55K transcription. Additionally, stress-induced U3-55K acetylation reduces U3 stability, while starvation pulls U3 out of the nucleoli, its main location, into the cytoplasm.

Recently, NGS evidence expanded small RNA regulation beyond the accepted miRNAs and introduced RNAs with double function. For instance tRNA and H/ACA snoRNAs can produce small RNAs that effectively bound to Ago proteins (The main miRNA's effector). While C/D snoRNAs produce small fragments, its functionality still unclear and contradicting. To characterize functionally C/D snoRNAs derived small RNAs (s-sno-RNAs) we generated small RNA libraries from Hek293, and analyzed snoRNP and miRNPs CLIP data.

To rule out that C/D snoRNAs perform miRNA activity, I checked for miRNAs biochemical, structural and functional features, and discarded intronic C/D s-snoRNA as a source of miRNAs. The most expressed ones in our profiles (U78, U44 and U74) have a mean length between 26 and 27 nts, are not found in miRNPs and do not perform posttranscriptional regulation. Meanwhile canonical miRNAs and ACA45 miR-snoRNA, -H/ACA with miRNA function- have a size between 19-21 nts, interact with Ago and show posttranscriptional repression (see below).

Surprisingly, I could not discard U3 as a source of miRNAs. We found products between 21 and 23 nt in our libraries, and in several Ago RIP and CLIP data, suggesting that it can produce functional miRNAs. But its low expression made classic functional assay (luciferase reporter) inquires unreliable, because their noise-to-signal resolution is lower than next generation sequencing techniques.

The luciferase resolution made small expressed miRNAs undetectable, misleading conclusions about functionality. To increase the resolution I designed a reporter-assay based on two

fluorescent proteins (Un-translated trans assay, UTA), where one works as reference while the other contains a miRNA-complementary target region.

Using analytical flow cytometry, which includes a molecular titration model and R environment data analysis, I increased the resolution to single cell level and evaluated it in combination with small RNA profiles derived from NGS. The model predicted three distinct groups of miRNAs with unique repression activities (low, mid and high) governed by expression and miRISC formation rates.

I confirmed the model's feasibility, employing controlled concentrations of a non-human targeting siRNA and increasing miRNA complementarity on the sensor protein. Using 30 individual human miRNA candidates with various expression levels, we validated this technique as a high-resolution analytical tool for miRNAs repression activity.

I used the model prediction for low functional miRNAs and tested it for endogenous snoRNA-derived miRNA, as predicted: the control (miRNA-451a) showed small changes after increasing bind and outcompeting with siRNAs. The snoRNA-derived miRNAs, ACA45 and U3, performed similarly to the canonical miRNA.

U3 5' domain derived-miRNAs have a secondary structure that fits as a Dicer substrate, the predicted loop from thermodynamic folding is poorly represented in Dicer CLIP pools. Furthermore, low base pairing energy at the distal U3 5' domain predicts its 3' end as guide strands and Ago CLIP data confirms this prediction.

Since U3 5' derived miRNAs are low expressed, I confirmed its functionality with my high-resolution tool and showed, as the mathematical model predicts, that U3 in basal conditions works a low functional miRNA. Additionally calling chimeric reads from CLASH data, I found four mRNA targets including Nop58 and ATXN.

Other experiments ought to be employed to confirm targets regulation, including U3 5' domain expression without the 3' domain (snoRNPs interaction domains) and using mutated C,C' and D boxes (U3-55K binding). Also, stress conditions that regulate ribosome biogenesis and U3 expression, need to be tested to rule out that U3 is a moonlight non-coding RNA that regulates cellular stress response by generating miRNAs.

In conclusion, I can assure that intronic snoRNAs are rarely a source of miRNAs, that analytical flow cytometry increases the resolution and evaluates post-transcriptional regulation of low expressed endogenous miRNAs. Additionally, I have evidence that predicts U3 5' domain as a source of miRNAs, which needs to be confirmed and mechanistically explored.

## 5 BIBLIOGRAPHY.

- 1 Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77-94, doi:10.1016/j.cell.2014.03.008 (2014).
- 2 Martens-Uzunova, E. S. *et al.* C/D-box snoRNA-derived RNA production is associated with malignant transformation and metastatic progression in prostate cancer. *Oncotarget* **6**, 17430-17444, doi:10.18632/oncotarget.4172 (2015).
- 3 Wang, Y., Li, H., Sun, Q. & Yao, Y. Characterization of Small RNAs Derived from tRNAs, rRNAs and snoRNAs and Their Response to Heat Stress in Wheat Seedlings. *PLoS one* **11**, e0150933, doi:10.1371/journal.pone.0150933 (2016).
- 4 Ender, C. *et al.* A human snoRNA with microRNA-like functions. *Molecular cell* **32**, 519-528, doi:10.1016/j.molcel.2008.10.017 (2008).
- 5 Hasler, D. *et al.* The Lupus Autoantigen La Prevents Mis-channeling of tRNA Fragments into the Human MicroRNA Pathway. *Molecular cell* **63**, 110-124, doi:10.1016/j.molcel.2016.05.026 (2016).
- 6 Watkins, N. J. & Bohnsack, M. T. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* **3**, 397-414, doi:10.1002/wrna.117 (2012).
- 7 Dupuis-Sandoval, F., Poirier, M. & Scott, M. S. The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscip Rev RNA* **6**, 381-397, doi:10.1002/wrna.1284 (2015).
- 8 Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509-524, doi:10.1038/nrm3838 (2014).
- 9 Scott, M. S. & Ono, M. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie* **93**, 1987-1992, doi:10.1016/j.biochi.2011.05.026 (2011).
- 10 Falaleeva, M. & Stamm, S. Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. *BioEssays : news and reviews in molecular, cellular and developmental biology* **35**, 46-54, doi:10.1002/bies.201200117 (2013).
- 11 Falaleeva, M., Welden, J. R., Duncan, M. J. & Stamm, S. C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *BioEssays : news and reviews in molecular, cellular and developmental biology* **39**, doi:10.1002/bies.201600264 (2017).
- 12 Matera, A. G., Terns, R. M. & Terns, M. P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**, 209-220, doi:10.1038/nrm2124 (2007).
- 13 Palazzo, A. F. & Lee, E. S. Non-coding RNA: what is functional and what is junk? *Front Genet* **6**, 2, doi:10.3389/fgene.2015.00002 (2015).
- 14 Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* **3**, doi:10.1101/cshperspect.a003707 (2011).
- 15 Shi, Y. The Spliceosome: A Protein-Directed Metalloribozyme. *J Mol Biol*, doi:10.1016/j.jmb.2017.07.010 (2017).
- 16 Ruvkun, G. Molecular biology. Glimpses of a tiny RNA world. *Science* **294**, 797-799, doi:10.1126/science.1066315 (2001).
- 17 Schuldt, A. Great expectations of small RNAs. *Nat Rev Mol Cell Biol* **11**, 676, doi:10.1038/nrm2987 (2010).
- 18 Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862-864, doi:10.1126/science.1065329 (2001).

- 19 Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858-862, doi:10.1126/science.1065062 (2001).
- 20 Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-858, doi:10.1126/science.1064921 (2001).
- 21 Steinkraus, B. R., Toegel, M. & Fulga, T. A. Tiny giants of gene regulation: experimental strategies for microRNA functional studies. *Wiley Interdiscip Rev Dev Biol* **5**, 311-362, doi:10.1002/wdev.223 (2016).
- 22 Heo, I. *et al.* Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **151**, 521-532, doi:10.1016/j.cell.2012.09.022 (2012).
- 23 Berezikov, E., Chung, W. J., Willis, J., Cuppen, E. & Lai, E. C. Mammalian mirtron genes. *Molecular cell* **28**, 328-336, doi:10.1016/j.molcel.2007.09.028 (2007).
- 24 Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83-86, doi:10.1038/nature05983 (2007).
- 25 Cifuentes, D. *et al.* A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694-1698, doi:10.1126/science.1190809 (2010).
- 26 Yang, J. S. *et al.* Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15163-15168, doi:10.1073/pnas.1006432107 (2010).
- 27 Xie, M. *et al.* Mammalian 5'-capped microRNA precursors that generate a single microRNA. *Cell* **155**, 1568-1580, doi:10.1016/j.cell.2013.11.027 (2013).
- 28 Kim, Y. K., Kim, B. & Kim, V. N. Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E1881-1889, doi:10.1073/pnas.1602532113 (2016).
- 29 Martinez, I. *et al.* An Exportin-1-dependent microRNA biogenesis pathway during human cell quiescence. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E4961-E4970, doi:10.1073/pnas.1618732114 (2017).
- 30 Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654-665, doi:10.1016/j.cell.2013.03.043 (2013).
- 31 Moore, M. J. *et al.* miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nature communications* **6**, 8864, doi:10.1038/ncomms9864 (2015).
- 32 Grosswendt, S. *et al.* Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Molecular cell* **54**, 1042-1054, doi:10.1016/j.molcel.2014.03.049 (2014).
- 33 Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W. & Pasquinelli, A. E. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Molecular cell* **64**, 320-333, doi:10.1016/j.molcel.2016.09.004 (2016).
- 34 Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **162**, 84-95, doi:10.1016/j.cell.2015.06.029 (2015).
- 35 Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055-1067, doi:10.1016/j.cell.2012.10.036 (2012).
- 36 Lai, X., Wolkenhauer, O. & Vera, J. Understanding microRNA-mediated gene regulatory networks through mathematical modelling. *Nucleic acids research* **44**, 6019-6035, doi:10.1093/nar/gkw550 (2016).

- 37 Vera, J., Lai, X., Schmitz, U. & Wolkenhauer, O. MicroRNA-regulated networks: the perfect storm for classical molecular biology, the ideal scenario for systems biology. *Advances in experimental medicine and biology* **774**, 55-76, doi:10.1007/978-94-007-5590-1\_4 (2013).
- 38 Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233, doi:10.1016/j.cell.2009.01.002 (2009).
- 39 Hausser, J. & Zavolan, M. Identification and consequences of miRNA-target interactions-beyond repression of gene expression. *Nature reviews. Genetics* **15**, 599-612, doi:10.1038/nrg3765 (2014).
- 40 Khorshid, M., Hausser, J., Zavolan, M. & van Nimwegen, E. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nature methods* **10**, 253-255, doi:10.1038/nmeth.2341 (2013).
- 41 Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nature methods* **9**, 840-846, doi:10.1038/nmeth.2078 (2012).
- 42 Brown, B. D. *et al.* Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nature biotechnology* **25**, 1457-1467, doi:10.1038/nbt1372 (2007).
- 43 Brown, B. D., Venneri, M. A., Zingale, A., Sergi Sergi, L. & Naldini, L. Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nature medicine* **12**, 585-591, doi:10.1038/nm1398 (2006).
- 44 Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature genetics* **43**, 854-859, doi:10.1038/ng.905 (2011).
- 45 Kiss, T., Fayet, E., Jady, B. E., Richard, P. & Weber, M. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb Symp Quant Biol* **71**, 407-417, doi:10.1101/sqb.2006.71.025 (2006).
- 46 Dieci, G., Preti, M. & Montanini, B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* **94**, 83-88, doi:10.1016/j.ygeno.2009.05.002 (2009).
- 47 Massenet, S., Bertrand, E. & Verheggen, C. Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA biology* **14**, 680-692, doi:10.1080/15476286.2016.1243646 (2017).
- 48 Hirose, T. & Steitz, J. A. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12914-12919, doi:10.1073/pnas.231490998 (2001).
- 49 Hirose, T., Shu, M. D. & Steitz, J. A. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Molecular cell* **12**, 113-123 (2003).
- 50 Hirose, T. *et al.* A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Molecular cell* **23**, 673-684, doi:10.1016/j.molcel.2006.07.011 (2006).
- 51 Szczepinska, T. *et al.* DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Res* **25**, 1622-1633, doi:10.1101/gr.189597.115 (2015).
- 52 Ooi, S. L., Samarsky, D. A., Fournier, M. J. & Boeke, J. D. Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *Rna* **4**, 1096-1110 (1998).
- 53 Kim, J. W. *et al.* Human RNA lariat debranching enzyme cDNA complements the phenotypes of *Saccharomyces cerevisiae* dbr1 and *Schizosaccharomyces pombe* dbr1 mutants. *Nucleic acids research* **28**, 3666-3673 (2000).
- 54 Larochelle, M., Lemay, J. F. & Bachand, F. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic acids research* **40**, 10240-10253, doi:10.1093/nar/gks838 (2012).



- 55 Lubas, M. *et al.* Interaction profiling identifies the human nuclear exosome targeting complex. *Molecular cell* **43**, 624-637, doi:10.1016/j.molcel.2011.06.028 (2011).
- 56 Andersen, P. R. *et al.* The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nature structural & molecular biology* **20**, 1367-1376, doi:10.1038/nsmb.2703 (2013).
- 57 Norbury, C. J. Regional specialization: the NEXT big thing in nuclear RNA turnover. *Molecular cell* **43**, 502-504, doi:10.1016/j.molcel.2011.08.001 (2011).
- 58 Sloan, K. E., Schneider, C. & Watkins, N. J. Comparison of the yeast and human nuclear exosome complexes. *Biochem Soc Trans* **40**, 850-855, doi:10.1042/BST20120061 (2012).
- 59 Macias, S., Cordiner, R. A., Gautier, P., Plass, M. & Caceres, J. F. DGCR8 Acts as an Adaptor for the Exosome Complex to Degrade Double-Stranded Structured RNAs. *Molecular cell* **60**, 873-885, doi:10.1016/j.molcel.2015.11.011 (2015).
- 60 Boulon, S. *et al.* PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Molecular cell* **16**, 777-787, doi:10.1016/j.molcel.2004.11.013 (2004).
- 61 Leary, D. J., Terns, M. P. & Huang, S. Components of U3 snoRNA-containing complexes shuttle between nuclei and the cytoplasm and differentially localize in nucleoli: implications for assembly and function. *Mol Biol Cell* **15**, 281-293, doi:10.1091/mbc.E03-06-0363 (2004).
- 62 Terns, M. P. & Terns, R. M. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr* **10**, 17-39 (2002).
- 63 Knox, A. A. *et al.* A weak C' box renders U3 snoRNA levels dependent on hU3-55K binding. *Mol Cell Biol* **31**, 2404-2412, doi:10.1128/MCB.05067-11 (2011).
- 64 Chen, S. *et al.* SIRT7-dependent deacetylation of the U3-55k protein controls pre-rRNA processing. *Nature communications* **7**, 10734, doi:10.1038/ncomms10734 (2016).
- 65 Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonne, R. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Molecular cell* **9**, 891-901 (2002).
- 66 Pradet-Balade, B. *et al.* CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *The EMBO journal* **30**, 2205-2218, doi:10.1038/emboj.2011.128 (2011).
- 67 Girard, C. *et al.* Characterization of a short isoform of human Tgs1 methylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *The Journal of biological chemistry* **283**, 2060-2069, doi:10.1074/jbc.M704209200 (2008).
- 68 Sienna, N., Larson, D. E. & Sells, B. H. Altered subcellular distribution of U3 snRNA in response to serum in mouse fibroblasts. *Exp Cell Res* **227**, 98-105, doi:10.1006/excr.1996.0254 (1996).
- 69 Watkins, N. J., Lemm, I. & Luhrmann, R. Involvement of nuclear import and export factors in U8 box C/D snoRNP biogenesis. *Mol Cell Biol* **27**, 7018-7027, doi:10.1128/MCB.00516-07 (2007).
- 70 Kawaji, H. *et al.* Hidden layers of human small RNAs. *BMC genomics* **9**, 157, doi:10.1186/1471-2164-9-157 (2008).
- 71 Scott, M. S. *et al.* Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic acids research* **40**, 3676-3688, doi:10.1093/nar/gkr1233 (2012).
- 72 Ono, M. *et al.* Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic acids research* **39**, 3879-3891, doi:10.1093/nar/gkq1355 (2011).
- 73 Scott, M. S., Avolio, F., Ono, M., Lamond, A. I. & Barton, G. J. Human miRNA precursors with box H/ACA snoRNA features. *PLoS Comput Biol* **5**, e1000507, doi:10.1371/journal.pcbi.1000507 (2009).
- 74 Brameier, M., Herwig, A., Reinhardt, R., Walter, L. & Gruber, J. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic acids research* **39**, 675-686, doi:10.1093/nar/gkq776 (2011).

- 75 Hoyningen-Huene, P. Systematicity: The Nature of Science. *Philosophia* **36**, 167-180, doi:10.1007/s11406-007-9100-x (2008).
- 76 Falaleeva, M. *et al.* Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E1625-1634, doi:10.1073/pnas.1519292113 (2016).
- 77 He, X. *et al.* An Lnc RNA (GAS5)/SnoRNA-derived piRNA induces activation of TRAIL gene by site-specifically recruiting MLL/COMPASS-like complexes. *Nucleic acids research* **43**, 3712-3725, doi:10.1093/nar/gkv214 (2015).
- 78 Zhong, F. *et al.* A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. *Nucleic acids research* **43**, 10474-10491, doi:10.1093/nar/gkv954 (2015).
- 79 Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA Interactions. *Molecular cell* **62**, 618-626, doi:10.1016/j.molcel.2016.04.030 (2016).
- 80 Ono, M. *et al.* Analysis of human small nucleolar RNAs (snoRNA) and the development of snoRNA modulator of gene expression vectors. *Mol Biol Cell* **21**, 1569-1584, doi:10.1091/mbc.E10-01-0078 (2010).
- 81 Ausubel F.M., B. R. E., Kingston D.D., Seidmann J.R., Smith J.A., Struhl K. *Current Protocols in Molecular Biology.*, (Green Publishing Associates and John Wiley and Sons Inc., 1993).
- 82 Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. K. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *The Journal of biological chemistry* **252**, 1102-1106 (1977).
- 83 Weber K, P. J., Osborn M. Measurement of molecular weights by electrophoresis on SDS-crylamide gel. *PtC:3-27* **26** (1972).
- 84 Field, D. *et al.* Open software for biologists: from famine to feast. *Nature biotechnology* **24**, 801-803, doi:10.1038/nbt0706-801 (2006).
- 85 Team, R. C. A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* (2014).
- 86 RStudioTeam. *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA* <<http://www.rstudio.com/>> (2015).
- 87 Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* **5**, R80, doi:10.1186/gb-2004-5-10-r80 (2004).
- 88 Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* **42**, D68-73, doi:10.1093/nar/gkt1181 (2014).
- 89 Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic acids research* **34**, D158-162, doi:10.1093/nar/gkj002 (2006).
- 90 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 91 Travis, A. J., Moody, J., Helwak, A., Tollervey, D. & Kudla, G. Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods* **65**, 263-273, doi:10.1016/j.ymeth.2013.10.015 (2014).
- 92 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
- 93 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
- 94 Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database, C. The sequence read archive. *Nucleic acids research* **39**, D19-21, doi:10.1093/nar/gkq1019 (2011).

- 95 Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* **1**, 895-905, doi:10.3390/biology1030895 (2012).
- 96 Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620, doi:10.1093/bioinformatics/btt593 (2014).
- 97 Bai, B., Liu, H. & Laiho, M. Small RNA expression and deep sequencing analyses of the nucleolus reveal the presence of nucleolus-associated microRNAs. *FEBS Open Bio* **4**, 441-449, doi:10.1016/j.fob.2014.04.010 (2014).
- 98 Kishore, S. *et al.* Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome biology* **14**, R45, doi:10.1186/gb-2013-14-5-r45 (2013).
- 99 Macias, S. *et al.* DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nature structural & molecular biology* **19**, 760-766, doi:10.1038/nsmb.2344 (2012).
- 100 Rybak-Wolf, A. *et al.* A variety of dicer substrates in human and *C. elegans*. *Cell* **159**, 1153-1167, doi:10.1016/j.cell.2014.10.040 (2014).
- 101 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141, doi:10.1016/j.cell.2010.03.009 (2010).
- 102 Flores, O., Kennedy, E. M., Skalsky, R. L. & Cullen, B. R. Differential RISC association of endogenous human microRNAs predicts their inhibitory potential. *Nucleic acids research* **42**, 4629-4639, doi:10.1093/nar/gkt1393 (2014).
- 103 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.*, (Springer-Verlag New York, 2009).
- 104 gplots: Various R Programming Tools for Plotting Data (2016).
- 105 Granneman, S., Petfalski, E., Swiatkowska, A. & Tollervey, D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *The EMBO journal* **29**, 2026-2036, doi:10.1038/emboj.2010.86 (2010).
- 106 Hahne, F. *et al.* flowCore: a Bioconductor package for high throughput flow cytometry. *BMC bioinformatics* **10**, 106, doi:10.1186/1471-2105-10-106 (2009).
- 107 Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nature methods* **12**, 115-121, doi:10.1038/nmeth.3252 (2015).
- 108 Sladitschek, H. L. & Neveu, P. A. Bidirectional Promoter Engineering for Single Cell MicroRNA Sensors in Embryonic Stem Cells. *PLoS one* **11**, e0155177, doi:10.1371/journal.pone.0155177 (2016).
- 109 Luna, J. M. *et al.* Hepatitis C virus RNA functionally sequesters miR-122. *Cell* **160**, 1099-1110, doi:10.1016/j.cell.2015.02.025 (2015).
- 110 Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids research* **41**, D991-995, doi:10.1093/nar/gks1193 (2013).
- 111 Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9613-9618, doi:10.1073/pnas.0901997106 (2009).
- 112 Pundhir, S. & Gorodkin, J. Differential and coherent processing patterns from small RNAs. *Scientific reports* **5**, 12062, doi:10.1038/srep12062 (2015).
- 113 James, G., Witten, D., Hastie, T. & Tibshirani, R. *An introduction to statistical learning : with applications in R.* (Springer, 2013).
- 114 Friedlander, M. R. *et al.* Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome biology* **15**, R57, doi:10.1186/gb-2014-15-4-r57 (2014).
- 115 Granneman, S. *et al.* The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA in Vitro. *The Journal of biological chemistry* **277**, 48490-48500, doi:10.1074/jbc.M206631200 (2002).

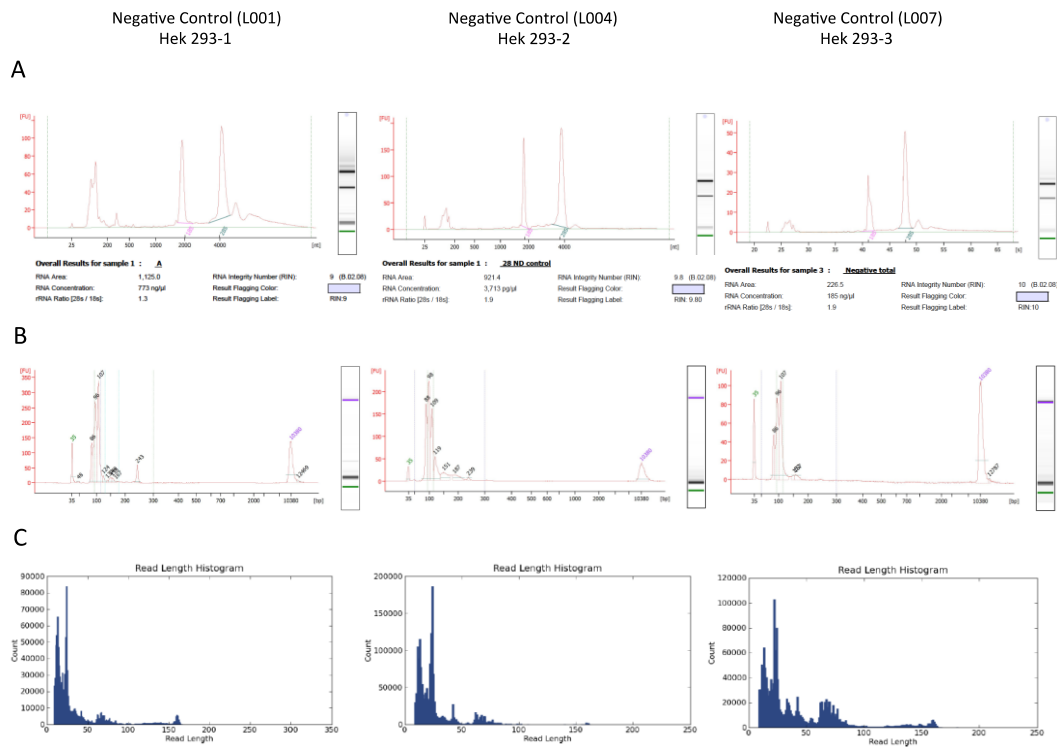
- 116 Pluk, H., Soffner, J., Luhrmann, R. & van Venrooij, W. J. cDNA cloning and characterization of the human U3 small nucleolar ribonucleoprotein complex-associated 55-kilodalton protein. *Mol Cell Biol* **18**, 488-498 (1998).
- 117 Buchler, N. E. & Louis, M. Molecular titration and ultrasensitivity in regulatory networks. *J Mol Biol* **384**, 1106-1119, doi:10.1016/j.jmb.2008.09.079 (2008).
- 118 Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. *PLoS Biol* **5**, e229, doi:10.1371/journal.pbio.0050229 (2007).
- 119 Denzler, R. *et al.* Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Molecular cell* **64**, 565-579, doi:10.1016/j.molcel.2016.09.027 (2016).
- 120 Lubas, M. *et al.* The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* **10**, 178-192, doi:10.1016/j.celrep.2014.12.026 (2015).
- 121 Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851-1854, doi:10.1126/science.1164096 (2008).
- 122 Lemus-Diaz, N. *et al.* Dissecting miRNA gene repression on single cell level with an advanced fluorescent reporter system. *Scientific reports* **7**, 45197, doi:10.1038/srep45197 (2017).
- 123 Bartel, D. P. & Chen, C. Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature reviews. Genetics* **5**, 396-400, doi:10.1038/nrg1328 (2004).
- 124 Iwasaki, Y. W., Siomi, M. C. & Siomi, H. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* **84**, 405-433, doi:10.1146/annurev-biochem-060614-034258 (2015).
- 125 Watkins, N. J. *et al.* Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Molecular cell* **16**, 789-798, doi:10.1016/j.molcel.2004.11.012 (2004).
- 126 Lukowiak, A. A. *et al.* Interaction of the U3-55k protein with U3 snoRNA is mediated by the box B/C motif of U3 and the WD repeats of U3-55k. *Nucleic acids research* **28**, 3462-3471 (2000).
- 127 Nabavi, S., Nellimarla, S. & Nazar, R. N. Post-transcriptional regulation of the U3 small nucleolar RNA. *The Journal of biological chemistry* **283**, 21404-21410, doi:10.1074/jbc.M802189200 (2008).
- 128 Winter, J. & Diederichs, S. Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA biology* **8**, 1149-1157, doi:10.4161/rna.8.6.17665 (2011).
- 129 Diederichs, S. & Haber, D. A. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097-1108, doi:10.1016/j.cell.2007.10.032 (2007).
- 130 Martinez, N. J. & Gregory, R. I. Argonaute2 expression is post-transcriptionally coupled to microRNA abundance. *Rna* **19**, 605-612, doi:10.1261/rna.036434.112 (2013).
- 131 Smibert, P., Yang, J. S., Azzam, G., Liu, J. L. & Lai, E. C. Homeostatic control of Argonaute stability by microRNA availability. *Nature structural & molecular biology* **20**, 789-795, doi:10.1038/nsmb.2606 (2013).
- 132 Janas, M. M. *et al.* Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. *Rna* **18**, 2041-2055, doi:10.1261/rna.035675.112 (2012).
- 133 Vasudevan, S. Functional validation of microRNA-target RNA interactions. *Methods* **58**, 126-134, doi:10.1016/j.ymeth.2012.08.002 (2012).

- 134 Thomson, D. W. *et al.* Assessing the gene regulatory properties of Argonaute-bound small RNAs of diverse genomic origin. *Nucleic acids research* **43**, 470-481, doi:10.1093/nar/gku1242 (2015).
- 135 Qi, H. H. *et al.* Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature* **455**, 421-424, doi:10.1038/nature07186 (2008).
- 136 Wu, C. *et al.* Hypoxia potentiates microRNA-mediated gene silencing through posttranslational modification of Argonaute2. *Mol Cell Biol* **31**, 4760-4774, doi:10.1128/MCB.05776-11 (2011).
- 137 Li, Z. *et al.* Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic acids research* **40**, 6787-6799, doi:10.1093/nar/gks307 (2012).
- 138 Keam, S. P. & Hutvagner, G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life* **5**, 1638-1651, doi:10.3390/life5041638 (2015).
- 139 Yu, F. *et al.* p53 Represses the Oncogenic Sno-MiR-28 Derived from a SnoRNA. *PloS one* **10**, e0129190, doi:10.1371/journal.pone.0129190 (2015).
- 140 Muller, S. *et al.* Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Molecular cancer* **14**, 94, doi:10.1186/s12943-015-0358-5 (2015).
- 141 Lykke-Andersen, S. *et al.* Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes & development* **28**, 2498-2517, doi:10.1101/gad.246538.114 (2014).
- 142 Hauptmann, J., Kater, L., Loffler, P., Merkl, R. & Meister, G. Generation of catalytic human Ago4 identifies structural elements important for RNA cleavage. *Rna* **20**, 1532-1538, doi:10.1261/rna.045203.114 (2014).

## 6 APPENDIX

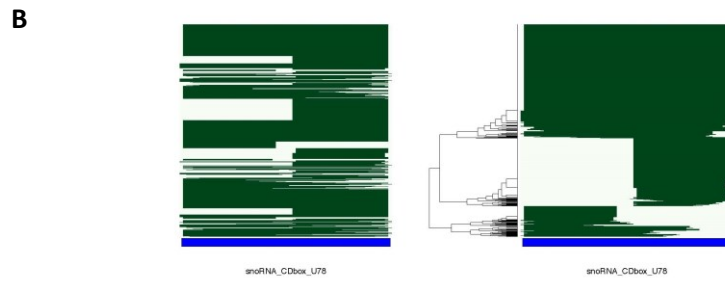
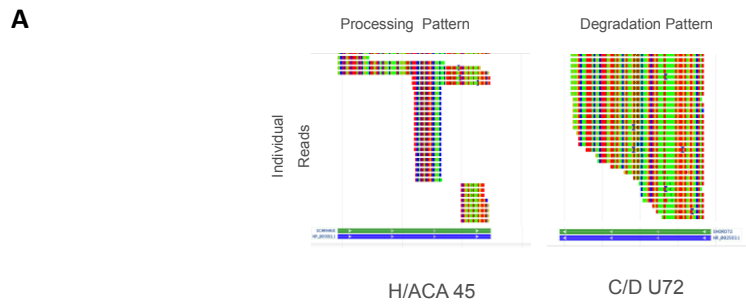
### SUPPLEMENTARY FIGURES.

Supplementary Figure 1 Total RNA and small RNA libraries quality from Hek293 cells.....	109
Supplementary Figure 2 Processing analysis of snoRNA locus .....	110
Supplementary Figure 3 snoRNA-derived miRNA processing and counts distribution .....	110
Supplementary Figure 4 U78 northern blot optimization. ....	111
Supplementary Figure 5 HeLa small RNA libraries quality and correlation.....	112
Supplementary Figure 6 U3 derived microRNA HeLa subcellular fractions .....	112
Supplementary Figure 7 Argonaute libraries correlation .....	113
Supplementary Figure 8 UTA-3. Dual Fluorescence Reporter System functionally characterized miRNA activity by microscopy. ....	114
Supplementary Figure 9 Analytical flow cytometry: Non cognate controls.....	114
Supplementary Figure 10 Transfection standardized conditions. ....	115
Supplementary Figure 11 Functional miRNA characterization.....	116
Supplementary Figure 12 Post-transcriptional miRNA activity from analytical flow cytometry in HeLa cell. .....	117
Supplementary Figure 13 Analytical flow cytometry .....	118



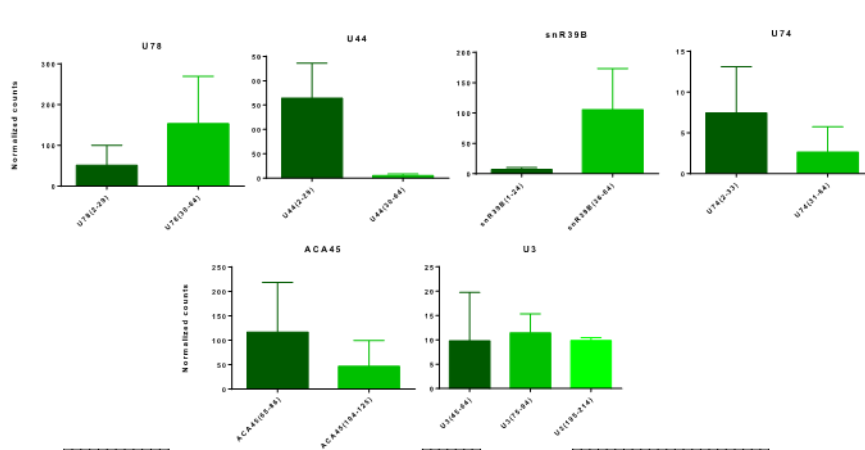
### Supplementary Figure 1 Total RNA and small RNA libraries quality from Hek293 cells

Three independent total RNA isolations were performed and RNA integrity tested by BioAnalyzer System (A). If RNA integrity number (RIN) was higher than 9 (B) Ion Torrent small RNA library preparation were performed and quantified and (C). Diluted library (18 pM) was sequenced via Ion-Torrent System.



**Supplementary Figure 2 Processing analysis of snoRNA locus**

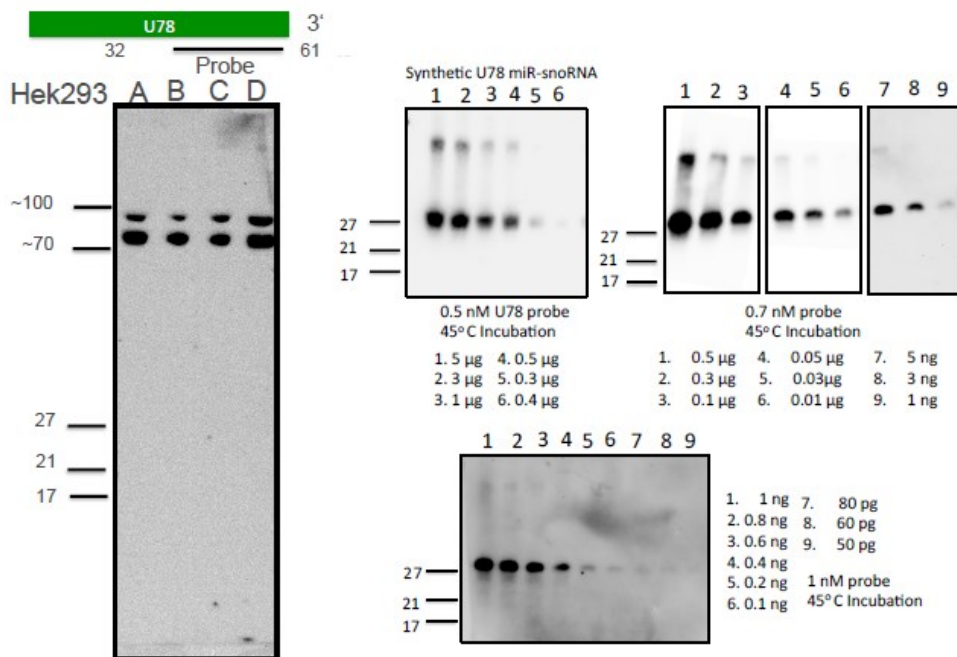
**(A)** Manual analysis on genomic tracks    **(B)** No clustered vs Clustered reads



**Supplementary Figure 3 snoRNA-derived miRNA processing and counts distribution**

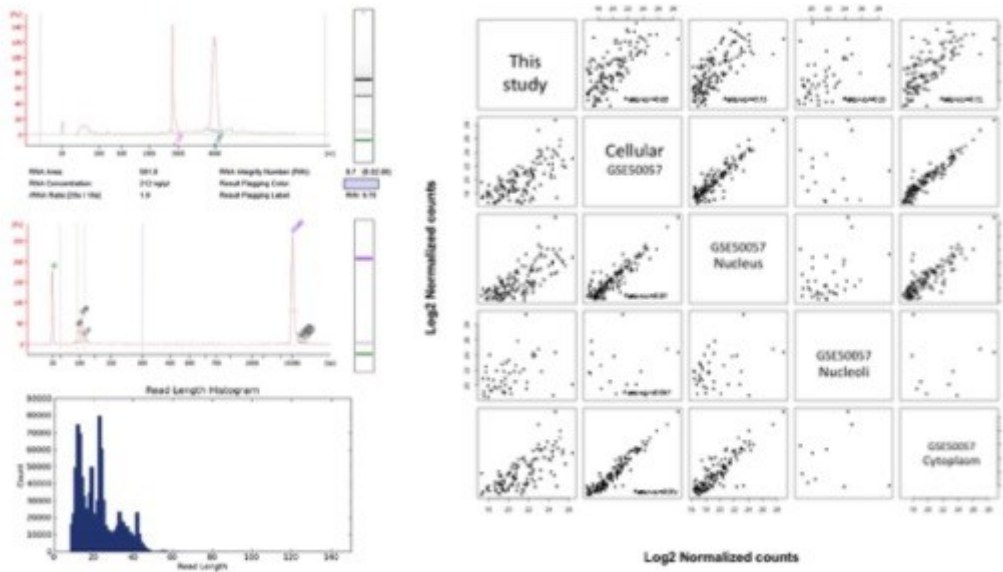
Counts distribution for three libraires, box plot showing normalized counts and location within the snoRNA locus, and density plots for miRNAs as example and the snoRNAs as read length.





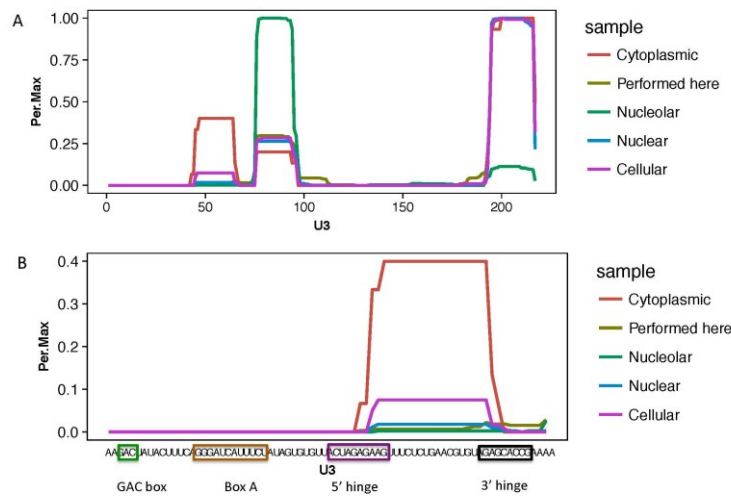
**Supplementary Figure 4 U78 northern blot optimization.**

Conditions used for U78 derived miRNA detection, using synthetic RNA, the blots show different probe concentrations and incubation temperatures. The blots plot different synthetic RNA and probe concentration. The concentration limit for U78 is pull down to 50 pg.



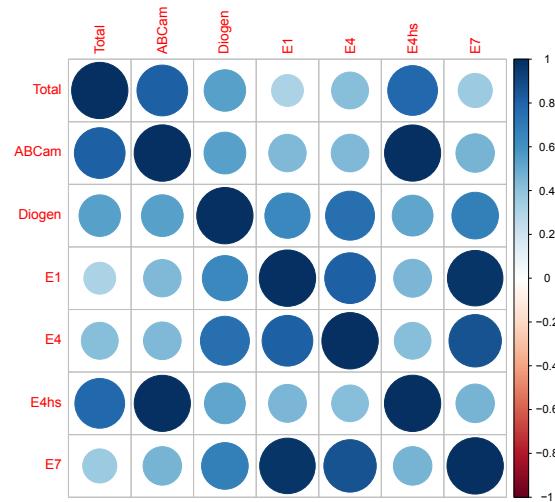
**Supplementary Figure 5 HeLa small RNA libraries quality and correlation.**

Total RNA quality and small RNA selection bioanalyzer, Read distribution from Ion-Torrent server and correlation between Dr. Böker and GSM50057 libraries.



**Supplementary Figure 6 U3 derived microRNA HeLa subcellular fractions**

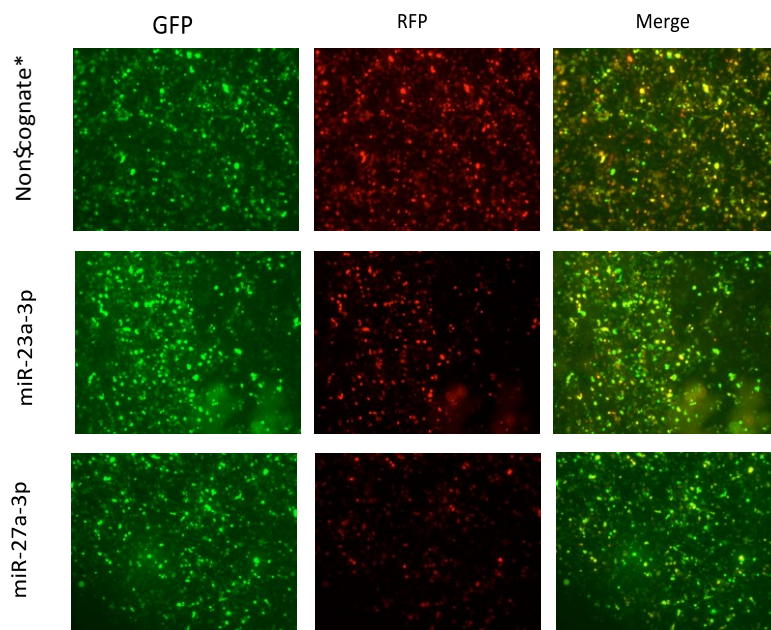
Density plots show read distributions from subcellular location derived NGS data (GSM50057) and the control performed in this thesis (Sequencing performed by Dr. Kai Böker). (A) Density plot U3 nucleotide positions as function of Percentage of Maximum and (B) U3 5' Domain density plot.

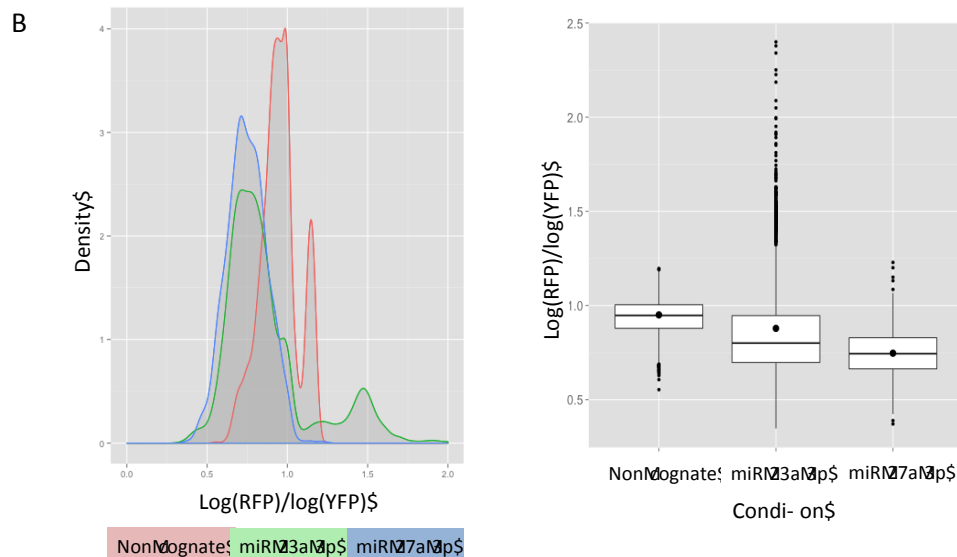


### Supplementary Figure 7 Argonaute libraries correlation

Correlation matrix plot of RIP and CLIP Argonaute data, circles represent the Pearson correlation and color code are located at the left side

A

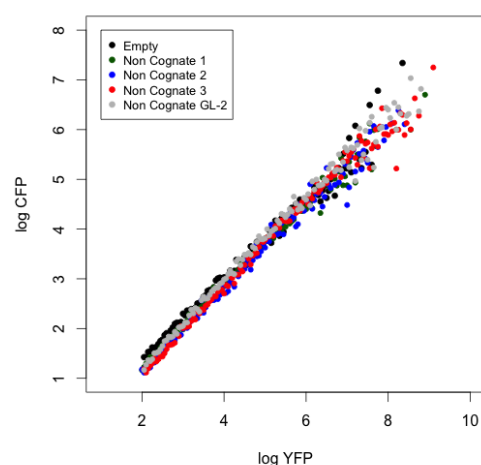




**Supplementary Figure 8 UTA-3. Dual Fluorescence Reporter System functionally characterized miRNA activity by microscopy.**

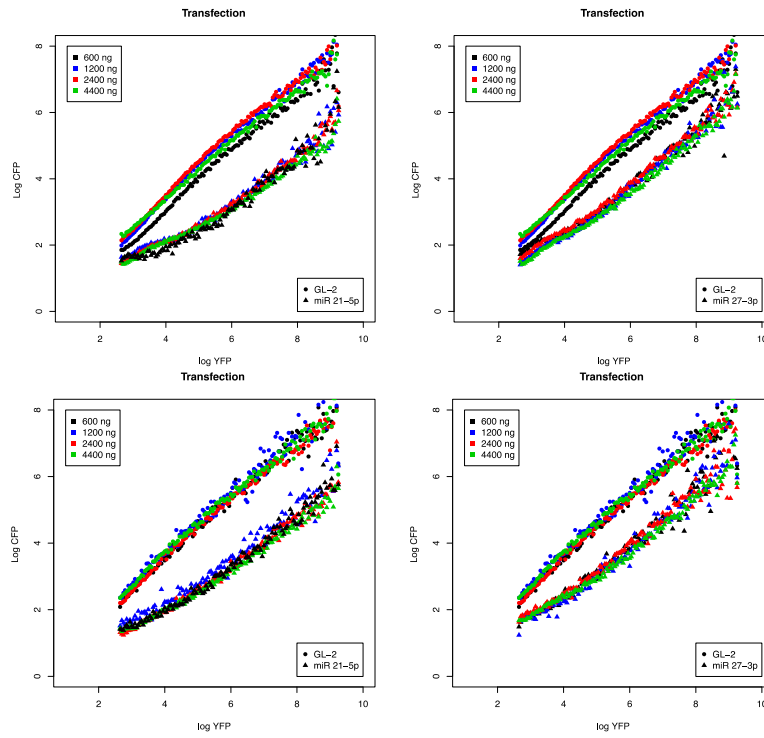
Untranslated Trans Assay (UTA) uses two independent fluorescent proteins expressed individually from two different promoters. RFP contained a perfect complementary target region for miRNAs within its 3' UTR while GFP is unaffected. Human HEK293 cells were transfected with three different sensor constructs (miR-23a-3p, miR-27a-3p and Non-cognate as control) and evaluated after 72 h. (A) Micrographs display variety of GFP and RFP expression in all the samples within a broad range of fluorescence intensities. Micrographs were taken with the same exposure time and with 10X objective. (B) Decreased ratios CFP/YFP for miRNA-targeted constructs from bulk quantitative analysis: pictures were analyzed with ImageJ and relative fluorescence intensity ratios ( RFP and GFP) for miRNA 23a-3p, miRNA 27a-3p and the non-cognate control were calculated using custom R scripts and imageJ.

(Transfections performed together with Ignacio Rodriguez during his Research internship under Lemus Supervision).



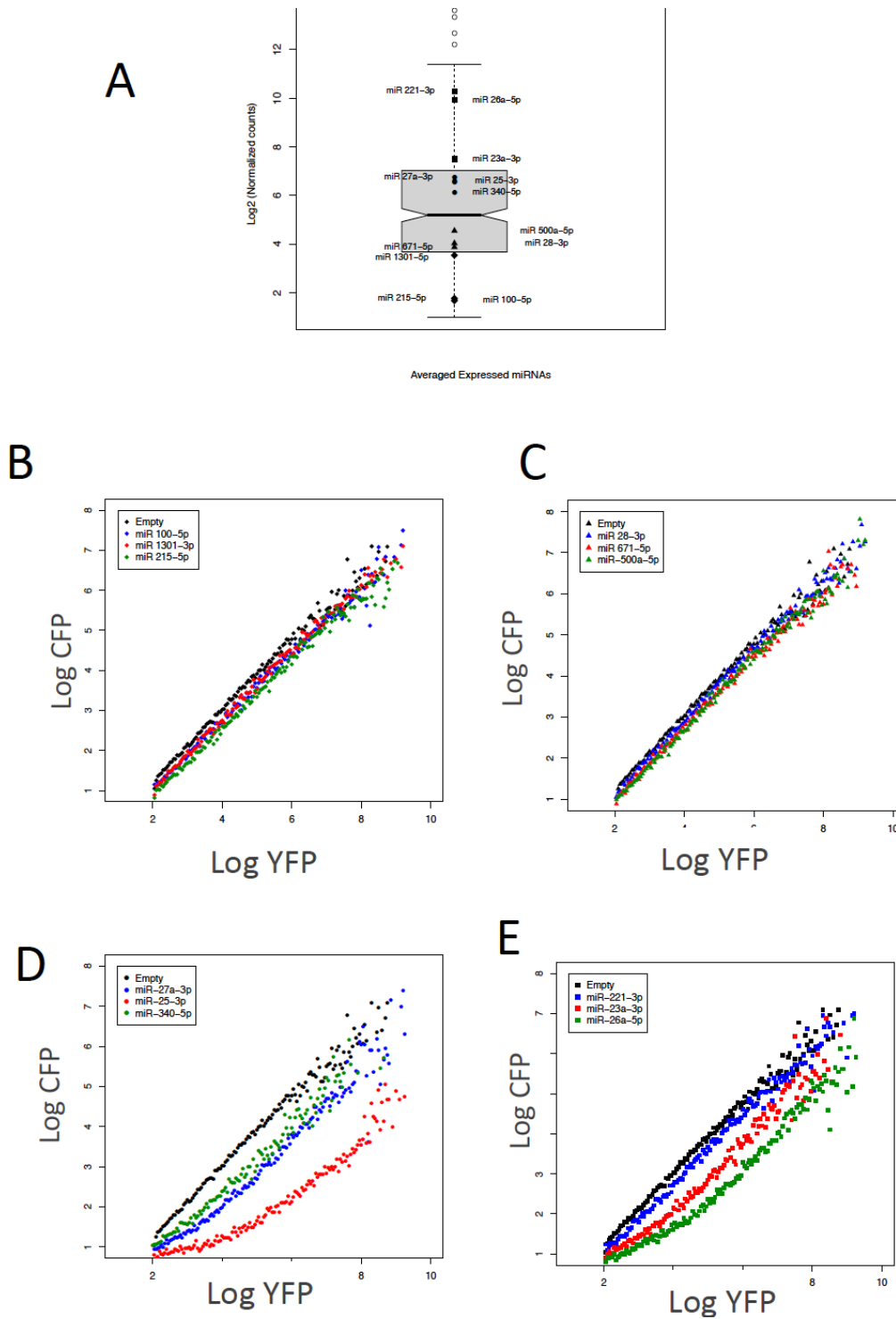
**Supplementary Figure 9 Analytical flow cytometry: Non cognate controls**

Several non-cognate controls were inserted on UTA-2 plasmids and evaluated on Hek293 cells 72 h post transfection



**Supplementary Figure 10 Transfection standardized conditions.**

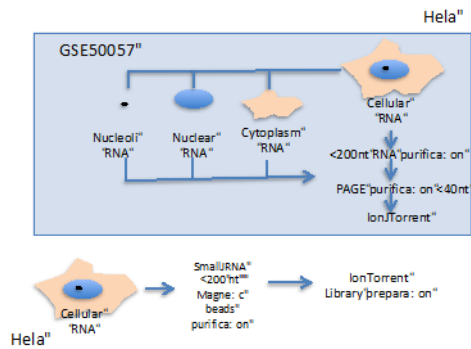
Several transfection using increasing concentration and collection times, Upper 24 and 48 hours, and bottom 72 h. Concentration are depicted.



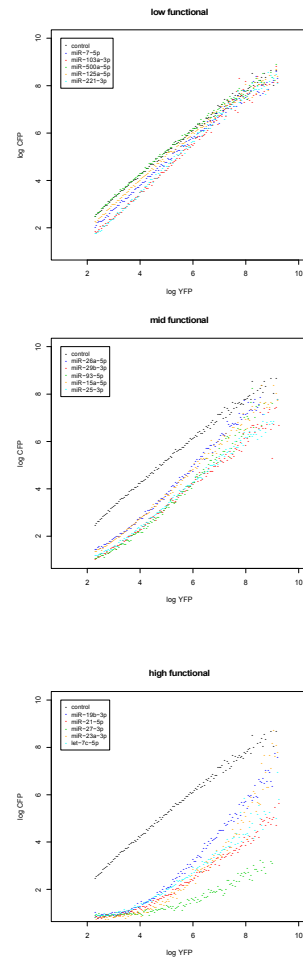
**Supplementary Figure 11 Functional miRNA characterization**

Hek293 cultures were sequenced by Ion-Torrent platform and (A) twelve different miRNAs were selected based on their expression levels (Mean Normalized Counts) and quartile position. 12 different reporters were generated and functionally tested using FACS. Analytical transfer function for candidates located: (B) below the first quartile, (C) between the median and the first quartile, (C) above median and the third quartile; (D) above the third quartile.

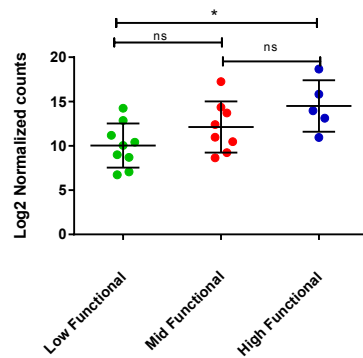
A



B

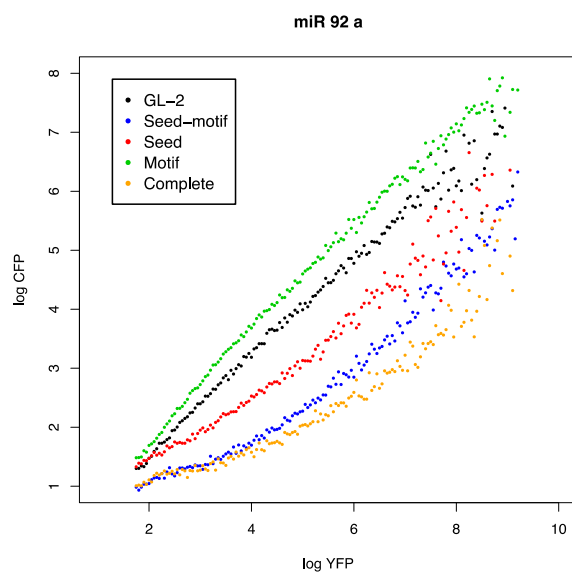


C



### Supplementary Figure 12 Post-transcriptional miRNA activity from analytical flow cytometry in HeLa cell.

GEO50057 retrieved raw data from HeLa small – RNA seq experiment was retrieved and compared with fresh produced total cell small RNA library (Raw data produced by Dr. Kai Böker). Fractionated cell compartment derived small RNA libraries were used for further analysis of nuclear, cytoplasmic and nucleolar content. (A) Experimental layout for HeLa small RNA library preparation in GSE50057 and Dr. Böker. miRNAs reporters were transfected into HeLa cells, evaluated after 72 h and (B) analytical functions were plotted. (C) Functions distributed into three ordinal variables as before (Figure 35) and cellular normalized counts were compared between groups (P < 0.01 Kruskal-Wallis test, Dunn's multiple comparisons test \* < 0.05)



### Supplementary Figure 13 Analytical flow cytometry

Seed, motif, seed+ motif or complete for miR92a were inserted in the dual fluorescent construct and tested for analytical flow cytometry.



PLASMID USED AND GENERATED IN THIS THESIS.

Plasmid	Observations
p.UTA.1.0 Empty	Also GL-2, miR23 and mir27
p.UTA.2.0 Empty.	All oligos in <b>Supplementary Table 4</b>
p.UTA.3.0 Empty	Oligos for 10 miRNAs
pGEM-Gas5	2 clones with distinct orientation
pTarget-Gas5	TA cloning
pGEM-IncGAS5	10 different splice variants
pAcGas5-GFP	As described in methods

Supplementary Table 1 Generated plasmid

Plasmid	Source
pLenti-CaMKIIa-eArchT 3.0-EYFP	gift from Karl Deisseroth (Addgene plasmid # 35513)
pcDNA3-CFP	Gift from Doug Golenbock (Addgene plasmid # 13030)
psiCheck2	Promega
pAcGFP-N1	Clontech
pGEM-T-Easy	Promega
pTarget	Promega
Plemir-RFP	Dharmacom
psiCheck2 mir23	Cloned by Henricke Höke
psiCheck2 mir27	Cloned by Henrike Höke

Supplementary Table 2 Used plasmids in this thesis

Plasmid	Addgene ID
p.UTA.2.0 miR-27a-3p	82438
p.UTA.2.0 miR-27a-3p 3 perfect copies.	82439
p.UTA.2.0-GL-2	82440
p.UTA.2.0-GL-2-3 perfect copies	82441
p.UTA.2.0 miR 19b-3p	82442
p.UTA.3.0 miR-21-5p	82445
p.UTA.2.0 Empty	82446
p.UTA.3.0 Empty	82447

Supplementary Table 3 Public available plasmids: submitted at <https://www.addgene.org>

## OLIGO SEQUENCES USED FOR MIRNA ANALYSIS AT SINGLE CELL LEVEL

<i>miRNA</i>	<i>oligos (5'-3' xhoI or NotI overhangs)</i>
<i>hsa-miR-100-5p</i>	<i>ggccagcttctttacagtgccttg</i> <i>tcgacaaggcagcactgtaagaagct</i>
<i>hsa-miR-1301-3p</i>	<i>ggcctgcagctgcctgggagtgacttc</i> <i>tcgagaagtcactcccaggcagctgca</i>
<i>hsa-miR-215-5p</i>	<i>ggccatgacctatgaattgacagac</i> <i>tcgagtctgtcaattcataggtcat</i>
<i>hsa-miR-28-3p</i>	<i>ggcccactagattgtgagctcctgga</i> <i>tcgatccaggagctcacaatctagtg</i>
<i>hsa-miR-29a-3p</i>	<i>ggcctagcaccatctgaatcggtta</i> <i>tcgataaccgatttcagatggtgcta</i>
<i>hsa-miR-24-3p</i>	<i>ggcctggctcagttcagcaggaacag</i> <i>tcgactgttctgctgaactgagcca</i>
<i>hsa-miR-451a</i>	<i>ggccaaaccgttaccattactgagtt</i> <i>tcgaaactcagtaatggaacggttt</i>
<i>hsa-miR 7-5p</i>	<i>ggcctggaagactagtgattttgtgt</i> <i>tcgaacaacaaaatcactagtcttcca</i>
<i>hsa-miR-15a-5p</i>	<i>ggcctagcagcacataatggtttgtg</i> <i>tcgacacaaaccattatgtgctgcta</i>
<i>hsa-miR-16-5p</i>	<i>ggcctagcagcacgtaaatattggcg</i> <i>tcgacgccaatatttcagtctgcta</i>
<i>hsa-miR-19b-3p</i>	<i>ggcctgtgcaaatccatgcaaaactga</i> <i>tcgatcagttttgcatggatttcaca</i>
<i>hsa-miR-26a-5p</i>	<i>ggccttcaagtaatccaggataggct</i> <i>tcgaagcctatcctggattactgaa</i>
<i>hsa-miR-29b-3p</i>	<i>ggcctagcaccatttgaatcagtggt</i> <i>tcgaaacactgattcaaatggtgcta</i>
<i>hsa-miR-93-5p</i>	<i>ggcccaaagtgtgttcgtgcaggtag</i> <i>tcgactacctgcacgaacagcactttg</i>
<i>hsa-miR-103a-3p</i>	<i>ggccagcagcattgtacagggtatga</i> <i>tcgatcatagccctgtacaatgctgct</i>
<i>hsa-miR-148a-3p</i>	<i>ggcctcagtgactacagaactttgt</i> <i>tcgaacaaagtctgtagtgcactga</i>
<i>hsa-miR-181a-5p</i>	<i>ggccaacattcaacgctgtcggtagt</i> <i>tcgaactcaccgacagcgttgaatgtt</i>
<i>hsa-miR-221-3p</i>	<i>ggccagctacattgtctgctgggttc</i> <i>tcgagaaaccagcagacaatgtagct</i>
<i>hsa-let-7c-5p</i>	<i>ggcctgaggtagtaggtgtatggtt</i> <i>tcgaaaccatacaacctactacctca</i>
<i>hsa-miR-18a-5p</i>	<i>ggcctaaggtgatctagtgcagatag</i> <i>tcgactatctgcactagatgcacctta</i>

<i>hsa-miR-19b-3p</i>	<i>ggcctgtgcaaatccatgcaaaactga</i>
	<i>tcgatcagtttgcatggattgcaca</i>
<i>hsa-miR-25-3p</i>	<i>ggccattgcacttgctcggctctga</i>
	<i>tcgatcagaccgagacaagtgaatg</i>
<i>hsa-miR-125a-5p</i>	<i>ggcctcctgagacccttaacctgtga</i>
	<i>tcgatcacaggtaaagggtctcaggga</i>
<i>hsa-miR-500a-5p</i>	<i>ggcctaactctgtctacctgggtgaga</i>
	<i>tcgatctcaccaggtagcaaggatta</i>
<i>hsa-miR-340-5p</i>	<i>ggccttataaagcaatgagactgatt</i>
	<i>tcgaaatcagtctcattgctttataa</i>
<i>hsa-miR-671-5p</i>	<i>ggccaggaagccctggaggggctggag</i>
	<i>tcgactccagcccctcagggttctct</i>
<i>hsa-miR-3607-5p</i>	<i>ggccgatgtgatgaagcaaatcagt</i>
	<i>tcgaactgattgcttcatcacatgc</i>
<i>hsa-miR-3607-3p</i>	<i>ggccactgtaaacgctttctgatg</i>
	<i>tcgacatcagaaagcgtttacagt</i>
<i>hsa-miR-21-5p</i>	<i>ggcctagcttatcagactgatgttga</i>
	<i>tcgatcaacatcagtctgataagcta</i>
<i>hsa-miR-27a 3p</i>	<i>ggccttcacagtggctaagttccgc</i>
	<i>tcgagcggaaacttagccactgtgaa</i>
<i>hsa-miR-27a 3-p b bulged</i>	<i>ggccttcacaggtctaagttccgc</i>
	<i>tcgagcggaaacttagacctgtgaa</i>
<i>GL-2</i>	<i>ggccccagtcacggaataacttcgaaa</i>
	<i>tcgatttcgaagtattccgcgtacgtg</i>
<i>GL-2 bulged</i>	<i>ggccccagtcacggttacttcgaaa</i>
	<i>tcgatttcgaagtagaccgcgtacgtg</i>
<i>U78</i>	<i>ggccatgtagacaaggtaaacct</i>
	<i>tcgaagtgttaccttctctacat</i>
<i>U3 1</i>	<i>ggccaagtttctctgaacgtgtag</i>
	<i>tcgactacacgttcagagaaactt</i>
<i>U3 2</i>	<i>ggccaccacgaggaagagaggtag</i>
	<i>tcgactacctctctcctcgtggt</i>
<i>U3 3</i>	<i>ggccggagagaacgcggtctgagt</i>
	<i>tcgaactcagaccggttctctcc</i>
<i>ACA45</i>	<i>ggccaaaggtagatagaacaggtc</i>
	<i>tcgagacctgttctatctacctt</i>
<i>U44</i>	<i>ggcccctggatgatgataagcaaa</i>
	<i>tcgatttgcttatcatcatccagg</i>
<i>GL-2 3 copies</i>	<i>ggccccagtcacggaataacttcgaaaccggcacgtacgcggaataacttcgaaaccggcacgtacgcggaataacttcgaaa</i>
	<i>tcgatttcgaagtattccgcgtacgtgccggttcgaagtattccgcgtacgtgccggttcgaagtattccgcgtacgtg</i>
<i>hsa-miR-27a 3p 3 copies</i>	<i>ggccttcacagtggctaagttccgcccgttcacagtggctaagttccgcccgttcacagtggctaagttccgc</i>
	<i>tcgagcggaaacttagccactgtgaaccggcggaacttagccactgtgaaccggcggaacttagccactgtgaa</i>
<i>hsa-miR-27a 3-p b bulged</i>	<i>tcgagcggaaacttagacctgtgaaccggcggaacttagacctgtgaaccggcggaacttagacctgtgaaccggcggaacttagacctgtgaa</i>
	<i>ggccttcacaggtctaagttccgcccgttcacaggtctaagttccgcccgttcacaggtctaagttccgcccgttcacaggtctaagttccgc</i>

U78	ggccatgtagacaaaggtaacact
	tcgaagtgttacctttgtctacat
hsa-miR-451a	ggccaaaccgttaccattactgagttccggaaccgttaccattactgagttccggaaccgttaccattactgagtt
	tcgaaactcagtaaatggaacggttccggaactcagtaaatggaacggttccggaactcagtaaatggaacggtt
hsa-miR-15a-5p	ggcctagcagcacataatggtttgtccggtagcagcacataatggtttgtccggtagcagcacataatggtttgtg
	tcgacacaaaccattatgtgctgctaccggcacaaccattatgtgctgctaccggcacaaccattatgtgctgctgta
hsa-miR-19b-3p	ggcctgtgcaaatccatgcaaaactgaccgggtgcaaatccatgcaaaactgaccgggtgcaaatccatgcaaaactga
	tcgatcagttttgcatggattgacacaccggcagttttgcatggattgacacaccggcagttttgcatggattgacaca
hsa-miR-148a-3p	ggcctcagtgactacagaactttgtccggtcagtgactacagaactttgtccggtcagtgactacagaactttgt
	tcgaacaaagttctgtagtgactgaccggacaaagttctgtagtgactgaccggacaaagttctgtagtgactgta
U3 1 3 copies	ggcceaagtttctgtaacgtgtagccggaagtttctgtaacgtgtagccggaagtttctgtaacgtgtag
	tcgactacacgttcagagaaactccggctacacgttcagagaaactccggctacacgttcagagaaact
U3 2 3 copies	ggccaccacgaggaagagaggtagccggaccacgaggaagagaggttagccggaccacgaggaagagaggttag
	tcgactacctcttctcctggtccggctacctcttctcctggtccggctacctcttctcctggt
U3 3 3 copies	ggccggagagaaacggtctgagtcggggagagaacggtctgagtcggggagagaacggtctgagt
	tcgaaactcagaccggttctctcccggactcagaccggttctctcccggactcagaccggttctctcccgg
ACA45 3 copies	ggcceaaggtagatagaacaggtccgggaaggtagatagaacaggtccgggaaggtagatagaacaggtc
	tcgagacctgttctatctacctttccggacctgttctatctacctttccggacctgttctatctaccttt
U44 3 copies	ggcccctggatgatgataagcaaacggcctggatgatgataagcaaacggcctggatgatgataagcaaa
	tcgatttgcttatcatcatccaggccggtttgcttatcatcatccaggccggtttgcttatcatcatccagg
U78 3 copies	ggccatgtagacaaaggtaacactccggatgtagacaaaggtaacactccggatgtagacaaaggtaacactaacgtaacact
	tcgaagtgttacctttgtctacatccggaggttacctttgtctacatccggaggttacctttgtctacat

Supplementary Table 4 Oligos used to insert into dual fluorescence reporter.

