HAL
open science

Identification and functional characterization of proteins involved in the regulation of viral microRNAs biogenesis

Antoine Creugny

To cite this version:

Antoine Creugny. Identification and functional characterization of proteins involved in the regulation of viral microRNAs biogenesis. Virology. Université de Strasbourg, 2019. English. NNT : 2019STRAJ063 . tel-03510328

HAL Id: tel-03510328

https://theses.hal.science/tel-03510328v1

Submitted on 4 Jan 2022

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

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

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

Institut de biologie moléculaire et cellulaire de Strasbourg

Architecture et réactivité de l'ARN – UPR9002

THÈSE présentée par :

Antoine CREUGNY

soutenue le : **15 octobre 2019**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaire et cellulaire de la biologie

Identification et caractérisation fonctionnelle de protéines impliquées dans la régulation de la biogénèse de miARN viraux

[Identification and functional characterization of proteins involved in the regulation of viral microRNAs biogenesis]

THÈSE dirigée par :

Mr PFEFFER Sébastien	Directeur de recherche, Université de Strasbourg

RAPPORTEURS externes :

Mr DIEDERICHS Sven	Professeur, DKFZ – Heidelberg; ZTZ - Freiburg
Mr DECROLY Etienne	Directeur de recherche, Université Aix Marseille

RAPPORTEUR interne :

Mme MEIGNIN Carine	Professeur, Université de Strasbourg

ACKNOWLEDGEMENTS

I'm grateful to my friends and family, who provide me unconditional support.

To all colleagues and scientists that gladly shared with me their time, knowledge and skills, I cannot thank you enough.

I would like to express my special appreciations and thanks to my advisors Aurélie Fender and Sébastien Pfeffer, you have been tremendous mentors for me.

A special gratitude goes to the jury members (Dr. E., Decroly Prof. S., Diederichs, Prof. C. Meignin) who took time to read and criticize my work.

Table of Contents

I. Introduction 1

I.A. miRNAs 3
I.A.1 miRNA history and nomenclature 3
I.A.2 miRNA biogenesis & regulation 4
I.A.2.a Non canonical miRNA biogenesis pathway 22
I.A.3 miRNA function 23
I.A.3.a miRNA translational repression 23
I.A.3.b miRNA target recognition 24
I.A.3.c miRNA and target abundance 24
I.A.4 Clustered miRNA 24
I.A.4.a miRNA evolution 25
I.A.4.b Regulation of miRNA cluster expression 25
I.A.4.c Regulation of clustered miRNA specific feature 25
I.A.5 miRNA & host-virus interaction 26
I.A.5.a Virus-mediated host miRNAs turnover and hijack 26
I.A.5.b Virus-encoded miRNAs 26

I.B. KSHV 27
I.B.1 Pathogenesis 28
I.B.2 Viral cycle 29
I.B.3 Latent cycle 30

I.C. KSHV's miRNA 30

I.D. RNA binding proteins 33
I.D.1 RNA binding domains: sequence, structure and context preferences 33
I.D.2 RNPs & membrane-less organelles 34

II. Thesis objectives 35

III. Results 36

III.A. RNA affinity chromatography for identification of factors binding specific KSHV MIRNA 36
III.A.1 Agarose RNA-pulldown 37
III.A.1.a SL-miRNAs production and quality 37
III.A.1.b SL-miRNAs beads binding 38
III.A.1.c Nuclear extract quality 39
III.A.1.d Agarose RNA-pulldown 39
III.A.1.e Mass spectrometry data exploration 40
III.A.2 Magnetics RNA-pulldown for sensitivity and specificity increase beads 48
III.A.2.a Chromatography set up 48
III.A.2.b Pre-adenylation step 48
III.A.2.c SL-miRNA and ssDNA ligation 49
III.A.2.d Magnetic beads binding, washing and elution 50
III.A.2.e Heat vs DNase treatment elution 52
III.A.3 Identification of microRNA precursor binding proteins by DNase-assisted RNA chromatography coupled with mass spectrometry 53
III.A.3.a Introduction to the DNase-assisted RNA chromatography experiment 53
III.A.3.b DNase-assisted RNA chromatography results 55
III.A.3.c DNase-assisted RNA chromatography assay and controls 56
III.A.3.d Mass spectrometry data analysis 58
III.A.3.d.1.1 Filtration and imputation 58
III.A.3.d.1.2 Descriptive analysis 60
III.A.3.d.1.3 Differential analysis 62
III.A.3.d.1.4 Discussion and conclusion 64
III.A.4 DNase assisted RNA chromatography with KSHV SL-miRNA 66
III.A.4.a Bait formation and elution control 66

III.A.4.b MSMS data exploration 68
III.A.4.b.1.1 B.CWM_MSMS "home-made" analysis. 68
III.A.4.b.1.2 ProStaR data mining 70
III.A.4.b.1.3 Day1: SL-miRNA-K1, -K2, -K3, -K5, -let7 70
III.A.4.b.1.4 Day2: SL-miRNA-K6, -K7, -K8, -K9, -let7 76
III.A.4.b.1.5 Day3: SL-miRNA-K11, -155, -let7 81
III.A.4.b.1.6 DG75: SL-miRNA-K11, -155, -let7 86
III.A.4.b.1.7 Day3 versus DG75 91
III.A.4.b.1.1 SL-miRNA-let7a biological triplicate 96

III.B. Candidates validation 99
III.B.1 Tools developed for the validation 99
III.B.1.a Modulation of candidate expression 99
III.B.1.a.1.1 siRNA & shRNA 100
III.B.1.a.1.2 Overexpression of HA-tag 101
III.B.1.b Luciferase reporters 102
III.B.1.b.1.1 DROSHA cleavage reporter 103
III.B.1.b.1.2 miRNA activity reporter 104
III.B.1.c Northern blot & qRT-PCR 105
III.B.2 Validation of protein involvement 107
III.B.2.a Confirmation of interaction 107
III.B.2.a.1.1 Endogenous proteins 107
III.B.2.a.1.2 HA-tagged proteins 110
III.B.2.b Impact on miRNA biogenesis 112
III.B.2.b.1.1 Luciferase screen assays 112
III.B.2.b.1.2 DROSHA cleavage reporter assay 112
III.B.2.b.1.3 psiCHECK-2: mature miRNA activity sensor 115
III.B.2.b.1.4 miRNA biogenesis variation upon individual candidate modulation 117
III.B.2.b.1.5 Heterologous system 117
III.B.2.b.1.6 KSHV naturally infected lymphocyte system 120

IV. Discussion 124

IV.A. RNA chromatography 125

IV.B. Luciferase screen assays 128

IV.C. RNPs, candidates and expressing vectors 129
IV.C.1 Selected candidates 130
IV.C.2 Candidates silencing and overexpression 131

IV.D. Candidate modulation and effect on miRNA 132

IV.E. RNPs fascinating world 134

V. Conclusion and perspectives 135

VI. Materials and methods 140

VI.A. Plasmids 140
VI.A.1 Plasmids expressing KSHV's miRNAs 140
VI.A.2 Plasmids over-expressing candidate 140
VI.A.3 Plasmids expressing candidate's shRNA 141
VI.A.4 Plasmids pmiRGLO: SL-cleavage reporter 142
VI.A.5 Plasmids psiCHECK-2: miRNA activity reporter 142
VI.A.6 Bacterial transformation and plasmids purification 143

VI.B. Cell culture 143
VI.B.1 Cells growth 143
VI.B.2 Transfection 143
VI.B.3 Lentivirus production 144
VI.B.4 Lentivirus transduction 144

VI.C. RNA pulldown 144
VI.C.1 PCR template 144

VI.C.2 In vitro transcription 145
VI.C.3 SL-miRNA binding to agarose beads 145
VI.C.4 SL-miRNA binding efficiency assessment 146
VI.C.5 Chimeric SL-miRNA generation and beads binding 146
VI.C.6 Nuclear extract 146
VI.C.7 Nuclear extract quality 147
VI.C.8 Pulldown & elution 147

VI.D. MSMS identification 147
VI.D.1 Spectral count 147
VI.D.2 MaxLFQ 148

VI.E. Protein analysis 149
VI.E.1 Protein extraction and quantification 149
VI.E.2 Western blot 149
VI.E.3 Silver nitrate staining 150

VI.F. RNA analysis 151
VI.F.1 RNA extraction and quantification 151
VI.F.2 Northern blot 151
VI.F.3 Reverse transcriptase and quantitative PCR 152

VI.G. Western blot and northern blot signal quantification 153

VI.H. Luciferase assays 153

VII. Bibliography 155

Table of figures

Figure 1. Central dogma 1
Figure 2. miRNA biogenesis.....22
Figure 3. KSHV viral cycle & viral miRNA. 29
Figure 4. KSHV latent genomic locus and miRNA precursors 31
Figure 5. SL-miRNA production and beads attachment 38
Figure 6. Nuclear extract preparation 39
Figure 7. Descriptive analysis of the four chromatography datasets with the SL-miRNA-K4 bait 41
Figure 8. SL-miRNA-K4 significantly enriched dataset by the 4 different cell lines 42
Figure 9. Descriptive analysis of the four chromatography datasets with the SL-miRNA-K11 bait 43
Figure 10. SL-miRNA-K11 significantly enriched dataset in the 4 different cell lines 44
Figure 11. Descriptive analysis of the four chromatography datasets with the SL-miRNA-let7 bait 45
Figure 12. SL-miRNA-let7 significantly enriched dataset by the 4 different cell lines 46
Figure 13. Pre-adenylation control.....49
Figure 14. Chimera formation.....50
Figure 15. Chimera binding, washing and elution efficiency 51
Figure 16. DNase and heat elution comparison 52
Figure 17. Graphical abstract of the DNase-assisted RNA chromatography and the mass spectrometry data analysis pipeline.....54
Figure 18. Bait formation and elution controls 57
Figure 19. Overview of the matrix dataset before and after the imputation procedure.....59
Figure 20. Descriptive analysis of the dataset.....61
Figure 21. Differential analyses performed on the dataset 63
Figure 22. Enriched protein dataset comparison.....64
Figure 23. Bait formation and elution controls 67
Figure 24. p-values-heatmap of the different RNA chromatography 69
Figure 25. Overview of the Day1 matrix dataset before and after imputation procedure 71
Figure 26. Descriptive analysis of the Day1 chromatography dataset 73
Figure 27. Differential analysis of the Day1 dataset 75
Figure 28. Overview of the matrix dataset before and after imputation procedure of the Day2 RNA chromatography.....77
Figure 29. Descriptive analysis of the Day2 chromatography dataset 78
Figure 30. Differential analysis of the Day2 dataset 80
Figure 31. Overview of the Day3 matrix dataset before and after imputation procedure 82
Figure 32. Descriptive analysis of the Day3 chromatography dataset 83
Figure 33. Differential analysis of the Day3 dataset 85
Figure 34. Overview of the DG75 matrix dataset before and after imputation procedure 87
Figure 35. Descriptive analysis of the DG75 chromatography dataset 88

Figure 36. Differential analysis of the DG75 dataset 90
Figure 37. Descriptive analysis of the Day3 and DG75 RNA chromatography 91
Figure 38. Differential analysis of the DG75 and Day3 RNA chromatography dataset for the SL-miRNA let-7a-1 93
Figure 39. Differential analysis of the DG75 and Day3 RNA chromatography dataset for the SL-miRNA -11 94
Figure 40. Differential analysis of the DG75 and Day3 RNA chromatography dataset for the SL-miRNA -155. 95
Figure 41. Descriptive analysis of the SL-miRNA-let7 biological triplicate dataset 97
Figure 42. Enriched dataset comparison of the SL-miRNA-let7 biological triplicate generated with the DNase assisted RNA chromatography 98
Figure 43. Silencing activity of the shRNA on the candidate mRNA 101
Figure 44. pLenti6.3 expression control 102
Figure 45. Schematic representation of the pmiRGLO reporter system 103
Figure 46. Schematic representation of the psiCHECK-2 reporter system 104
Figure 47. Schematic representation of the primer couples and pri-miRNA normalization 106
Figure 48. Western blot analysis of the magnetic RNA-pulldown 108
Figure 49. Agarose RNA-pulldown with HA-candidate 111
Figure 50. pmiRGLO luciferase assay 114
Figure 51. psiCHECK-2 luciferase assays 116
Figure 52. Candidate over-expression in heterologous system. 119
Figure 53. Candidate over-expression and silencing within infected BC-3 cell line 121
Figure 54. KSHV pri-miRNA assessment by QRT-PCR of transduced BC-3 123
Figure 55. Representation of specific subcellular localization of RNP complexes (granules, foci…) 135
Figure 56. Representation of experiments needed to assess cofactor involvement in specific miRNA biogenesis 137
Figure 57. Representation of the pri-miRNA fold dynamism upon cleavage 138

Abbreviations

A: Adenine
AGO: Argonaute
BCBL: Body Cavity Base Lymphoma
BLV: Bovine leukemia virus
bp: base pair
C: Cytosine
cDNA: complementary DNA
CLASH: Cross-Linking Ligation And Sequencing of Hybrids
CLIP: Cross-Linking ImmunoPrecipitation
CMV: CytoMegaloVirus
DAPI: 4',6'-diamidino-2-phénylindole
DG75: Burkitt's lymphoma derived B lymphocytes
DGCR8: DiGeorge syndrome Critical Region 8
DMEM: Dulbecco's Modified Eagle Medium
DNA: DeoxyriboNucleic Acid
DNase: Deoxyribonuclease
ds: double stranded
DTT: Dithiothreitol
EBV: Epstein Barr Virus
EDTA: EthyleneTiaminetetraAcetic acid
FC: Fold Change
G: Guanine
HEK293T: Human Embryonic Kidney transformed with SV40 large T antigen
HHV-8: Human Herpesvirus 8
HITS-CLIP: High-Throughput Sequencing Cross-Linking Immunoprecipitation
KS: Kaposi Sarcoma
KSHV: Kaposi's Sarcoma-associated HerpesVirus
LANA: Latency Associated Nuclear Antigen
LC-MS/MS: Liquid Chromatography coupled to tandem Mass Spectrometry
MEC: Missing on an Entire Condition
MHV68: Murine gamma-HerpesVirus 68
miRNA - miR: microRNA
mRNA: messenger RNA
NB: Northern Blot
NFκB: Nuclear Factor kappa-light-chain-enhancer of activated B cells
ncRNA: non-coding RNA
nts: nucleotides
ORF: Open Reading Frame
PBS: Phosphate Buffer Saline
PCI: Phenol Chloroform Isoamyl acid
PCR: Polymerase Chain Reaction
PEL: primary effusion lymphoma
piRNA: piwi-interacting RNA
POV: Partially Observed Value
pre-miRNA - pre-miR: precursor miRNA
pre-mRNA: precurcor messenger RNA
pri-miRNA - pri-miR: primary miRNA transcript
qRT-PCR: quantitative Reverse-Transcribed PCR
RBD: RNA binding domain
RBP: RNA Binding Protein
RISC: RNA Induced Silencing Complex
RNA: ribonucleic acid
RNAi: RNA interference
RNase: Ribonuclease
RNP: Ribonucleoprotein
RPMI: Roswell Park Memorial Institute medium
RPS21: Ribosomal Protein S21
rNTP: ribonucleotide triphosphate
sdRNAs: snoRNA-derived miRNAs
SDS: Sodium Dodecyl Sulfate
SFV: Simian foamy virus
shRNA: short hairpin RNA
siRNA: small interfering RNA
SL: Stem Loop
ss: single stranded
T: thymine
TAE: Tris Acetate EDTA buffer
TBE: Tris Borate EDTA buffer
tdRNAs: tRNA-derived miRNAs
tRNA: transfer RNA
U: Uracil
UTR: UnTranslated Region
WB: Western Blot
5p: 5' strand of miRNA duplex
3p: 3' strand of miRNA duplex

I. INTRODUCTION

Biology comes from the Greek words *bios* and *logos,* which respectively mean *life* and *study.* Biology has been explored throughout human history with different dogmas and tools on various levels from an entire population or organism down to its molecular aspect. Modern biology denotes the science that studies all life features, properties and evolutionary history of all life forms (bacteria, archaea, protozoa, plantae, fungi and metazoa). It is based on several fundamental notions: the cell is the basic unit of life; all life forms share a common ancestor; genes composed of deoxyribonucleic acid (DNA) are the support of life information and proteins are the products of life information. Based on it, life and living organisms can be defined with several properties among others: a capacity of self-replication inducing genetic variability (evolution), a capacity for self-regulation (homeostasis) and a capacity for responding to environmental stimuli. Taken together and at the cellular level, it creates the central dogma of modern biology (Figure 1). Two cellular machineries composed of protein and ribonucleic acid (RNA), the transcription and translation machinery, are indispensable to all living organisms. These machineries allow the cell to transform genetic information (DNA) into an intermediate molecule (RNA), which acts or not as a template to allow the generation of a polypeptide chain. Consequently, expression of the genetic information can be regulated at those different steps namely: transcription, post-transcription, translation, and post-translation. All of those regulation processes can be shared or specific to a previously cited phylum. Shared regulation processes between different phylum come with different flavours concerning their biogenesis and function. This manuscript focuses on mammals and their viruses, but other metazoan phylum will be referred to in the introduction.

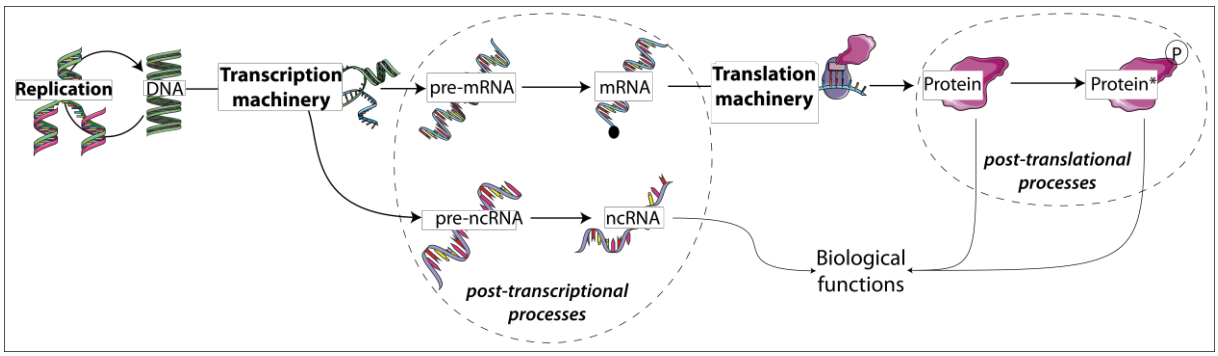


Figure 1. Central dogma. DNA is transcribed into pre-messenger RNA (pre-mRNA) or pre-non-coding RNA (pre-ncRNA) and maturated though post-transcriptional processes. Mature mRNA is then translated into protein which can undergo post-translation modifications. Both proteins and ncRNAs participate to biological functions.

As viruses are obligatory intracellular parasites, they are able to highjack the cellular machinery of their host cell as well as all the genetic expression regulation processes available. Here, I will only expose and describe one small part of the genetic post-transcriptional regulation process mediated by small non-coding RNAs called microRNAs (miRNAs). Those regulatory RNAs are one kind of effector molecules found in RNA silencing processes, which also contain small interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs). miRNA-mediated post transcriptional process is used by several virus family notably the herpesvirus

family including the Kaposi's sarcoma associated herpesvirus (KSHV). More precisely, this manuscript focuses on the regulation of KSHV miRNAs expression. First, I will introduce the miRNA pathway from its biogenesis to its function, and its regulation. I will then introduce the KSHV and its miRNAs, before presenting RNA binding proteins properties. Finally, I will present the thesis project and objectives.

I.A. miRNAs

miRNAs are short non-coding RNAs of ~22 nucleotides (nt) in length that guide ARGONAUTE (AGO) proteins to a target messenger RNA (mRNA) to mediate its translation inhibition and/or destabilization (Bartel, 2018). Since its discovery, gene silencing mediated by RNA molecules also called RNA silencing or RNA interference (RNAi), has emerged as a fundamental post-transcriptional regulation mechanism shared by a vast majority of eukaryotes, involved in almost, if not all, cellular processes from development to homeostasis (Meister and Tuschl, 2004).

I.A.1 miRNA history and nomenclature

The first miRNA, although it was not referred to as such, was described in 1993 in a paper by Ambros and collaborators, where they describe the role of the *lin-4* gene in the heterochronic development of *C. elegans* (Lee *et al.*, 1993). At this time, it was surprising to find that this *lin-4* gene does not produce an mRNA but rather a small non-coding RNA (ncRNA) of ~22 nts length, which possesses partial sequence complementarity for the *LIN-14* mRNA. This discovery together with another small ncRNA, Let-7, a few years later (Reinhart *et al.*, 2000), laid the basis of the current model in which these small ncRNAs are able to mediate translational repression of the targeted mRNA through miRNA/target base-pairing (reviewed in (Farh, 2005)). Rapidly, scientist looked for other small ncRNAs of ~22 nts and discovered a large number of those molecules, which they coined microRNAs (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee, 2001). Since then, the interest in miRNAs blew up and thousands of different miRNAs were identified in the metazoan phylum highlighting a new regulation process able to target any mRNAs. In addition, the discovery of conserved miRNAs among metazoan phylum together with their requirement for development, clearly suggest their fundamental importance (Friedman *et al.*, 2008). Indeed, miRNAs are involved in all fundamental cellular processes and their mis-expression is responsible or correlated to an important number of pathologies including cancer (Esteller, 2011; Lin and Gregory, 2015). Meanwhile, viral miRNAs have been first identify in herpesvirus family (Pfeffer *et al.*, 2004, 2005), other DNA virus family (Sullivan *et al.*, 2005), and in retroviruses (Kincaid *et al.*, 2012, 2014). The mechanism of miRNA-mediated translation inhibition is detailed further below.

The first miRNAs discovered were named after their mutant phenotype (e.g. let-7 stands for *lethal-7*). However, a multitude of miRNA have been identified without associated phenotype. Thus, for the sake of database understanding, they were identified with number sequentially attributed (e.g. miR-*number*). Efforts have been made to attribute similar number to orthologous miRNAs. For instance, *C. elegans* let-7 is shared with other species, consequently human orthologs are also labelled let-7. To differentiate the species, three letters have been added to the miRNA name, hsa-let-7 and cel-let-7 respectively correspond to the human and *C. elegans* let-7 miRNA. Paralog miRNAs within species are designed with letter suffixes (a, b, …) and distinct precursor sequences and genomic loci that express similar miRNAs are labelled with number suffixes (1, 2, …), for instance hsa-let-7a-1. Among the miRNA duplex (see part I.A.2 for biogenesis), 5' and 3' arms are referred to respectively as -5p and -3p. In addition, the miRNA arm which is less abundant due to AGO loading is sometimes named the miRNA *star* sequence (e.g. hsa-let-7a-1-3p*) (Ambros, 2003).

I.A.2 miRNA biogenesis & regulation

The miRNA biogenesis and its nuclear regulation is introduced in the following review.

Creugny_et_al-2018-FEBS_Letters
https://doi.org/10.1002/1873-3468.13067

FEBS Letters

FEBS PRESS
science publishing by scientists

REVIEW ARTICLE

Regulation of primary microRNA processing

Antoine Creugny, Aurélie Fender and Sébastien Pfeffer

Architecture and Reactivity of RNA, Institut de Biologie Moléculaire et Cellulaire du CNRS, Université de Strasbourg, France

Correspondence
A. Fender and S. Pfeffer, Architecture and Reactivity of RNA, Institut de Biologie Moléculaire et Cellulaire du CNRS, Université de Strasbourg, 15 rue René Descartes, 67084 Strasbourg France
Fax: +33 3 88 60 22 18
Tel: +33 3 88 41 70 60
E-mails: a.fender@unistra.fr and spfeffer@unistra.fr

(Received 23 March 2018, revised 12 April 2018, accepted 16 April 2018)

doi:10.1002/1873-3468.13067

Edited by Wilhelm Just

MicroRNAs (miRNAs) are evolutionarily conserved small regulatory RNAs that participate in the adjustment of many, if not all, fundamental biological processes. Molecular mechanisms involved in miRNA biogenesis and mode of action have been elucidated in the past two decades. Similar to many cellular pathways, miRNA processing and function can be globally or specifically regulated at several levels and by numerous proteins and RNAs. Given their role as fine-tuning molecules, it is essential for miRNA expression to be tightly regulated in order to maintain cellular homeostasis. Here, we review our current knowledge of the first step of their maturation occurring in the nucleus and how it can be specifically and dynamically modulated.

Keywords: biogenesis; Drosha; microRNA; regulation

Regulation of gene expression by small noncoding RNAs is at the heart of an ever-increasing number of biological pathways and can definitely not be overlooked by biologists whatever their field of research. There are different types of small regulatory RNA, which can be classified by their genomic origin or their function. They can also come in different flavors based on the kingdom, but for the sake of brevity, we will only mention here the different families that exist in animals. Despite some differences in their biogenesis, small RNAs share the same mode of action. Indeed, they act as sequence-specific guides for effector proteins, which belong to the Piwi/Argonaute (AGO) family [1]. Upon loading, they direct them toward their intended target RNAs. Broadly speaking, one can distinguish two main classes of small RNAs: (a) small interfering (si)RNAs, and micro (mi)RNAs, which are generated by the cleavage of varying size double-stranded (ds) RNA precursor molecules by type III ribonucleases, also called Dicer proteins [2] and (b) germline specific piwi-associated (pi)RNAs, which do not depend on the dicing of a dsRNA molecule (for a review see [3]). Although they share some common biogenesis factors, siRNAs and miRNAs are very different in terms of their biological role in the cell. The former can be seen as a defense system against foreign or unwanted double-stranded nucleic acids, whereas the later are constitutively expressed and play important roles as fine-tuners of gene expression. The focus of this review is on miRNAs, and we will therefore not dwell longer on si- and piRNAs.

The biogenesis of miRNAs, as described in Fig. 1, is a complex and compartmented process that begins with the transcription of a long primary transcript called pri-miRNA, which contains all the features of a coding mRNA. This transcription is mostly performed

Abbreviations
ADAR, adenosine deaminase acting on RNA; AGO, Argonaute; CED, central domain; ChIP, chromatin immunoprecipitation; CTT, C-term tail; dsRBD, double-stranded RNA-binding domain; KSRP, KH-type splicing regulatory protein; lncRNAs, long noncoding RNAs; MAPK, mitogen-activated protein kinase; miRNA, microRNA; PAZ, Piwi/Argonaute/Zwille; RBPs, RNA-binding proteins; Rhed, RNA-binding heme domain; RIIID, RNase III domain; RISC, RNA-induced silencing complex; snRNPs, small nuclear ribonucleoproteins; TDP-43, Tar DNA-binding protein 43.

by RNA polymerase II (RNA pol II) [4], but there are some cases of virus-encoded miRNAs that are transcribed by RNA polymerase III (RNA pol III) [5,6]. Upon recognition of a stem-loop structure within the pri-miRNA by the RNase III Drosha [7] and its cofactor DGCR8 [8], i.e., the Microprocessor complex, the ~ 65-nucleotide long precursor (pre) miRNA is cleaved and taken in charge by the Exportin 5 factor [9] to be translocated to the cytoplasm. There, the pre-miRNA undergoes a second cleavage event, which is mediated by the RNase III Dicer [10], with the help of its cofactor TRBP [11]. The resulting small RNA duplex is then assembled into one AGO protein, where it is unwound to keep only one of the two strands [12], which becomes the mature, 22-nucleotide long, miRNA. This process has been shown to require the help of chaperones such as Hsp90 [13]. In humans, there are four AGO proteins, which can indiscriminately accommodate miRNAs (for a review see [14]). The AGO protein loaded with a miRNA, also referred to as RNA-induced silencing complex (RISC) [15], scans the population of mRNA molecules within the cell until it finds a sequence match. The targeting process is complex and has been the subject of a tremendous amount of work by several groups, and we will not cover it into details here. Briefly, the recognition of the target by the RISC involves a handful (6–8) of nucleotides located 5′ of the miRNA, and coined the seed [16]. Because the requirement for miRNA–mRNA pairing that results in an efficient regulation by AGO is so limited, it is no wonder that the vast majority of the coding genome can be regulated by miRNAs. Indeed, there are currently almost 2000 miRNA genes reported for human alone [17], and the conservative estimates are that at least 60% of mRNAs are miRNA targets [18]. The mechanism by which the miRISC regulates its target mRNA requires a review on its own. Suffice to say that it involves the recruitment of an adaptor protein called GW182 or TNRC6 in human that in turn will interact with a number of other proteins ultimately leading to the inhibition of translation initiation and destabilization of the mRNA by deadenylation (for a review see [19]).

Although we described here the key steps involved in the canonical miRNA biogenesis, there are a number of alternate ways that have been reported in the literature by which these small RNAs can also be matured. We already referred to the involvement of RNA pol III in the transcription of pri-miRNA, which to date has only been reported in few viruses such as the murid herpesvirus 4 (MuHV4), which synthesizes a tRNA–pre-miRNA hybrid matured by tRNase Z [6], or the bovine leukemia virus [5]. The maturation step by Drosha is not mandatory to make a miRNA; there are a number of Drosha-independent ways to synthesize them. The most well-known are the mirtrons, which are generated by splicing of the pre-miRNA out of an mRNA [20,21]. Other miRNAs are generated in a Dicer-independent manner, although they are much rarer. In this case, the pre-miRNA is directly loaded into AGO2, which cleaves one arm of the hairpin, before the resulting RNA gets shortened by an exoribonuclease [22].

These alternate pathways for miRNA biogenesis highlight the various steps that can be diverted and that are therefore under tight control by the cell. Given the regulatory power of miRNAs, it is of prime importance to maintain their expression in check and to ensure quality control at each and every step along the way. We now know that regulation of the miRNA biogenesis does occur from the transcription of the pri-miRNA all the way down to the stability of the final mature miRNA product (for a general review on miRNA biogenesis regulation see [23]). Here, we will review the first step of miRNA maturation mediated by the Microprocessor complex in the nucleus. We will describe how it occurs before focusing on its regulation by various cofactors that help to control cellular homeostasis or stress response. We will more specifically detail the protein cofactors and their mode of action, but recent findings on alternative modes of pri-miRNA processing regulation will also be discussed.

Nuclear processing of primary miRNAs

The Microprocessor complex

Drosha functions as a core complex together with its essential cofactor DGCR8 (or Pasha in Drosophila and *Caenorhabditis elegans*), called the Microprocessor, but additional RNA-associated proteins, such as helicases or heterogeneous nuclear ribonucleoproteins (hnRNPs), are also found in close association with this complex [24–26]. At first, the role of these additional proteins was unclear as the processing activity was attributed to the sole Drosha/DGCR8 complex. Recently, it became apparent that they play important roles in regulating the activity of the Microprocessor, both globally and specifically, as it will be discussed later on. Drosha is a protein of 159 kDa, mainly nuclear, composed of proline-rich (P-rich) and arginine/serine-rich (R/S-rich) domains at the N-terminal (N-term) part, followed by a central domain (CED), a tandem of RNase III domains (RIIIDa and RIIIDb), and a double-stranded RNA-binding domain (dsRBD) at the C-terminal (C-term) part (Fig. 2A). The N-term

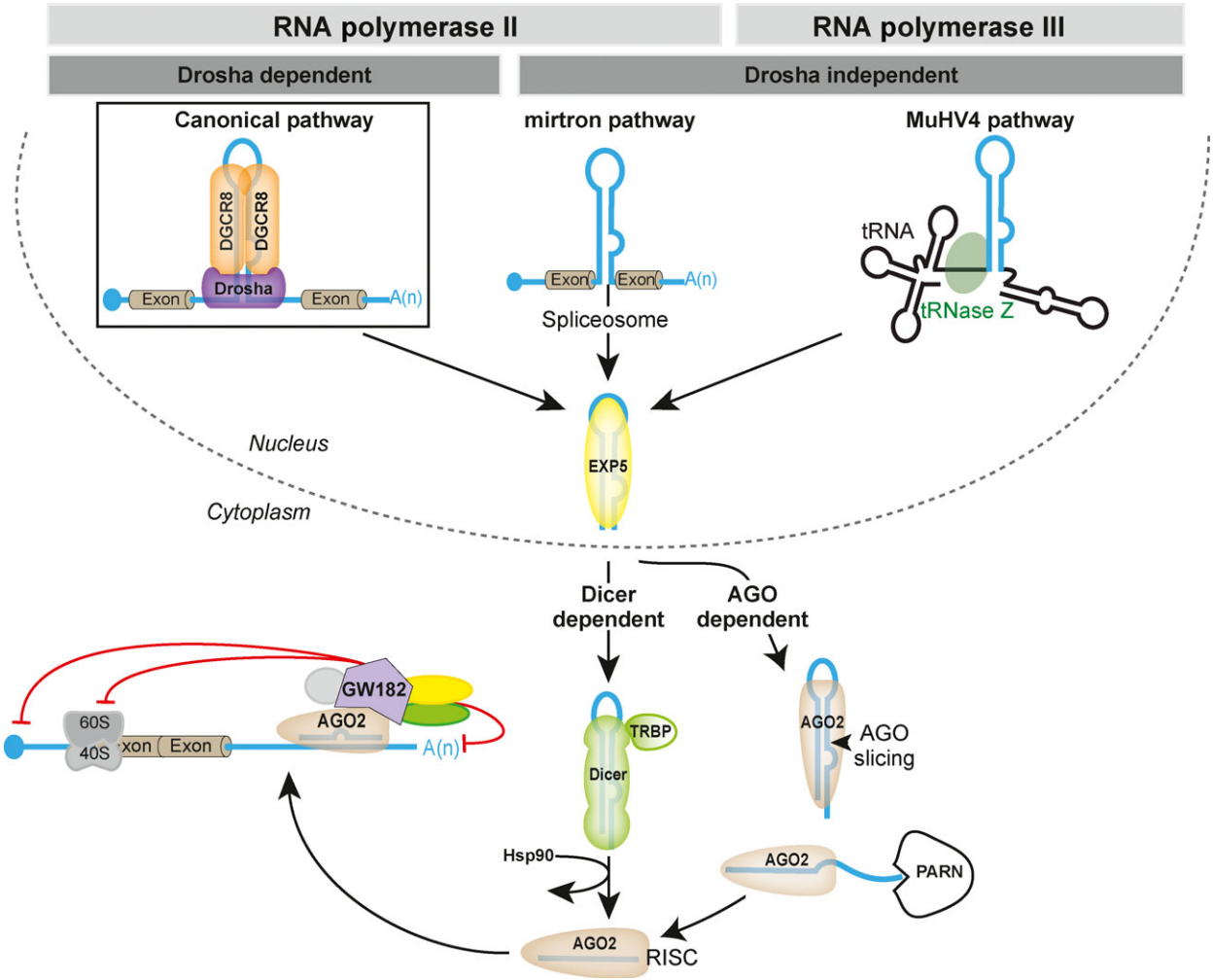


Fig. 1. Biogenesis and mode of action of miRNAs. Both canonical and alternate pathways are emphasized. MuHV4, Murid Herpesvirus 4. See main text for more details.

part is dispensable for pri-miRNA processing activity *in vitro*, however it harbors the nuclear localization signal and two phosphorylation sites necessary for targeting Drosha to the nucleus [27]. The CED is essential for Drosha processing activity and is composed of a platform and a Piwi/Argonaute/Zwille (PAZ)-like domain that may help the recognition of the pri-miRNA [28]. The two RIIIDs form an intramolecular dimer to create one processing center containing two catalytic sites. The RIIIDa cuts the 3′ strand of the pri-miRNA stem, whereas the RIIIDb cleaves the 5′ strand, resulting in a pre-miRNA with a typical RNase III signature of a two-nucleotide 3′ overhang [8]. Mutations of E1045 and E1222 residues at the catalytic core are sufficient to abolish the cleavage activity. Individual mutations impair only one of the two cuts [29].

DGCR8 stands for DiGeorge syndrome Chromosomal Region 8, since the monoallelic deletion of the associated genomic region is related to complicated clinical phenotypes including DiGeorge syndrome [30]. DGCR8 is a nuclear protein of 86 kDa composed of an N-term part containing a nuclear localization signal, followed by the central RNA-binding heme domain (Rhed), two dsRBDs, and the C-term tail region (CTT) (Fig. 2A). The Rhed domain contains a WW motif and is responsible for homodimerization [31] and for heme binding through two cysteine residues [32]. It is not clear whether heme is involved or not in DGCR8 dimerization [33,34]. However, different studies agree on the necessity of heme-bound DGCR8 for the full activity of Drosha in mammalian cells [31–35]. The CTT region is necessary and sufficient for Drosha binding [29].

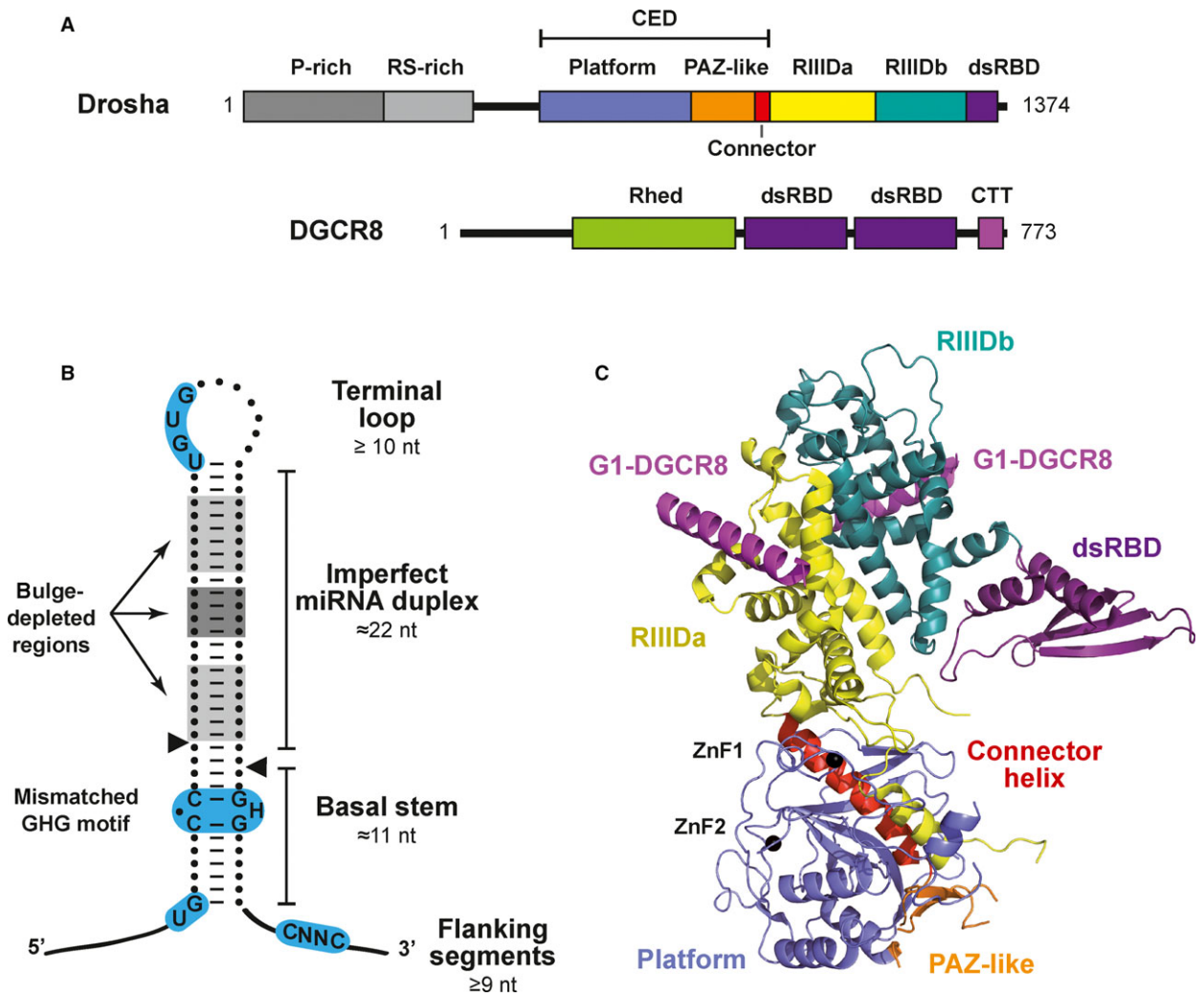


Fig. 2. (A) Domain composition of human Drosha and DGCR8. P-rich, proline-rich; RS-rich, arginine/serine-rich. (B) Schematic representation of a typical pri-miRNA structure, with the different domains emphasized (see main text for details). Sequence/structural motifs determined by the Bartel group [42,46] are depicted in blue, and bulge-depleted regions in gray (light gray according to Ref. [47] and dark gray according to Ref. [48]). H, all nucleotides except G. (C) Structure of Drosha bound to G1-DGCR8 according to [28]; PDB ID: 5B16, www.rcsb.org [127]. Picture was obtained using version 1.6 of PYMOL (Schrödinger, New York, NY, USA). Protein domains are colored according to (A). Zinc molecules are in black. The structure of the putative PAZ-like domain was only partially determined.

Molecular and structural features of pri-miRNAs

miRNA sequences are located within various genomic contexts. In human, they are found in intronic, exonic, and intergenic regions (both from coding or noncoding transcripts). These pri-miRNAs are capped and polyadenylated [36] and can reach several kilobases in length; they contain one or several clustered miRNAs [37]. Their processing by the Microprocessor is a key step and requires a high degree of precision because it defines the miRNA ends and even a single-nucleotide deviation can affect which mRNAs are targeted. In addition, global analysis of pri-miRNA processing *in vitro* and *in vivo* showed that the nuclear processing event predominantly explains the differential accumulation of miRNAs [38–40]. It is therefore essential to understand how the Microprocessor discriminates between authentic substrates and the rest of the countless stem-loop RNA structures in the cell.

Structural features are of prime importance in defining how efficiently a pri-miRNA is recognized and processed by the Microprocessor. Earlier studies have identified such criteria, including a stem structure of about three helical turns (≈ 33–35 nucleotide long) containing the miRNA duplex at its apical part, terminated by a flexible loop (≥ 10 nucleotide) and flanked

by single-stranded segments at its basal stem (≥ 9 nucleotide long) (Fig. 2B). The miRNA duplex (≈ 22 nucleotide) is imperfectly base-paired, containing G:U wobble pairs, single-nucleotide bulges, and mismatches, whereas the basal stem (≈ 11 nucleotide) is mainly perfectly base-paired, especially at its extremities forming stable platforms [41–45]. This basal stem together with flanking single-stranded regions forms a key junction for Microprocessor recognition [42–44]. Another key feature is the terminal loop: it should be flexible but not too large since pri-miRNAs with bigger loop (≥ 15 nucleotide) were shown to be less efficiently processed [41,45]. Additionally, a large loop may mimic, along with the miRNA duplex, the basal junction recognized by the Microprocessor and may lead to abortive cleavage from the wrong end of the hairpin [43].

More recently, these secondary features have been refined in two studies from the Bartel group [42,46]. They conducted unbiased approaches by creating libraries of variants of representative human pri-miRNAs either differing in the nucleotides composition of the apical loop and the single-stranded flanking segments or containing randomized three base pair (bp) blocks of nucleotides at different positions along the stem region. Notably, the authors identified for the first time that primary sequences are involved in pri-miRNA recognition. These are a UG motif at position −14 and −13 from the 5′ cleavage site, respectively; a GUG/UGU motif in the apical loop; and a CNNC motif (downstream of the basal junction) recognized by SRp20 accessory protein. In their second study, they further identified a 'mismatched GHG' on the 3′ strand 7–9 nucleotides from the basal junction (in which H is any nucleotide but G), a preference for pairing throughout the remainder of the stem, and a narrow stem-length preference of 35 ± 1 bp, as key structural elements. Finally, by combining different lengths and number of unstructured features (mismatches and bulges in the stem), they found that primary sequence motifs had an additive effect and that they are essential for efficient processing of suboptimal structures. Interestingly, primary sequence motifs seem to be species specific since their absence in *C. elegans* pri-miRNAs leads to inefficient processing by the human Microprocessor. However, only a subset of human pri-miRNAs contains them (about 79% of the conserved human miRNAs contain at least one of the motif), which strongly suggests that individual pri-miRNAs are composed of different modules/determinants and that the combination of these defines the specificity and efficiency of Drosha processing. Along the same line, Roden *et al.* [47] developed a computational approach that systematically evaluated miRNA hairpins in comparison to non-miRNA transcripts from Refseq sequences. The authors found that stem-length is a key distinguishing factor and that optimal pri-miRNA stems contain bulge-enriched and -depleted subdomains. They also showed that the distance between the CNNC motif and the basal junction of the hairpin is important. Finally, Sperber *et al.* [48] showed that mismatches in the central region of pri-miRNA stems (9–12 nucleotides from the cleavage site) are antideterminants of processing when Drosha levels is limiting in the cells. Furthermore, they showed that Drosha levels vary between tissues and throughout cellular development, and that miRNAs without mismatches in the 9–12 nucleotides region are over-represented in cells with low levels of Drosha.

Altogether these studies have defined critical sequence/secondary features important in specifying authentic pri-miRNAs, in a context of unique hairpin structure. However, none was based on actual structural studies and the involvement of additional *cis* elements contained in the rather large pri-miRNAs (several kilobases) was not tested. Genome-wide interrogation of RNA structures is under fast development and its use may greatly benefit this specific research field. To date, secondary structures were determined only for two pri-miRNAs, using in solution probing approach and *in vitro* transcripts. These are the human pri-miR-17-92 (containing six miRNAs) [49,50] and the virus-encoded pri-miR-K10/12 from Kaposi's sarcoma-associated herpesvirus (containing 10 miRNAs, [51]). Chakraborty and Krishnan [50] showed that most hairpin structures of pri-miR-17-92 were neither optimally folded nor accessible to solvent, suggesting the requirement of transacting factors to remodel the pri-miRNA tertiary organization. Using single molecule electron microscopy, Chaulk *et al.* [49] showed that pri-miR-17-92 adopted a compact globular tertiary structure where miRNAs internalized within the core structure are processed less efficiently. These constraints happen to be released by endonuclease activity, liberating an intermediate product termed 'progenitor-miRNA' [52]. Our group showed that the secondary structure of the viral pri-miR-K10/12 correlated with miRNA abundance in infected cells. We also showed that swapping hairpin structures containing low and highly expressed miRNAs inverted miRNA expression levels. This suggests that sequence and/or structural contexts are important for miRNA expression. Along the same line, it was showed that deletion of miR-11 within the miR-11-998 cluster abolished miR-998 expression [53]; and that Epstein–Barr herpesvirus miR-BHRF1-3 expression decreased in B

cells infected with a virus that lacks pre-miR-BHRF1-2 suggesting a sequential processing of the two miRNAs from the same transcript [54].

Mode of action of the Microprocessor and recognition of pri-miRNAs

A long-standing debate was to determine how the Microprocessor recognizes pri-miRNAs and orients itself asymmetrically on the stem-loop structure to select the right cleavage site. During the last decades, many research groups have put effort to determine the critical features within the Microprocessor ensuring specific recognition of pri-miRNAs. At least three different models have been proposed until now.

In an early model, the Microprocessor was proposed to bind a large (> 10 nucleotides) terminal loop to position Drosha's catalytic center two helical turns (about 22 nucleotides) from the stem-loop junction, resulting in the liberation of the pre-miRNA [41]. Another model suggested that the terminal loop is dispensable for Microprocessor recognition. Instead DGCR8 recognizes the hairpin's basal single-stranded and dsRNA junction and recruits Drosha for cleavage [43]. Two more recent studies agree to show that neither model alone is sufficient to account for precise cleavage position of pri-miRNAs in cells [55,56]. Instead, both the basal and the apical junctions cooperatively coordinate cleavage position and processing efficiency. Indeed, changing the size by even 1 or 2 bp on either side of the stem-loop causes alternative processing *in vitro* and influences its efficiency in cells. Additionally, Ma *et al.* [55] showed that naturally occurring alternative Drosha processing of pri-miR-142 can be reversed by changing the distances of the lower and upper stem–ssRNA junctions from the cleavage site.

Discrepancies among these three models most probably come from the fact that different pri-miRNA backbones were used for mutational and functional analysis, and that these have an impact *per se* on how the Microprocessor recognizes the different RNA determinants. Thus, altering the length of the upper stem of pri-miR-16 has less impact than altering the lower stem and leads to less alternative cleavage than for pri-miR-150 or pri-miR-122 used in Ma *et al.* study. Finally, primary sequence motifs also play an additional and important role [42]. Altogether this suggests that the way the Microprocessor recognizes pri-miRNAs depends upon their individual sequence/structural features, which also explains how isoforms of miRNAs occur naturally due to alternative or inaccurate cleavage [57].

Another topic of debate was to determine whether either Drosha or DGCR8 was responsible for RNA recognition. For a long time, it was firmly believed that DGCR8 was essential for substrate recognition, followed by the recruitment of Drosha for the catalytic step [43]. Biochemical studies on DGCR8 showed that it is indeed able to directly and stably bind to pri-miRNA through its tandem dsRBDs, whereas Drosha dsRBD alone showed weak or no affinity for RNA [58–60]. However, this model was unlikely to be fully correct since DGCR8 can bind single-stranded, double-stranded, and random hairpin transcripts *in vitro* with similar affinity [61]. In addition, it was shown that DGCR8 alone multimerizes and covers the entire pri-miRNA hairpin structure [62], which cannot lead to a specific recruitment of Drosha on the correct cleavage site. These discrepancies were also due to a technical limitation, as it seems that full Microprocessor is required to achieve proper recognition and processing of RNA substrates and there was no good approach to make Drosha recombinant protein. Recently, this issue was solved by the Woo laboratory who managed to produce it in HEK293 cells [29]. It turned out that coexpressing a fragment of DGCR8 composed of only the final 23 amino acids of the CTT region (called G1-DGCR8) with Drosha was sufficient to get a soluble and active complex. This also indicated that Drosha alone is able to determine the cleavage site. However, cooperation of Drosha and the other domains of DGCR8 improve the efficiency (DGCR8 dsRBDs) and accuracy (DGCR8 Rhed) of processing. In particular, interaction of DGCR8 at the apical loop may prevent abortive cleavages by orienting Drosha toward the basal stem when pri-miRNAs contains fewer determinants. In this study, it was also shown that the Microprocessor operates as a heterotrimer formed before RNA recognition by the assembly of one molecule of Drosha and a dimer of DGCR8. This was later confirmed by the Steitz group [63].

Recently, the crystal structure of a catalytically inactive human Drosha, N-term truncated, and in complex with the CTT (G1) domain of DGCR8 was solved at a 3.2 Å resolution [28]. Drosha is an elongated protein composed of two main structural modules: the catalytic center on the top formed by the two RIIIDs connected by a long α-helix (Connector helix) to the CED which arranges as a platform containing a non-canonical Zn-finger motif and a PAZ-like domain. RIIIDa interacts with CED, whereas RIIIDb interacts with dsRBD. Two G1-DGCR8 helices bind asymmetrically on each RIIID of Drosha (Fig. 2C).

The domain arrangement and overall folding of Drosha, which belongs to the animal-specific class II of the RNase III family is very similar to those of Giardia Dicer [64], which is part of the class III RNase

III proteins (found in the three kingdoms). This suggests that both proteins may have evolved from an early class III metazoan ancestor. Modeling of RNA substrate recognition in this crystal structure revealed that Drosha covers the basal part of the pri-miRNA, while the two G1-DGCR8 are oriented toward the apical side. Notably, various modules of Drosha (Bump and MB helices from the nonconserved insertion within RIIIDa, Platform, and PAZ-like domain) may work together to recognize the basal junction and single-stranded RNA flanking segments and maintain their positions for the measurement of 11 bp toward the cleavage site. The dsRBD does not seem to contact RNA, but it may move upon interaction. The RNA-binding ability of this domain is controversial and it has been suggested that it has rather evolved another function that supports pri-miRNA processing [58–60].

From this 3D view of the Microprocessor, Drosha appears to be both the 'scissor' and the 'molecular ruler'. However, we will have a complete understanding of pri-miRNA recognition only when the full Microprocessor, bound to RNA, will be structurally characterized. Indeed, there is only fragmented information about the mode of recognition by DGCR8 so far [65]. It was shown that the tandem dsRBDs fold into a butterfly shape and may contact both ends of the RNA stem, forcing the middle part of the helix to kink. However, this is not consistent with a consecutive interaction with the upper part of the stem-loop structure when associated to Drosha. In fact, Kwon *et al.* [28] suggested that the two DGCR8 monomers may be located parallel to each other, so that their dsRBDs and Rhed interact with the upper stem and apical loop of pri-miRNAs. Recently, Partin *et al.* [34] showed that heme induces a conformational change in DGCR8 dimer, which in turn activates DGCR8 to recognize the terminal loop of pri-miRNAs *in vitro*. They suggest that heme-dependency varies among pri-miRNA substrates and seems to be essential for miRNAs in which the basal junction cannot outcompete the apical junction for the recruitment of Drosha. For these particular pri-miRNAs, in the presence of a mutant DGCR8 defective in heme binding, abortive cleavage products were more observed *in vitro*, reflecting the prevalent binding of Drosha at the apical junction. Altogether, heme and dimerization of DGCR8 improve processing accuracy of site cleavage selection, and are necessary for full Drosha activity both *in vitro* and in cells [31–35].

Global analysis of pri-miRNA processing

With the goal to globally define the identity of pri-miRNA transcripts and their pre-miRNA output, several laboratories have sequenced RNAs either from cells in which Drosha was silenced or expressing a catalytically inactive form of Drosha [66] or after coimmunoprecipitation of Drosha and/or DGCR8 [57,67,68]. Thus, Kim *et al.* found that many pri-miRNAs undergo alternative processing, yielding multiple miRNA isoforms. They also found that cleavage sites do not necessarily correspond to the ends of mature miRNAs, suggesting widespread end modifications at subsequent miRNA maturation steps [57]. Other studies also led to the discovery of novel RNAs interacting with the Microprocessor components (e.g., hairpins embedded in open reading frame, or RNAs deriving from retrotransposons), emphasizing alternative roles for Drosha and DGCR8 (see Refs [69,70] for example reviews).

Three other analyses focused on the kinetic aspect of pri-miRNA processing. Feng *et al.* [39] examined the *in vitro* processing of a total of 247 human *in vitro* transcribed pri-miRNAs. They showed that processing efficiency varies among the different pri-miRNA substrates and correlates with endogenous miRNA expression and that conserved pri-miRNAs are better substrates of the Microprocessor. Recently, the Orom laboratory developed a transcriptome-wide approach using next-generation sequencing to determine *in vivo* kinetics of pri-miRNA processing [38,40]. In a first study, they sequenced pri-miRNAs purified by chromatin immunoprecipitation. Later on, they used a pulse-chase approach to obtain nascent RNAs at different time points after labeling with bromouridine (BrdU). In that case, RNAs were purified by BrdU-immunoprecipitation. The main conclusions were that pri-miRNAs exhibit different processing kinetics ranging from fast to slow processing and that pri-miRNA processing may occur both cotranscriptionally and post-transcriptionally. A similar study was performed to measure half-lives of both pri- and mature miRNAs. Although the focus was more on the stability of the end product, they also reported that features such as genomic organization or transcription rates could influence the steady state level of the mature miRNA [71]. Altogether, global kinetics studies suggest that pri-miRNA maturation by the Microprocessor may act as a key regulatory step in miRNA biogenesis.

Regulation of pri-miRNA processing by accessory proteins

Now that we have seen in detail how pri-miRNAs are matured into pre-miRNAs by the Microprocessor, we will describe how this process can be regulated. Indeed, miRNA accumulation needs to be tightly

regulated in time and space in order to exert their regulatory function adequately [72,73]. As a result, global or specific deregulation of miRNA expression is often linked to numerous diseases [74]. At the pri-miRNA level, this regulation can be achieved by a large number of accessory proteins that can act on specific subset of miRNAs by binding to the apical loop or the flanking single-stranded sequences [75,76]. Alternatively, they can affect a broad range of miRNA precursors by directly modulating the Microprocessor function [77–79]. Of note, different studies also showed that these cofactors are at the crossroad of cellular signaling pathways, pre-mRNA maturation and miRNA biogenesis [80,81].

Methods to identify accessory proteins

Depending on whether one wants to identify a global regulator of pri-miRNA biogenesis or a factor involved in the maturation of one specific pri-miRNA transcript, one has to turn either to a protein-centric (i.e., Drosha or DGCR8) [24] or an RNA-centric (the pri-miRNA of interest) approach [82], coupled to mass spectrometry analysis. Both methods rely on protein–protein or protein–RNA complex formation within the cell or in solution, followed by affinity purification. One drawback is that RNA–protein or protein–protein complexes can be more or less specific, as well as stable. As a consequence, isolation of RNA–protein or protein–protein complexes in solution gives higher rate of false-positive interactions compared to *in vivo* complex formation, more particularly when the latter is coupled with cross-linking. This also means that once the list of proteins has been obtained, the validation phase should not be overlooked and has to be performed thoroughly. Eventually, a transcriptomic analysis following the inactivation of the cofactor gene can be performed to highlight a wider effect. To go further and globally identify the RNA partners and binding site sequences, cross-linking coupled to immunoprecipitation (CLIP) approaches are now routinely used [83]. To measure the impact of the identified factor on miRNA expression, classical techniques such as northern blot analysis, RT-qPCR or microarray have been used, as well as luciferase reporter or *in vitro* processing assays.

Modes of action of accessory proteins

Within the two main modes of action on either the Microprocessor or the pri-miRNA, several distinct mechanisms can be distinguished to more precisely explain how an accessory protein can impact on miRNA processing. Namely, for proteins directly binding the Microprocessor, we can distinguish between Microprocessor post-translational modification [84] and direct binding competition between Drosha and DGCR8 [85]. Proteins interacting with the pri-miRNA can either act by regulating its binding by the Microprocessor either negatively [78] or positively [86], by remodeling the transcript structure [72], or by inducing pri-miRNA post-transcriptional modifications [87]. Finally, the regulation of processing can also occur cotranscriptionally and will be influenced by splicing [88].

Transcription, splicing, and miRNA processing

The identification of a number of pre-mRNA maturation RNA-binding proteins (RBPs) as accessory proteins in the miRNA biogenesis, as well as the intronic localization of a majority of miRNA precursors, strongly suggests that pri-miRNAs are processed cotranscriptionally (Fig. 3A). Chromatin immunoprecipitation (ChIP) using Drosha-specific antibody revealed an association with miRNA genes. The Microprocessor seems to have some affinity for specific histone proteins. Thus, histone H1-like chromatin protein (HP1BP3) promotes global miRNA biogenesis by recruiting or retaining Drosha at active miRNA loci [89]. The epigenetic mark H3K4me3 together with superenhancer DNA sequence has been demonstrated to help the recruitment of the Microprocessor on nascent transcripts strongly supporting a cotranscriptional pri-miRNA processing [90]. Recently, a direct interaction between RNA pol II and Drosha has been observed in Drosophila cells. The phosphorylation of the RNA polymerase C-terminal domain by Cdk9 seems to increase the processing of pri-miRNA lacking the apical loop UGU motif [91]. Another study, based on ChIP assays followed by RNA sequencing, allowed to measure pri-miRNA half-lives [38]. Detection of short-lived pri-miRNAs suggests they are cotranscriptionally processed at the chromatin interface. However, not all pri-miRNAs disappear at such rate, and some seem to be much more stable indicating that in this case processing requires an active regulation or takes place post-transcriptionally [40].

Intron cotranscriptional cleavage assay coupled with hybrid selection, circularization, and rapid amplification of cDNA ends allowed to establish that Drosha-mediated cleavage does not interfere with splicing and occurs on nascent transcripts before splicing event [88]. More recently, another study showed that inhibition of splicing favors pri-miRNA processing, and vice versa [92]. Thus, it appears that the spliceosome complex

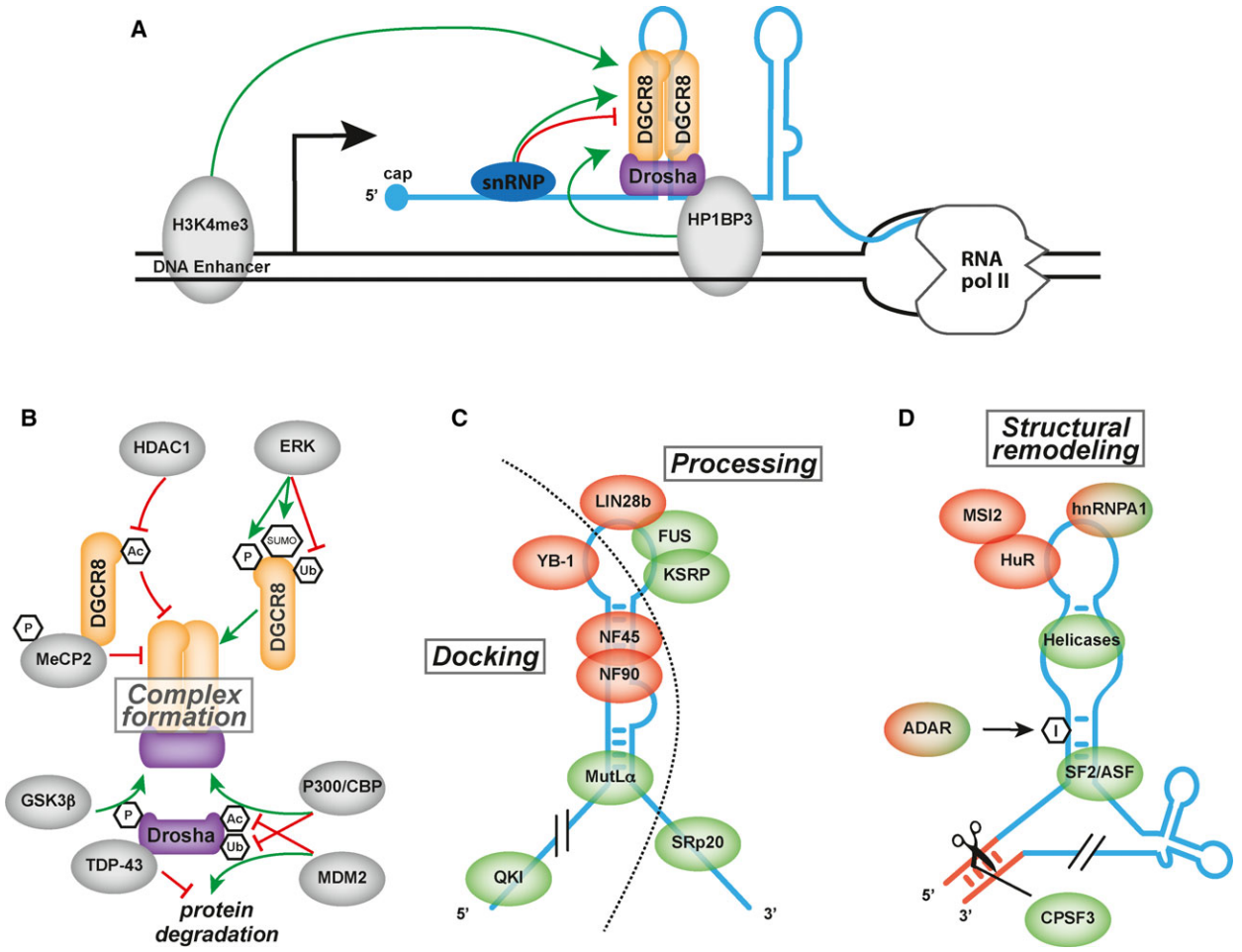


Fig. 3. (A) Schematic representation of cotranscriptional pri-miRNA processing. Positive (green arrows) and negative (red arrow) regulatory networks are displayed. The Microprocessor can be modulated by proteins like (a) snRNPs (in dark blue) and histone HP1BP3 (gray oval), and (b) DNA enhancer element in concert with epigenetic marks (H3K4me3, gray oval). (B) Direct action of accessory proteins on the Microprocessor. Post-translational modifications are in hexagons (P, phosphate group; Ac, acetyl group; SUMO, sumoyl group; Ub, ubiquitin). Green and red arrows stand for positive or negative action. (C) Accessory proteins interacting with the pri-miRNA to regulate docking (left side of the dotted line) or slicing activity (right side) of the Microprocessor. Accessory proteins can have a positive effect (green ovals) or a negative effect (red ovals). (D) Accessory proteins interacting with the pri-miRNA to induce structural remodeling. Cofactors can have positive (green ovals) or negative impact (red ovals) or both depending on the regulated pri-miRNA (red and green ovals). I, inosine (in hexagon). See main text for details.

may be involved in recruiting Drosha on nascent transcripts. Inactivation of the 5′ splicing site within the pri-miRNA-21 induces a decrease of its processing, suggesting the implication of small nuclear ribonucleoproteins (snRNPs) complex to directly or indirectly influence recruitment of the Microprocessor [93].

Direct action on the Microprocessor

Post-translational modifications are known to modulate subcellular localization, protein half-life or binding capacity. Figure 3B summarizes the various modifications that affect the Microprocessor components and thus impacts on miRNA processing. Acetylation can either enhance or repress protein decay. In the case of Drosha, several acetyltransferases (P300, CBP, and GCN5) have been described to acetylate various lysine residues within the N-terminal domain of Drosha, thereby repressing its ubiquitin-mediated proteasome decay [94]. Removal of acetyl groups has also been observed for the Microprocessor, as deacetylation of the dsRBD of DGCR8 by histone deacetylase 1 has been described to enhance its affinity for a subset of pri-miRNAs resulting in an increased processing [95]. Other modifications, such as protein phosphorylation, have also been linked to an increase in Microprocessor

complex formation and activity. For example, Drosha is phosphorylated by the protein kinase Glycogen synthase kinase 3β (GSK3β), implicated also in a large number of signaling pathways involving proteins such as Hedgehog, Notch, and WNT/β-catenin. Interestingly, GSK3β is only able to act in an RNA-dependent manner, since it cannot directly bind Drosha or DGCR8 [84]. Moreover, at least 23 phosphorylated amino acids have been described on DGCR8. Modification of some of these results from the extracellular signal-regulated kinases (ERK1 and ERK2) activity, linked to the mitogen-activated protein kinase (MAPK) signaling pathway. DGCR8 phosphorylation appears to increase the stability of the protein without affecting its subcellular localization or ability to interact with Drosha [96]. Interestingly, ERK kinase activity is also linked to SUMOylation of the K707 residue of DGCR8 by the SUMO1 protein; this modification prevents ubiquitination of the protein and therefore its decay *via* the proteasome [97]. Tar DNA-binding protein 43 (TDP-43) is a hnRNP involved in processing of many RNAs, and identified within the large Microprocessor complex [25]. At least during neuronal differentiation, TDP-43 is able to stabilize Drosha by inhibiting its proteasome-dependent decay [98]. On the contrary, ubiquitination of a lysine within Drosha N-terminal domain by MDM2 has been demonstrated to induce Drosha degradation through mTOR signaling pathway [99]. Competition between lysine modification by stabilizing acetylation and decay-inducing ubiquitination within Drosha N-terminal domain thereby regulates its concentration and consequently modulates global miRNA level. Finally, in cells deficient for Methyl-CpG-binding protein 2 (MeCP2), a transcriptional repressor involved in alternative RNA splicing, miRNA expression is globally enhanced. MeCP2 phosphorylated at Ser80 has been described as a partner of DGCR8 that competes for its binding with Drosha, therefore repressing pri-miRNA maturation [85].

Accessory proteins interacting with the pri-miRNA

Pri-miRNA-binding proteins can interact with the single-stranded regions flanking the pre-miRNA, the stem of the pre-miRNA, or its apical loop (Fig. 3C). Interestingly, 14% of miRNA loop sequences are conserved. Thus, they have been considered as proteins recruitment platform to specifically modulate miRNA biogenesis both positively and negatively [100]. Nuclear factors 45 and 90 (NF45, NF90), interleukin-2-dependent transcription factors also involved in splicing, have been among the first described accessory proteins [24,25]. An *in vitro* study demonstrated that NF90–NF45 complex is able to bind a large subset of pri-miRNA with a stronger affinity than DGCR8. Thus, this complex is able to prevent pri-miRNA processing by competing with the Microprocessor for pri-miRNA binding [78]. A similar mechanism has been described for the Y box-binding protein (YB-1), which is able to bind the consensus sequence UYAUC of both pri- and pre-miRNA of miR-24-1 and miR-29b to inhibit Drosha binding [101]. Similarly, LIN28b competes with the Microprocessor for the binding of let-7 primary transcript in *C. elegans* [102]. On the contrary, TDP-43, which we talked about before, acts by specifically recognizing the terminal loop of pri-miR-132, pri-miR-143, pri-miR-558, and pri-miR-574 to facilitate recruitment of the Microprocessor [98].

Interestingly, several of those pri-miRNA-binding proteins can display feedback loop regulation with their targeted miRNA. Thus, fused in sarcoma protein (FUS) has been described to facilitate Drosha recruitment during transcription of a subset of pri-miRNA genes including the precursor of miR-200, which is known to be involved in the regulation of FUS [86]. Similarly, the KH-type splicing regulatory protein (KSRP) is able to bind single-stranded AU-rich element within the apical loop of some pri-miRNAs, including Let-7, to increase pri-miRNA processing. Let-7 in turn is able to regulate the expression of KSRP in a feedback loop [76]. Finally, the mismatch repair mutant protein (MutLα), which has been described to compensate for the absence of a basal segment of several pri-miRNA, including pri-miR-422a, is also implicated in a regulatory feedback loop with this miRNA [103].

As we have seen in the first part, structural features are fundamental for the good recognition of the pri-miRNA by the Microprocessor. In cases where these features are not present or not optimal, the shape of the primary transcript can be altered in a positive or negative manner by helicases, RBPs, and post-transcriptional modifications (Fig. 3D). Silencing of DDX5 and DDX17 resulted in a lower expression of a large number of mature miRNAs. Moreover, their helicase ATPase activity is necessary to promote pri-miRNA cleavage. Studies proposed that these helicases improve structural conformation to increase pri-miRNA recognition and cleavage by Drosha/DGCR8 [104,105]. Other ATP-dependent RNA helicases such as DDX1 [106], DDX23 [107], DDX3X [79] can act with a similar mechanism on different pri-miRNAs. Tissue-specific expression of miR-7 has been described to be regulated by Musashi homolog 2 (MSI2) through the binding of Hu antigen R (HuR) within the AU-rich element sequence found in pri-miR-7 apical loop. This

binding results in a structural rigidity that inhibits Drosha catalytic activity [72]. At the same time, the splicing factor SF2/ASF has been described to have the exact opposite effect of HuR/MSI2 on pri-miR-7, since its binding in the basal stem induces potential structural rearrangement to favor Drosha cleavage [108]. Similar to the SF2/ASF protein, the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA-1) reshapes pri-miR-18a after binding the consensus sequence UAGGGA/U within the apical loop. This also results in an increase of Drosha slicing activity [100].

Structural rearrangements can also be the result of an intermediate processing of the pri-miRNA. This concept is fundamental for clustered miRNAs and suggests a step-wise maturation of the pri-miRNA. For instance, sequences located 5′ and 3′ of the pri-miR-17-92 have been described as being complementary and could interact to induce a specific folding of the pri-miRNA, in a conformation preventing its maturation by the Microprocessor. This repressed state can be opened by the cleavage and polyadenylation specificity factor subunit 3 (CPSF3) therefore allowing the Microprocessor recruitment [52] and highlighting the crosstalk between pre-mRNA maturation and miRNA biogenesis. Interestingly, the staphylococcal nuclease domain-containing protein Tudor-SN (SND1) has also been described to participate to the specific maturation of miR-92 within its cluster [109].

Other regulators

Beside accessory proteins, other types of regulators can be involved in the regulation of pri-miRNA processing. The number of reported examples is still limited, but it will without a doubt increase in the next few years. These regulators are mainly constituted of RNA molecules as we will see below, but chemical modification of the primary transcript has also recently emerged as an important factor in this field.

Noncoding RNAs

miRNAs themselves were shown to directly bind pri-miRNAs and regulate miRNA biogenesis. In *C. elegans*, the Let-7 miRNA can be translocated to the nucleus to promote the processing of its own pri-miRNA transcript by guiding the AGO protein ALG-1 to a complementary site located about 0.5 kb downstream of the miRNA hairpin structure. AGO also associates with a subset of pri-let-7 RNAs in human cells [110]. Similarly, the mouse miRNA miR-709, which is predominantly found in the nucleus, binds to a 19-nucleotide sequence located 0.8 kb downstream of the pri-miR-15a/16-1. In this case, the consequence is that subsequent pri-miRNA maturation events are inhibited as well as processing of the pri-miRNA hosting Dleu2 mRNA [111]. Recently, human miR-122 was shown to bind a 19-nucleotide region just upstream of the pri-miR-21 stem-loop structure [112], comprising the UG motif normally found in optimal pri-miRNA sequences [42]. This interaction is conserved in mouse, and directly interferes with the Microprocessor activity. However, the actual mechanism of action is not yet fully understood for any of these examples.

Another class of RNAs that plays a role in modulating the Microprocessor activity is the one constituted by long noncoding RNAs (lncRNAs). For example, the lncRNA Uc.283+3, transcribed from an ultraconserved region, binds to the lower stem of pri-miR-195 and impairs its processing by the Microprocessor [113]. Intriguingly, the same miRNA is also regulated by another lncRNA, Uc.372, which binds to the apical region of both pri-miR-195 and pri-miR-4668 and suppresses their maturation [114]. Finally, Jiang *et al.* discovered a totally new layer of pri-miRNA regulation, i.e., spatial regulation. They showed that the lncRNA NEAT1 may orchestrate a 'microprocessing factory'. Indeed, NEAT1 contains a number of pre-miRNA-like elements in its sequence that allows it to recruit the Microprocessor complex, as well as the global pri-miRNA processing enhancer NONO-PSF heterodimer and various RBPs, in specific nuclear subdomains called paraspeckles. This subcellular localization is directly linked to a more efficient processing of a number of pri-miRNAs [115].

Chemical modifications

Adenosine deaminase acting on RNA (ADAR) can modify pri-miRNAs to modulate, depending on the localization of the modification, their processing by Drosha, the efficiency of RISC loading or the repertoire of mRNA targets. A study demonstrated that editing the +4 and +5 nucleotides downstream the 5′ cleavage site within the pri-miR-142 destabilized the stem-loop structure thus inhibiting the binding of the Microprocessor [116]. However, ADAR activity on clustered miRNA can display a differential expression of mature miRNA, adding a level of complexity to accurately decipher the involvement of ADAR in miRNA biogenesis [117]. In addition, it should be noted that editing-independent effects of ADAR1 and 2 on miRNA biogenesis have been reported, which indicates that these proteins can also bind to pri-

miRNAs in a noncatalytic manner to regulate their processing [118].

Another type of RNA modification that has recently attracted a lot of interest in the RNA biology field is the N^6-methyl-adenosine, which has been shown to occur on mRNAs to modify their expression (see [119] for review). N^6-methyl-adenosine has been identified also within pri-miRNAs, in the vicinity of the miRNA hairpin structure, and was shown to play an important role in specifying authentic pri-miRNA substrates for the Microprocessor. METTL3 methylates a subset of pri-miRNAs, at specific RGAC motifs. These are then landing marks for hnRNPA2B1 that recruits the Microprocessor and promotes its activity [87,120].

Conclusion and perspectives

One important question that could be asked to conclude this review is: why so many layers of regulation instead of a single straightforward regulation of the miRNA gene at the transcriptional level? Indeed, it could be argued that given the efficiency of transcriptional control, an 'on and off' switch should be sufficient to limit the availability of a given miRNA in specific conditions. However, this does not take into account the fact that the downstream intermediates can take some time to accumulate or on the contrary are too stable to be eliminated fast enough. It is therefore of prime importance to be able to control each step of the miRNA biogenesis process to increase the possibilities to modulate and achieve the right concentration of a mature miRNA at the right time and in the right place.

There are many examples of temporal expression of individual miRNAs, the most well described being Let-7, which accumulation is exquisitely controlled during the different developmental stages of *C. elegans* [121]. This regulation involves among others, LIN28, which expression is repressed during the third larval stage of the nematode morphogenesis to allow the expression of Let-7 [73]. But the processing of miRNAs also needs to be controlled in space, which can clearly be illustrated by the numerous tissue-specific miRNAs that accumulate very differently throughout the organism [122]. miR-7 is a good example of a spatially regulated miRNA. Its pri-miRNA is located within intron 15 of hnRNP-K pre-mRNA, which is ubiquitously expressed. However, its mature form is mainly found in neuronal and pancreatic tissues. Studies rule out the sole control of transcription of hnRNP-K and as mentioned earlier, HuR and MSI2 are responsible for the repression of pre-miR-7 maturation [72]. The miR-124 miRNA offers another example of both space- and time-dependent regulation. It has been described to be post-transcriptionally regulated during erythroid differentiation by the RBP Quaking 5 (QKI5). QKI5 act as a positive factor for pri-miR-124-1 maturation through recognition of the QKI response element ~ 300 nucleotide upstream of the pre-miRNA stem-loop. RNA–RNA interaction within the pri-miRNA is suspected to bring QKI close to the stem-loop of pre-miR-124-1 to efficiently recruit the Microprocessor. During erythroid development, the expression of QKI5 is inhibited, thus decreasing recruitment of the Microprocessor on the pri-miR-124-1 for its maturation [123].

Modulation of the Microprocessor activity will also be impacted by various signaling pathways involved in normal development, cellular maintenance and homeostasis. For example, important components of the Microprocessor undergo post-translational modifications that alter their subcellular localization as well as their partner binding efficiency thus inducing modulation of miRNA expression. Such phenomena can be dependent on p38 MAPK, estrogen receptor, and TGFβ signaling pathways [80,104,124].

In addition to the regulation of pri-miRNA processing, all the other levels of the miRNA biogenesis are under control as well and, although we have made tremendous progress in that field recently, we have not yet elucidated the multiple facets of this important question. We now have at our disposal powerful techniques that allow us to identify at unprecedented scale all possible RBPs and to study their involvement in miRNA biogenesis. At the same time, it is equally important to fully decipher the importance of other regulators such as lncRNAs since this aspect remains largely underexplored. The next task will be to integrate all these data in order to characterize the regulatory networks at play, and then to assess their importance in pathological conditions. In the long run, this might pave the way to the development of new therapeutic strategies.

Finally, we should not overlook the possibility that miRNA primary transcripts might contain important elements besides the mature miRNA, which could be equally important. The recent discovery of regulatory peptides encoded by pri-miRNA transcripts that play key roles in plant development [125], or of longer (80-100 nucleotides) RNAs that can associate with AGO [126], indicates that there are still many discoveries to be made in the field of regulatory RNAs.

Acknowledgements

The authors thank Diane Bortolamiol-Becet and Erika Girardi for their critical comments. Our work was

funded by the European Research Council (ERC-CoG-647455 RegulRNA) and was performed under the framework of the LABEX: ANR-10-LABX-0036_NETRNA, which benefits from a funding from the state managed by the French National Research Agency as part of the Investments for the future program. AC was funded by a doctoral fellowship from the French Ministère de la Recherche et de l'Enseignement Supérieur.

References

1 Meister G (2013) Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**, 447–459.
2 Carmell MA and Hannon GJ (2004) RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol* **11**, 214–218.
3 Le Thomas A, Tóth KF and Aravin AA (2014) To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biol* **15**, 204.
4 Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH and Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051–4060.
5 Kincaid RP, Burke JM and Sullivan CS (2012) RNA virus microRNA that mimics a B-cell oncomiR. *Proc Natl Acad Sci USA* **109**, 3077–3082.
6 Bogerd HP, Karnowski HW, Cai X, Shin J, Pohlers M and Cullen BR (2010) A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral microRNAs. *Mol Cell* **37**, 135–142.
7 Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419.
8 Han J, Lee Y, Yeom K-H, Kim Y-K, Jin H and Kim VN (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016–3027.
9 Bohnsack MT, Czaplinski K and Gorlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191.
10 Hutvágner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T and Zamore PD (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838.
11 Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K and Shiekhattar R (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744.
12 Kwak PB and Tomari Y (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat Struct Mol Biol* **19**, 145–151.
13 Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T and Tomari Y (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39**, 292–299.
14 Höck J and Meister G (2008) The Argonaute protein family. *Genome Biol* **9**, 210.
15 Hammond SM, Bernstein E, Beach D and Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296.
16 Lewis BP, Burge CB and Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20.
17 Kozomara A and Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68–D73.
18 Friedman RC, Farh KK, Burge CB and Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105.
19 Jonas S and Izaurralde E (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**, 421–433.
20 Berezikov E, Chung WJ, Willis J, Cuppen E and Lai EC (2007) Mammalian mirtron genes. *Mol Cell* **28**, 328–336.
21 Okamura K, Hagen JW, Duan H, Tyler DM and Lai EC (2007) The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89–100.
22 Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND *et al.* (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**, 1694–1698.
23 Ha M and Kim VN (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509–524.
24 Denli AM, Tops BBJ, Plasterk RHA, Ketting RF and Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235.
25 Gregory RI, Yan K, Amuthan G, Chendrimada T, Doratotaj B, Cooch N and Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240.
26 Landthaler M, Yalcin A and Tuschl T (2004) The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* **14**, 2162–2167.
27 Tang X, Zhang Y, Tucker L and Ramratnam B (2010) Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. *Nucleic Acids Res* **38**, 6610–6619.
28 Kwon SC, Nguyen TA, Choi Y-G, Jo MH, Hohng S, Kim VN and Woo J-S (2016) Structure of human DROSHA. *Cell* **164**, 81–90.

29 Nguyen TA, Jo MH, Choi Y-G, Park J, Kwon SC, Hohng S, Kim VN and Woo J-S (2015) Functional anatomy of the human Microprocessor. *Cell* **161**, 1374–1387.

30 Shiohama A, Sasaki T, Noda S, Minoshima S and Shimizu N (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochem Biophys Res Commun* **304**, 184–190.

31 Senturia R, Faller M, Yin S, Loo JA, Cascio D, Sawaya MR, Hwang D, Clubb RT and Guo F (2010) Structure of the dimerization domain of DiGeorge critical region 8: the dimerization domain of DGCR8. *Protein Sci* **19**, 1354–1365.

32 Barr I, Smith AT, Senturia R, Chen Y, Scheidemantle BD, Burstyn JN and Guo F (2011) DiGeorge critical region 8 (DGCR8) is a double-cysteine-ligated heme protein. *J Biol Chem* **286**, 16716–16725.

33 Faller M, Matsunaga M, Yin S, Loo JA and Guo F (2007) Heme is involved in microRNA processing. *Nat Struct Mol Biol* **14**, 23–29.

34 Partin AC, Ngo TD, Herrell E, Jeong B-C, Hon G and Nam Y (2017) Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs. *Nat Commun* **8**, 1737.

35 Weitz SH, Gong M, Barr I, Weiss S and Guo F (2014) Processing of microRNA primary transcripts requires heme in mammalian cells. *Proc Natl Acad Sci USA* **111**, 1861–1866.

36 Cai X, Hagedorn CH and Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957–1966.

37 Lee Y, Jeon K, Lee JT, Kim S and Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663–4670.

38 Conrad T, Marsico A, Gehre M and Orom UA (2014) Microprocessor activity controls differential miRNA biogenesis in vivo. *Cell Rep* **9**, 542–554.

39 Feng Y, Zhang X, Song Q, Li T and Zeng Y (2011) Drosha processing controls the specificity and efficiency of global microRNA expression. *Biochim Biophys Acta* **1809**, 700–707.

40 Louloupi A, Ntini E, Liz J and Ørom UA (2017) Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs. *RNA* **23**, 892–898.

41 Zeng Y, Yi R and Cullen BR (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **24**, 138–148.

42 Auyeung VC, Ulitsky I, McGeary SE and Bartel DP (2013) Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**, 844–858.

43 Han J, Lee Y, Yeom K-H, Nam J-W, Heo I, Rhee J-K, Sohn SY, Cho Y, Zhang B-T and Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901.

44 Zeng Y and Cullen BR (2005) Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**, 27595–27603.

45 Zhang X and Zeng Y (2010) The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res* **38**, 7689–7697.

46 Fang W and Bartel DP (2015) The menu of features that define primary microRNAs and enable de novo design of microRNA genes. *Mol Cell* **60**, 131–145.

47 Roden C, Gaillard J, Kanoria S, Rennie W, Barish S, Cheng J, Pan W, Liu J, Cotsapas C, Ding Y *et al.* (2017) Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. *Genome Res* **27**, 374–384.

48 Sperber H, Beem A, Shannon S, Jones R, Banik P, Chen Y, Ku S, Varani G, Yao S and Ruohola-Baker H (2014) miRNA sensitivity to Drosha levels correlates with pre-miRNA secondary structure. *RNA* **20**, 621–631.

49 Chaulk SG, Thede GL, Kent OA, Xu Z, Gesner EM, Veldhoen RA, Khanna SK, Goping IS, MacMillan AM, Mendell JT *et al.* (2011) Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biol* **8**, 1105–1114.

50 Chakraborty S and Krishnan Y (2017) A structural map of oncomiR-1 at single-nucleotide resolution. *Nucleic Acids Res* **45**, 9694–9705.

51 Contrant M, Fender A, Chane-Woon-Ming B, Randrianjafy R, Vivet-Boudou V, Richer D and Pfeffer S (2014) Importance of the RNA secondary structure for the relative accumulation of clustered viral microRNAs. *Nucleic Acids Res* **42**, 7981–7996.

52 Du P, Wang L, Sliz P and Gregory RI (2015) A biogenesis step upstream of Microprocessor controls miR-17 ~ 92 expression. *Cell* **162**, 885–899.

53 Truscott M, Islam ABMMK and Frolov MV (2016) Novel regulation and functional interaction of polycistronic miRNAs. *RNA* **22**, 129–138.

54 Haar J, Contrant M, Bernhardt K, Feederle R, Diederichs S, Pfeffer S and Delecluse H-J (2016) The expression of a viral microRNA is regulated by clustering to allow optimal B cell transformation. *Nucleic Acids Res* **44**, 1326–1341.

55 Ma H, Wu Y, Choi J-G and Wu H (2013) Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc Natl Acad Sci USA* **110**, 20687–20692.

56 Burke JM, Kelenis DP, Kincaid RP and Sullivan CS (2014) A central role for the primary microRNA stem in guiding the position and efficiency of Drosha processing of a viral pri-miRNA. *RNA* **20**, 1068–1077.

57 Kim B, Jeong K and Kim VN (2017) Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates. *Mol Cell* **66**, 258–269.e5.

58 Yeom K-H, Lee Y, Han J, Suh MR and Kim VN (2006) Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res* **34**, 4622–4629.

59 Zhang X, Li P, Lin J, Huang H, Yin B and Zeng Y (2017) The insertion in the double-stranded RNA binding domain of human Drosha is important for its function. *Biochim Biophys Acta* **1860**, 1179–1188.

60 Kranick JC, Chadalavada DM, Sahu D and Showalter SA (2017) Engineering double-stranded RNA binding activity into the Drosha double-stranded RNA binding domain results in a loss of microRNA processing function. *PLoS One* **12**, e0182445.

61 Roth BM, Ishimaru D and Hennig M (2013) The core Microprocessor component DiGeorge syndrome critical region 8 (DGCR8) is a nonspecific RNA-binding protein. *J Biol Chem* **288**, 26785–26799.

62 Faller M, Toso D, Matsunaga M, Atanasov I, Senturia R, Chen Y, Zhou ZH and Guo F (2010) DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA* **16**, 1570–1583.

63 Herbert KM, Sarkar SK, Mills M, Delgado De la Herran HC, Neuman KC and Steitz JA (2016) A heterotrimer model of the complete Microprocessor complex revealed by single-molecule subunit counting. *RNA* **22**, 175–183.

64 Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD and Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198.

65 Sohn SY, Bae WJ, Kim JJ, Yeom K-H, Kim VN and Cho Y (2007) Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* **14**, 847–853.

66 Chang T-C, Pertea M, Lee S, Salzberg SL and Mendell JT (2015) Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res* **25**, 1401–1409.

67 Seong Y, Lim D-H, Kim A, Seo JH, Lee YS, Song H and Kwon Y-S (2014) Global identification of target recognition and cleavage by the Microprocessor in human ES cells. *Nucleic Acids Res* **42**, 12806–12821.

68 Macias S, Plass M, Stajuda A, Michlewski G, Eyras E and Cáceres JF (2012) DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* **19**, 760–766.

69 Lee D and Shin C (2018) Emerging roles of DROSHA beyond primary microRNA processing. *RNA Biol* **15**, 186–193.

70 Macias S, Cordiner RA and Cáceres JF (2013) Cellular functions of the Microprocessor. *Biochem Soc Trans* **41**, 838–843.

71 Marzi MJ, Ghini F, Cerruti B, de Pretis S, Bonetti P, Giacomelli C, Gorski MG, Kress T, Pelizzola M, Muller H *et al.* (2016) Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res* **26**, 554–565.

72 Choudhury NR, de Lima Alves F, de Andres-Aguayo L, Graf T, Caceres JF, Rappsilber J and Michlewski G (2013) Tissue-specific control of brain-enriched miR-7 biogenesis. *Genes Dev* **27**, 24–38.

73 Lee H, Han S, Kwon CS and Lee D (2016) Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein Cell* **7**, 100–113.

74 Reddy KB (2015) MicroRNA (miRNA) in cancer. *Cancer Cell Int* **15**, 38.

75 Kim Y-J, Park S-J, Choi EY, Kim S, Kwak HJ, Yoo BC, Yoo H, Lee S-H, Kim D, Park JB *et al.* (2011) PTEN modulates miR-21 processing via RNA-regulatory protein RNH1. *PLoS One* **6**, e28308.

76 Michlewski G and Cáceres JF (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol* **17**, 1011–1018.

77 Moy RH, Cole BS, Yasunaga A, Gold B, Shankarling G, Varble A, Molleston JM, tenOever BR, Lynch KW and Cherry S (2014) Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell* **158**, 764–777.

78 Sakamoto S, Aoki K, Higuchi T, Todaka H, Morisawa K, Tamaki N, Hatano E, Fukushima A, Taniguchi T and Agata Y (2009) The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol Cell Biol* **29**, 3754–3769.

79 Zhao L, Mao Y, Zhao Y and He Y (2016) DDX3X promotes the biogenesis of a subset of miRNAs and the potential roles they played in cancer development. *Sci Rep* **6**, 32739.

80 Blahna MT and Hata A (2013) Regulation of miRNA biogenesis as an integrated component of growth factor signaling. *Curr Opin Cell Biol* **25**, 233–240.

81 Mori M, Triboulet R, Mohseni M, Schlegelmilch K, Shrestha K, Camargo FD and Gregory RI (2014) Hippo signaling regulates Microprocessor and links cell-density-dependent miRNA biogenesis to cancer. *Cell* **156**, 893–906.

82 Treiber T, Treiber N, Plessmann U, Harlander S, Daiß J-L, Eichner N, Lehmann G, Schall K, Urlaub H and Meister G (2017) A compendium of RNA-binding proteins that regulate microRNA biogenesis. *Mol Cell* **66**, 270–284.e13.

83 Nussbacher JK and Yeo GW (2018) Systematic discovery of RNA binding proteins that regulate microRNA levels. *Mol Cell* **69**, 1005–1016.e7.

84 Fletcher CE, Godfrey JD, Shibakawa A, Bushell M and Bevan CL (2016) A novel role for GSK3β as a modulator of Drosha Microprocessor activity and MicroRNA biogenesis. *Nucleic Acids Res* **45**, 2809–2828.

85 Cheng T-L, Wang Z, Liao Q, Zhu Y, Zhou W-H, Xu W and Qiu Z (2014) MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. *Dev Cell* **28**, 547–560.

86 Dini Modigliani S, Morlando M, Errichelli L, Sabatelli M and Bozzoni I (2014) An ALS-associated mutation in the FUS 3′-UTR disrupts a microRNA–FUS regulatory circuitry. *Nat Commun* **5**, 4335.

87 Alarcón CR, Lee H, Goodarzi H, Halberg N and Tavazoie SF (2015) N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482–485.

88 Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I and Proudfoot NJ (2008) Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* **15**, 902–909.

89 Liu H, Liang C, Kollipara RK, Matsui M, Ke X, Jeong B-C, Wang Z, Yoo KS, Yadav GP, Kinch LN *et al.* (2016) HP1BP3, a chromatin retention factor for co-transcriptional microRNA processing. *Mol Cell* **63**, 420–432.

90 Suzuki HI, Young RA and Sharp PA (2017) Super-enhancer-mediated RNA processing revealed by integrative microRNA network analysis. *Cell* **168**, 1000–1014.e15.

91 Church VA, Pressman S, Isaji M, Truscott M, Cizmecioglu NT, Buratowski S, Frolov MV and Carthew RW (2017) Microprocessor recruitment to elongating RNA polymerase II is required for differential expression of microRNAs. *Cell Rep* **20**, 3123–3134.

92 Agranat-Tamir L, Shomron N, Sperling J and Sperling R (2014) Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. *Nucleic Acids Res* **42**, 4640–4651.

93 Janas MM, Khaled M, Schubert S, Bernstein JG, Golan D, Veguilla RA, Fisher DE, Shomron N, Levy C and Novina CD (2011) Feed-forward microprocessing and splicing activities at a microRNA–containing intron. *PLoS Genet* **7**, e1002330.

94 Tang X, Wen S, Zheng D, Tucker L, Cao L, Pantazatos D, Moss SF and Ramratnam B (2013) Acetylation of Drosha on the N-terminus inhibits its degradation by ubiquitination. *PLoS One* **8**, e72503.

95 Wada T, Kikuchi J and Furukawa Y (2012) Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8. *EMBO Rep* **13**, 142–149.

96 Herbert KM, Pimienta G, DeGregorio SJ, Alexandrov A and Steitz JA (2013) Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile. *Cell Rep* **5**, 1070–1081.

97 Zhu C, Chen C, Huang J, Zhang H, Zhao X, Deng R, Dou J, Jin H, Chen R, Xu M *et al.* (2015) SUMOylation at K^{707} of DGCR8 controls direct function of primary microRNA. *Nucleic Acids Res* **43**, 7945–7960.

98 Chen X, Fan Z, McGee W, Chen M, Kong R, Wen P, Xiao T, Chen X, Liu J, Zhu L *et al.* (2017) TDP-43 regulates cancer-associated microRNAs. *Protein Cell* https://doi.org/10.1007/s13238-017-0480-9.

99 Ye P, Liu Y, Chen C, Tang F, Wu Q, Wang X, Liu C-G, Liu X, Liu R, Liu Y *et al.* (2015) An mTORC1-Mdm2-Drosha axis for miRNA biogenesis in response to glucose- and amino acid-deprivation. *Mol Cell* **57**, 708–720.

100 Michlewski G, Guil S, Semple CA and Cáceres JF (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* **32**, 383–393.

101 Wu S-L, Fu X, Huang J, Jia T-T, Zong F-Y, Mu S-R, Zhu H, Yan Y, Qiu S, Wu Q *et al.* (2015) Genome-wide analysis of YB-1-RNA interactions reveals a novel role of YB-1 in miRNA processing in glioblastoma multiforme. *Nucleic Acids Res* **43**, 8516–8528.

102 Van Wynsberghe PM, Kai ZS, Massirer KB, Burton VH, Yeo GW and Pasquinelli AE (2011) LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in *Caenorhabditis elegans*. *Nat Struct Mol Biol* **18**, 302–308.

103 Mao G, Lee S, Ortega J, Gu L and Li G-M (2012) Modulation of microRNA processing by mismatch repair protein MutLα. *Cell Res* **22**, 973.

104 Hong S, Noh H, Chen H, Padia R, Pan ZK, Su S-B, Jing Q, Ding H-F and Huang S (2013) Signaling by p38 MAPK stimulates nuclear localization of the Microprocessor component p68 for processing of selected primary microRNAs. *Sci Signal* **6**, ra16.

105 Suzuki HI, Yamagata K, Sugimoto K, Iwamoto T, Kato S and Miyazono K (2009) Modulation of microRNA processing by p53. *Nature* **460**, 529–533.

106 Han C, Liu Y, Wan G, Choi HJ, Zhao L, Ivan C, He X, Sood AK, Zhang X and Lu X (2014) The RNA-binding protein DDX1 promotes primary microRNA maturation and inhibits ovarian tumor progression. *Cell Rep* **8**, 1447–1460.

107 Chu Y-D, Chen H-K, Huang T and Chan S-P (2016) A novel function for the DEAD-box RNA helicase DDX-23 in primary microRNA processing in *Caenorhabditis elegans*. *Dev Biol* **409**, 459–472.

108 Wu H, Sun S, Tu K, Gao Y, Xie B, Krainer AR and Zhu J (2010) A splicing-independent function of SF2/ASF in microRNA processing. *Mol Cell* **38**, 67–77.

109 Heinrich E-M, Wagner J, Krüger M, John D, Uchida S, Weigand JE, Suess B and Dimmeler S (2013) Regulation of miR-17-92a cluster processing by the microRNA binding protein SND1. *FEBS Lett* **587**, 2405–2411.
110 Zisoulis DG, Kai ZS, Chang RK and Pasquinelli AE (2012) Autoregulation of microRNA biogenesis by let-7 and Argonaute. *Nature* **486**, 541–544.
111 Tang R, Li L, Zhu D, Hou D, Cao T, Gu H, Zhang J, Chen J, Zhang C-Y and Zen K (2012) Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system. *Cell Res* **22**, 504.
112 Wang D, Sun X, Wei Y, Liang H, Yuan M, Jin F, Chen X, Liu Y, Zhang C-Y, Li L *et al.* (2018) Nuclear miR-122 directly regulates the biogenesis of cell survival oncomiR miR-21 at the posttranscriptional level. *Nucleic Acids Res* **46**, 2012–2029.
113 Liz J, Portela A, Soler M, Gómez A, Ling H, Michlewski G, Calin GA, Guil S and Esteller M (2014) Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region. *Mol Cell* **55**, 138–147.
114 Guo J, Fang W, Sun L, Lu Y, Dou L, Huang X, Tang W, Yu L and Li J (2018) Ultraconserved element uc.372 drives hepatic lipid accumulation by suppressing miR-195/miR4668 maturation. *Nat Commun* **9**, 612.
115 Jiang L, Shao C, Wu Q-J, Chen G, Zhou J, Yang B, Li H, Gou L-T, Zhang Y, Wang Y *et al.* (2017) NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nat Struct Mol Biol* **24**, 816–824.
116 Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R and Nishikura K (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* **13**, 13–21.
117 Chawla G and Sokol NS (2014) ADAR mediates differential expression of polycistronic microRNAs. *Nucleic Acids Res* **42**, 5245–5255.
118 Heale BSE, Keegan LP, McGurk L, Michlewski G, Brindle J, Stanton CM, Caceres JF and O'Connell MA (2009) Editing independent effects of ADARs on the miRNA/siRNA pathways. *EMBO J* **28**, 3145–3156.
119 Roignant J-Y and Soller M (2017) m6A in mRNA: an ancient mechanism for fine-tuning gene expression. *Trends Genet* **33**, 380–390.
120 Alarcón CR, Goodarzi H, Lee H, Liu X, Tavazoie S and Tavazoie SF (2015) HNRNPA2B1 is a mediator of m6A-dependent nuclear RNA processing events. *Cell* **162**, 1299–1308.
121 Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR and Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906.
122 Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414.
123 Wang F, Song W, Zhao H, Ma Y, Li Y, Zhai D, Pi J, Si Y, Xu J, Dong L *et al.* (2017) The RNA-binding protein QKI5 regulates primary miR-124-1 processing via a distal RNA motif during erythropoiesis. *Cell Res* **27**, 416–439.
124 Klinge CM (2012) miRNAs and estrogen action. *Trends Endocrinol Metab* **23**, 223–233.
125 Lauressergues D, Couzigou J-M, Clemente HS, Martinez Y, Dunand C, Bécard G and Combier J-P (2015) Primary transcripts of microRNAs encode regulatory peptides. *Nature* **520**, 90–93.
126 Hansen TB, Venø MT, Jensen TI, Schaefer A, Damgaard CK and Kjems J (2016) Argonaute-associated short introns are a novel class of gene regulators. *Nat Commun* **7**, 11538.
127 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The protein data bank. *Nucleic Acids Res* **28**, 235–242.

I.A.2.a Non canonical miRNA biogenesis pathway

Since the characterization of the canonical biogenesis pathway, alternative ways to produce mature miRNA, which are microprocessor or DICER independent, have emerged (Figure 2).

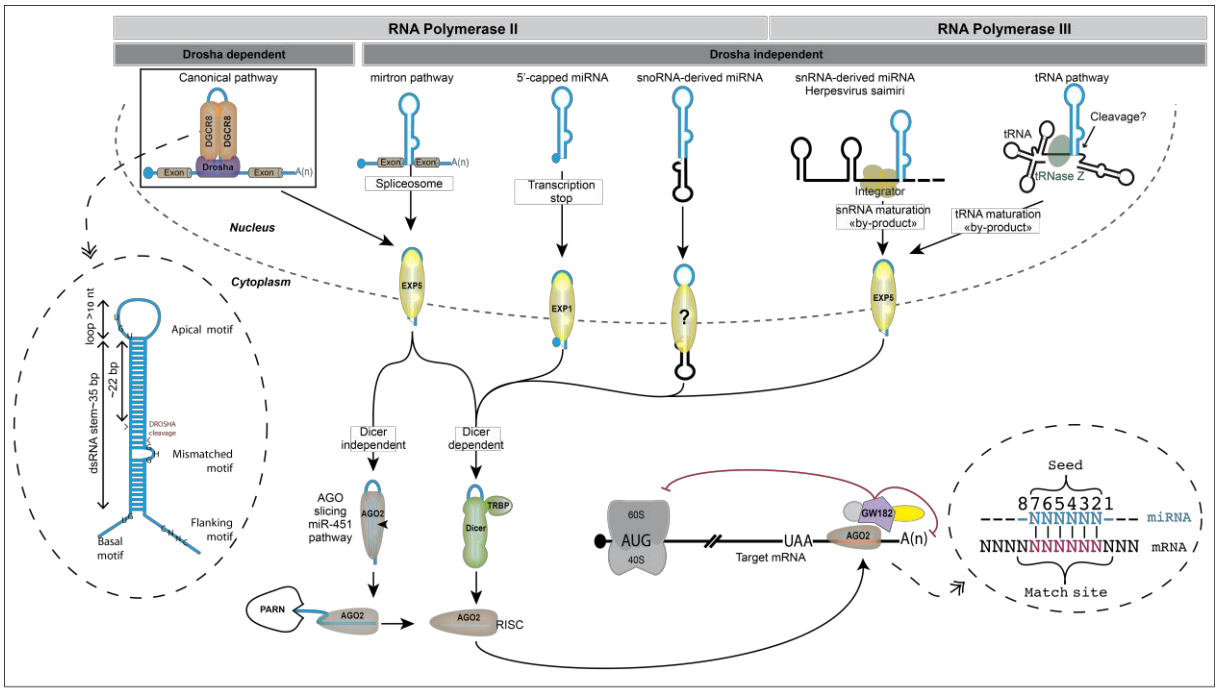


Figure 2. miRNA biogenesis. In the canonical miRNA biogenesis pathway pri-miRNAs containing a single or a cluster of stem-loop precursors are transcribed by the polymerase II. The microprocessor (DROSHA/DGCR8) cleaves the pri-miRNA to release pre-miRNA which is exported by the exportin 5 (EXP5). Pri-miRNA has to display several structural and sequence features to be correctly recognized by the microprocessor (left dashed circle). The pre-miRNA is then processed by DICER into mature miRNA duplexes of which one strand is loaded into an argonaute protein (AGO) forming the RNA-induced silencing complex (RISC). RISC is directed to targeted mRNAs though miRNA seed match binding (right dashed circle). Mirtron, 5'-capped miRNA, snoRNA-derived miRNA are microprocessor independent miRNAs transcribed by the polymerase II. Spliceosome and transcription stop are respectively responsible of the pre-miRNA generation of the mirtron and 5'-capped miRNA. snRNA-derived and tRNA-derived pre-miRNAs are generated upon snRNA and tRNA maturation respectively. Almost all pre-miRNAs are DICER dependent. Pre-miR-451 is processed by AGO2 to generate one mature miRNA.

Deep sequencing technology allowed several groups to identify miRNA duplex mapping on short hairpin intron within *D. melanogaster* and *C elegans* genome (Chung *et al.*, 2011). For those miRNAs, which were coined mirtrons, their extremity corresponds to splice acceptor and donor group meaning they can bypass the microprocessor cleavage (Ruby *et al.*, 2007). The spliceosome generates a pre-miRNA which enter in the canonical pathway as EXP-5 and DICER substrate. If the splicing site do not correspond directly to one of the miRNA extremity, they are called tailed mirtrons and require exoribonucleases activity to shorten the 5' or 3'

extremity (Flynt *et al.*, 2010). AGO immunoprecipitation followed by deep sequencing has allowed the discovery of other microprocessor independent miRNAs. Thus, some small nucleolar RNAs (snoRNAs) can mimics a DICER substrate and generate functional mature miRNA (Ender *et al.*, 2008). Those miRNAs are called snoRNA-derived miRNAs (sdRNAs). Analogous to the microprocessor independent, DICER dependent sdRNAs, some tRNA-derived miRNAs (tdRNAs) have been detected with deep sequencing approaches. For example, the tRNA-Ile precursor can fold in different way, one of which is a DICER substrate and therefore produce mature tRNA-derived miRNAs (Babiarz *et al.*, 2008). In addition during the tRNA maturation some small RNA species can be generated by the RNase Z cleavage and enter the miRNA biogenesis pathway after bypassing microprocessor activity (Haussecker *et al.*, 2010). The implication of the RNase Z in the production of miRNA has been validated by studying a miRNA expressed by the murine gammaherpesvirus 68 (MHV68) for which RNase Z cleaves the pre-miRNA 5' end (Bogerd *et al.*, 2010). Essential miRNA pathway component has been mutated individually to identify another alternative miRNA production. In a mutant microprocessor context, 5'-capped miRNA precursors have been characterized to produce mature miRNAs. In this case, the 5' end of the miRNA precursor correspond to the transcription start site and the 3' end correspond to a transcription termination. Those capped-pre-miRNAs are exported in the cytoplasm by the EXP-1 and cleaved by DICER to generate a functional 3p-miRNA (Xie *et al.*, 2013). The majority of the non-canonical miRNA biogenesis bypass the microprocessor. However, a few shorter pre-miRNA, such as pre-miR451, are DICER independent and requires the slicing ability of AGO2 to generate a functional mature miRNA (Cheloufi *et al.*, 2010).

I.A.3 miRNA function

Here, I focus on the mechanisms by which the mature miRNA recognizes and regulates mRNA translation and abundance.

I.A.3.a miRNA translational repression

The loading of the mature miRNA onto one of the four AGO protein allows the formation of the RISC complex (*RNA induced silencing complex*), which can then exert its translation-repression function (Ipsaro and Joshua-Tor, 2015) (Figure 2). Targeted mRNAs can either be sliced via AGO2 (Meister *et al.*, 2004) or destabilized via the other AGOs (Jonas and Izaurralde, 2015) depending on the sequence complementarity with the miRNAs (Bartel, 2009; Becker *et al.*, 2019). mRNA translation inhibition and destabilization is the main mechanism used by the animal miRNAs due to their imperfect complementarity with their target (Guo *et al.*, 2010). In this case, targeted mRNA destabilization happens via the recruitment of the adaptor protein GW182 (Eulalio *et al.*, 2008) in Drosophila, or TNRC6 in mammals (Fabian *et al.*, 2011), which is able to recruit deadenylase complexes (Chen and Shyu, 2011) as well as the CCR4-NOT complex (Sheu-Gruttadauria and MacRae, 2018). AGO2-mediated mRNA slicing, for its part, is possible thanks to the presence of its two Nter domains in concert with the PIWI domain together with perfect binding of the miRNA on its target (Schürmann *et al.*, 2013).

I.A.3.b miRNA target recognition

The identification of miRNA targets is easier in plants because they display near perfect complementarity with the targeting miRNA (Rhoades *et al.*, 2002). However, it is not the case for animal miRNA (Guo *et al.*, 2010). Two features have been described to explain animal miRNA target recognition. First, animal miRNAs require Watson-Crick base paring with their targets, more precisely in the 5' proximal region of the miRNA (nucleotides 2 to 7) called the miRNA seed (Brennecke *et al.*, 2005) (Figure 2). This limited base pairing means that a given animal miRNA can theoretically target hundreds of potential mRNA targets, and that extensive binding of the miRNA 3' extremity could explain target specificity within a miRNA family (Lewis *et al.*, 2005). Indeed, pairing of the miRNA seed region is sufficient for mRNA translation inhibition (Lewis *et al.*, 2005). Finally, miRNAs and seed-matches, generally present in the mRNA 3' UTR, are preferentially conserved among phylum (Friedman *et al.*, 2008). Multiple different kind of additional base pairing on the 3' part of the miRNA are also involved in the target recognition (Bartel, 2009), and allows among other things imperfect seed-match compensation, or targetome extension (Grimson *et al.*, 2007). Taken together, it clearly suggests that miRNA-mediated mRNA translation inhibition is a complex process, which allows the cell to precisely fine-tune individual protein output.

I.A.3.c miRNA and target abundance

In human, at least 60 % of the mRNAs are considered as potential miRNA target (Friedman *et al.*, 2008). In addition, mRNA targets can display several identical or different miRNA binding sites. It is accepted that the number of binding sites present within the transcriptome affect the miRNAs activity. For example, an artificial mRNA or circular RNAs bearing an important number of binding sites can titrate miRNAs away from their natural targets (Memczak *et al.*, 2013). Consequently, site-containing transcript can act as competitor and regulate other transcript bearing seed match. Furthermore, the affinity between a given miRNA and its target also affect target abundance regardless of molecules concentration (Denzler *et al.*, 2014). Recent study proposes that affinity feature predominantly explains target repression. However, miRNA level seems to be important to characterize the magnitude at which a target is initially repressed (Denzler *et al.*, 2016). To add even more complexity, cooperative action of miRNA on the same target can be an important determinant to explain mRNA translation inhibition (Denzler *et al.*, 2016).

I.A.4 Clustered miRNA

Animal miRNAs are present in intergenic regions as well as pre-mRNA introns, and within ncRNA genes (Rodriguez, 2004). As miRNAs are fundamental for a multitude of cellular processes, it is not surprising to identify lineage-specific miRNA suggesting their importance in life evolution history (Fromm *et al.*, 2015). Interestingly, more than 25 % of mammals miRNA precursors group on the same transcript units forming clustered miRNAs (Altuvia, 2005). The latest version of miRBase (v22.1) contains 1,917 human miRNA precursors of which more than 470 mature miRNA are distributed in 153 clusters (http://www.mirbase.org/) (Kozomara and Griffiths-Jones, 2014).

I.A.4.a miRNA evolution

It has been reported that the apparition of new miRNAs occurs preferentially within clustered miRNAs (Hertel *et al.*, 2006). Clustered miRNAs represents an interesting strategy to bring several target genes under a coordinate miRNA control (Mohammed *et al.*, 2014; Wang *et al.*, 2016) similar to prokaryote operons. However, a recent study claims that there is no statistical evidence that all clustered miRNAs have evolved to target common genes even though a few exception exists (Marco, 2019). The origin of new miRNA genes seems to be usually achieved by duplication and sequence divergence, similar to protein-coding genes (Marco *et al.*, 2013) or by *de novo* formation (Meunier *et al.*, 2013). miRNA *de novo* formation can more easily generate new miRNA compare to duplication since transcription naturally generate RNA stem-loop structures (Wang *et al.*, 2016). It is clear that miRNA duplicates will share different features from their maturation to their target, at least few generations after the duplication. It is different for the miRNA created *de novo*, which can possess totally different features. In any case, it seems that new miRNA preferentially appears within transcripts already bearing miRNA. Indeed, among a transcript, a pre-existing functional miRNA could act as protective sequence allowing new downstream miRNA to evolve. If the pre-existing miRNA is important for the cell, its transcription and maturation cannot be disrupted by the new miRNA. Thus, it could give time to the new miRNA precursor to develop new function (Wang *et al.*, 2016).

I.A.4.b Regulation of miRNA cluster expression

As for all biological processes, the regulation of expression can happen at different levels. Concerning the regulation of a miRNA cluster *per se,* obviously transcriptional regulation though transcription factors (Yan *et al.*, 2016) or epigenetics (Yang *et al.*, 2015) is the main mechanism involved. In addition, genetic alteration, by gene duplication or suppression, is an important factors responsible of miRNA cluster mis-expression which are commonly found and associated with pathology, as reported for example for the miR-17/92 cluster (Fuziwara and Kimura, 2015). It is important to highlight that a miRNA can target its own transcription factor, thereby inducing a feedback loop regulation and a homeostasis control of both transcription factors and miRNAs. Thus, if one of the two molecules display altered function, due to nucleotide or amine acid polymorphism, it can partially explain the apparition of a cellular pathology (Kent *et al.*, 2016).

I.A.4.c Regulation of clustered miRNA specific feature

Clustered miRNA are interesting molecules to highlight the importance of the post-transcriptional regulation of miRNA expression. Indeed, even though miRNA precursors are present within the same transcript, their mature forms are usually expressed at different levels. Structural and sequence features of miRNA precursors implicated in their maturation as well as the involvement of specific cofactor proteins, are described in part I.A.2 and Figure 2. However, clustered miRNAs are subjected to another post-transcriptional regulation process. Indeed, apart from the miRNA precursors *per se*, flanking RNA structures and sequences are critical for their expression regulation (Chaulk *et al.*, 2011; Contrant *et al.*, 2014; Baldrich *et*

al., 2016; Haar *et al.*, 2016; Kotaki *et al.*, 2017). In other words, at the post-transcriptional level any event (nucleotide modification, protein binding, cleavage…) occurring upstream or downstream a miRNA precursor can impact its expression. Because miRNA maturation is a co-transcriptional process (Morlando *et al.*, 2008), it does make sense to think this mechanism occurs. Thus, due to RNA structural dynamics (Ganser *et al.*, 2019), the order and the maturation kinetic of a miRNA precursor embedded within a miRNA cluster directly impacts on downstream or upstream miRNA precursors, as highlighted with the Epstein-Barr virus miRNA cluster (Haar *et al.*, 2016) or the human cluster miR-11/998 (Truscott *et al.*, 2016). This concept could be really important for both virus-encoded and human miRNA clusters.

I.A.5 miRNA & host-virus interaction

Due to their nature, viruses are masters of gene regulation. Their small size genome, relative to their host, forces them to deeply hijack all host machinery as well as regulation mechanisms (Fields *et al.*, 2013). Since miRNAs are implicated in most, if not all, cellular processes, it is not surprising that miRNA pathways play fundamental role in viral infection and host antiviral responses. This miRNA-mediated host-virus interactions is involved in, among other, viral replication, viral cycle control, immune evasion, cell transformation as well as host miRNA biogenesis and turnover impairment. Depending on the virus genetic nature and its ability to express its own miRNAs, it can impair or hijack the miRNA pathway as well as specifically use or repress host miRNAs (Girardi *et al.*, 2018).

I.A.5.a Virus-mediated host miRNAs turnover and hijack

Some viruses express transcripts displaying sequence complementarity to host miRNAs to promote either viral replication or control of specific miRNA degradation. Some host miRNA can be specifically degraded by this mean: miR-27 is a good example. This miRNA is degraded by different herpesviruses through expression of viral transcripts inducing miRNA decay (Marcinowski *et al.*, 2012; Guo *et al.*, 2014; Haas *et al.*, 2016) via target RNA-directed miRNA degradation process (Ameres *et al.*, 2010; de la Mata *et al.*, 2015). miRNA-mediated translation inhibition can also be used by viruses to limit their replication and avoid strong immune responses and viral clearance (Trobaugh *et al.*, 2014). Surprisingly, cellular miRNAs can display a positive effect on the viral replication and gene expression by direct binding of the viral genome. Thus, the liver specific miR-122 is essential for hepatitis C virus (HCV) replication cycle, through binding to the viral 5' UTR (Jopling *et al.*, 2005; Li *et al.*, 2015).

Obviously, the impact of host miRNAs on viral infection is fundamental and influence viral tropism, antiviral responses and viral replication. In addition, in a more global point of view, viral genome and transcripts can be considered as competing nucleic acids, which destabilize host network balances leading to host transcriptome and proteome variation generally in favour of the viral infection (Li *et al.*, 2014).

I.A.5.b Virus-encoded miRNAs

Other viruses, mainly DNA viruses, hijack the host miRNA machinery to express their own miRNAs to take advantages of this regulation pathway. Indeed, miRNAs display several

advantages: they are not antigenic, have important regulatory potential, can be inserted into multifunctional transcript and can evolve rapidly. Indeed, a point mutation within the miRNA seed region alter its target repertoire. Within the miRNA precursor, mutations can also alter their maturation. Taken together miRNAs represent a terrific viral tool to establish a favourable cellular environment. It is not surprising that most viruses that encode miRNAs have a DNA genome. Indeed, they have to go through a nuclear stage during their replication cycle, where they have access to the miRNA maturation enzyme DROSHA. In addition, since the miRNA precursor is not directly part of the genome, its cleavage will not be deleterious for the virus, as it would be for a virus with an RNA genome.

Herpesviruses represent the most striking example of canonical miRNA pathway hijack. Their ability to possess two distinct cycles, latent and lytic, which correspond to two distinct herpesvirus genes expression leading to respectively the maintenance of the viral genome upon cell division and the production of new viral particles leading to the cell lysis (Fields *et al.*, 2013) is surely one key explanation for this miRNA pathway hijacking. During latency, herpesvirus greatly restrict their coding RNA expression. It is also during latency, that most herpesviruses express their miRNAs, which are able to target both viral and cellular transcripts in order to help create a favourable micro-environment. Herpesviruses miRNAs have indispensable functions for the virus and act on a multitude of cellular fundamental pathways from cell cycle and apoptosis control to immune response (Piedade and Azevedo-Pereira, 2016). KSHV's miRNA targetome and function are described part I.C.

As mentioned, miRNAs can be produced by alternative pathway (Yang and Lai, 2011; Xie and Steitz, 2014), which is also true for viral-miRNAs. Other DNA viruses co-transcribed pre-miRNA downstream of other ncRNA such as the MHV68 to bypass the microprocessor maturation steps. Saimiri herpesvirus releases three pre-miRNA during the maturation of a snRNA like precursor viral transcript (Cazalla *et al.*, 2011). This ncRNA-pre-miRNA chimeric strategies could be considered as an optimization and an economic way for the virus to use ncRNA maturation by-products as miRNA precursors (Figure 2).

Even though it could be deleterious for an RNA virus to express viral-miRNA, it is now clear that few retroviruses are able to generate their own miRNAs (Kincaid *et al.*, 2012, 2014). They mainly use the RNA polymerase III to transcribe viral-miRNA precursors from their integrated genome. The bovine leukemia virus (BLV) expresses pre-miRNAs which 5' and 3' ends are defined by the polymerase III initiation and termination sites. Those pre-miRNAs are then exported in the cytoplasm to be process by DICER (Burke *et al.*, 2014). Simian foamy virus (SFV) expresses its miRNA through polymerase III transcription. Those viral-pri-miRNAs fold in different way including DROSHA substrate fold. Start and termination sites of the polymerase III can also generate SFV pre-miRNAs (Kincaid *et al.*, 2014).

I.B. KSHV

The Kaposi's sarcoma-associated herpesvirus (KSHV), also known as human herpesvirus 8 (HHV8), is the model used during my thesis. It is the last human herpesvirus identified in 1994 in a patient with acquired immunodeficiency patient displaying Kaposi's sarcoma pathology (Chang *et al.*, 1994). *Herpesviridae* family display characteristic features. They possess a monopartite linear dsDNA genome of ~180 kbp protected within an enveloped

spherical to pleomorphic icosahedric capsid (T=16). All *Herpesviridae* express enzymes involved in nucleic acid metabolism. Transcription, replication and nucleocapsid formation occur in the nucleus. Viral production leads to the host cell destruction. Finally, *Herpesviridae* are capable of lifelong persistence in their host via latency. *Herpesviridae* subfamily are mainly based on the cell tropism and associated pathology. The KSHV is part of the gammaherpesvirinae subfamily, which also contains the Epstein Barr virus (EBV). Their privileged host cells are lymphocytes, as well as epithelial and endothelial cells.

These different characteristics make the *Herpesviridae* family one of the most successful human pathogen. Almost all human adults are at least infected by one *Herpesviridae.* This success comes from a tight coevolution history between *Herpesviridae* and human as well as their ability to establish latency which can be considered as the ultimate form of immune evasion. One can consider *Herpesviridae* as part of the normal microbiotic flora, even though *Herpesviridae* subfamily were not present in all human's population since the beginning of their history. Indeed, it is interesting to precise that *Herpesviridae* primo infection at a later stage of human life is responsible of different diseases and complications compared to early stage of life primo infection. The most striking example is the mononucleosis induced by Epstein-Barr virus. It clearly highlights a fundamental balance between *Herpesviridae* activity and host immune status. *Herpesviridae* are able to express homologs of cellular interleukins, chemokines and associated receptors (Sin and Dittmer, 2012). In addition, on the close relationship between *Herpesviridae* and their host, it has been established that in mice, gammaherpesvirus latency provides resistance to bacterial pathogens *Listeria monocytogenes* and *Yersinia pestis* (Barton *et al.*, 2007). Here, I wanted to highlight that as opposed to other viral replicative strategies such as the "hit and run" strategy corresponding to rapid viral burst and host clearance (e.g. Influenza virus), viral latency is far more tightly intertwined to its host cells, host tissue and host organism status. Consequently, *Herpesviridae* fitness and homeostasis become incredibly complex. *Retroviridae* also display this incredible entwined relationship with their host.

I.B.1 Pathogenesis

Even though KSHV is perfectly adapted to its human host, it is still responsible of several pathologies from inflammatory conditions to malignancies. Kaposi sarcoma, primary effusion lymphoma, multicentric Castleman disease, lymphoma and inflammatory cytokine syndrome are now accepted as related to KSHV infection and are generally found in immune-impaired patients (Goncalves *et al.*, 2017). These different pathologies are associated with uncontrolled cell proliferation upon KSHV latent infection (Feldman *et al.*, 2014). Kaposi sarcoma is a highly vascularized endothelium tumour localized first in the skin, lymph nodes and organs. In this pathology, herpesvirus viral miRNAs increased linearly with the degree of transformation (O'Hara *et al.*, 2009). Primary effusion lymphoma is present as lymphomatous effusions in body cavities, usually the pleural, peritoneal or pericardial cavities. Latent KSHV genome is always detected from 20 to 200 copies in PEL cells (Jenner *et al.*, 2001). In this pathology more than half of the miRNAs present in the infected cell correspond to viral miRNAs (Hoshina *et al.*, 2016).

I.B.2 Viral cycle

Schematic representation of the viral cycle is represented in Figure 3. KSHV viral particles bind host cell heparan sulfate, integrins, DC-sign receptors and enter in the cell via endocytosis. The capsid is released after membrane fusion and transported to the nuclear pore where the viral DNA genome is delivered. Following nuclear genome release, viral infection outcome and gene expression vary depending on the cell host, but generally latent viral cycle is established (Yan *et al.*, 2019). During latency, the viral genome is maintained as a circular chromosome, which is replicated by the host cell machinery and evenly distributed between daughter cells. Latent cycle is disrupted by several cellular stresses and lead to the sequential expression of viral lytic genes. Lytic cycle leads to the cell division as well as viral structural protein expression. Virus assembly begins in the nucleus. The viral envelope is obtained through nucleus membrane budding. Viral progeny is then released from the host cell.

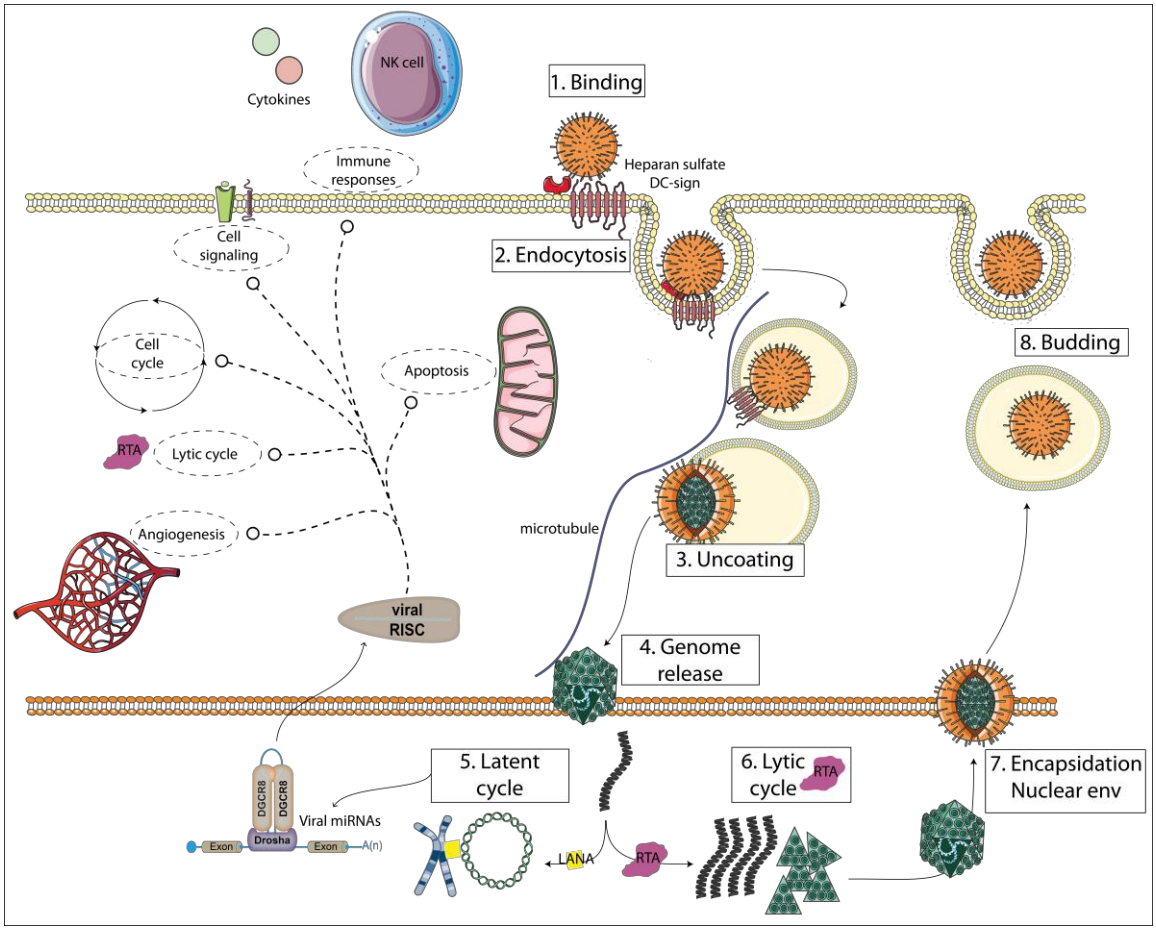


Figure 3. KSHV viral cycle & viral miRNA. KSHV recognizes its host cell through receptor binding. After endocytosis, viral envelop fuses with the endosome to release the capsid. Through microtubule transport, capsid is directed to the nucleus pore. KSHV is release into the nucleus. During the latent cycle, the viral genome is circularized and bound to host chromosomes via the viral protein LANA. Upon reactivation, viral genome and structural protein are strongly generated. Viral capsid formation takes place in the nucleus. Viral envelop is retrieve from the nucleus membrane. Finally, viral particle is release from the cell through budding. During the latency, the KSHV expresses multiple miRNAs which destabilized fundamental host biological processes to its advantages.

I.B.3 Latent cycle

Following primary infection, KSHV preferentially establishes a latent cycle. It is important to precise that latency differs from chronically or persistently infection in that no viral progeny is produced. During the latency, KSHV highly restricts its gene expression to limit immune exposure and enhance persistent host infection. To maintain its genome, the virus expresses 12 miRNA precursors and a handful of latent proteins (Hussein *et al.*, 2019). KSHV latency is associated with cancer, endothelial neoplasm (Kaposi sarcoma) and lymphoproliferative disorders (primary effusion lymphoma and multicentric Castleman's disease). Major latency transcript arises from specific genomic region. This latency region allows the coordinated expression of miRNAs and a handful of viral proteins namely the latency-associated nuclear antigen (LANA), vFLIP, vCyclin and Kaposins A-C. These viral protein act, among other functions, as respectively KSHV genome maintenance during cell division, NFκB signalling pathway control and cell cycle progression control.

I.C. KSHV's miRNA

KSHV expresses miRNAs from 12 different precursors expressed as a cluster within intronic regions of three different transcripts generated by alternative splicing (Cai and Cullen, 2006). Their mature forms are labelled miR-K12-1 to miR-K12-12 based on their genomic proximity to the K12 ORF. Ten of them from miR-K12-1 to miR-K12-9 and miR-K12-11 are part from the same intronic transcript (Figure 4). miR-K12-10 and miR-K12-12 are present within the ORF and the 3' UTR of the Kaposin A gene. Clustered KSHV pri-miRNA is controlled by two different latent promotors, which also control synthesis of LANA, v-cyclin and v-FLIP via alternative splicing.

KSHV miRNA targets have been essentially identified with experimental studies combining gene expression profiling upon individual miRNA expression (Suffert *et al.*, 2011; Gottwein, 2012). More global approaches, high-throughput sequencing cross-linking immunoprecipitation (HITS-CLIP) or cross-linking ligation and sequencing of hybrids (CLASH), have been used recently to go deeper into KSHV's miRNA targetome (Dölken *et al.*, 2010; Haecker and Renne, 2014; Gay *et al.*, 2018). KSHV's miRNA targets are diverse and numerous. They are able to control latent to lytic transition, cellular epigenetic machinery, cell cycle, apoptosis, innate immune responses, cytokines production (Table 1).

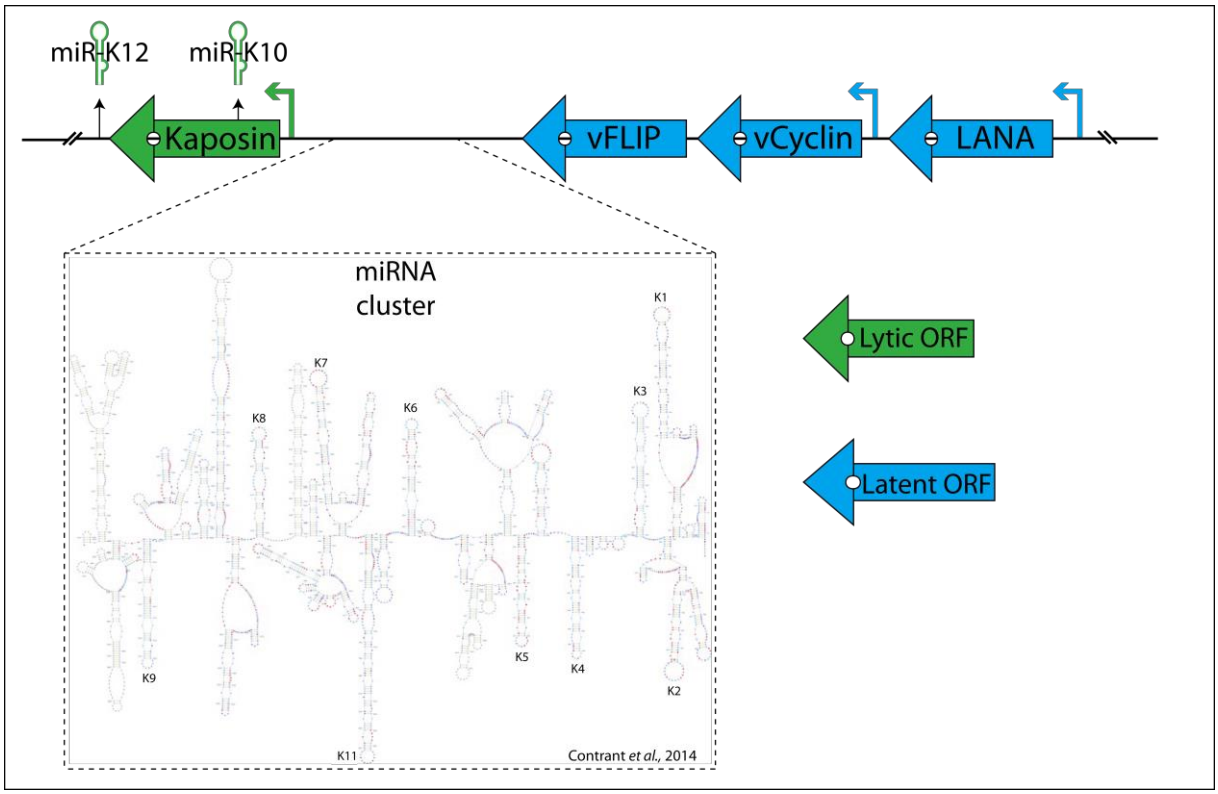


Figure 4. KSHV latent genomic locus and miRNA precursors. Relative position of KSHV miRNAs and latent genes. KSHV miRNA cluster is localized within the 3' UTR of LANA, vCyclin or vFLIP. Pre-miR-K12 & -K10 are respectively present in the 3' UTR and ORF of the Kaposin gene.

Table 1. KSHV miRNA targetome (non-exhaustive).

Target	miRNAs	Functional consequences	Reference
BACH1	K11	Increased viability under oxidative stress	(Gottwein *et al.*, 2007)
BCLAF1	K5	Inhibit caspase activity	(Ziegelbauer *et al.*, 2009)
CASP3	K1, K3, K4-3p	Apoptosis inhibition	(Suffert *et al.*, 2011)
CDKN1A/p21	K1	Oncogenesis	(Gottwein and Cullen, 2010)
C/EBP	K11, K3, K7	Inhibition cytokine secretion	(Boss *et al.*, 2011)
GRK2	K3	Promote angiogenesis	(Li, Jia, *et al.*, 2016)
IKBKE	K11	Interferon signalling inhibition	(Liang *et al.*, 2011)
IRAK1	K9	Decreased TLR activity	(Abend *et al.*, 2012)
MAF1	K1, K6-5p, K11	Endothelial cell reprogramming	(Hansen *et al.*, 2010)
MICB	K7	Favour immune evasion	(Nachmani *et al.*, 2009)
MYD88	K5	Decreased TLR activity	(Abend *et al.*, 2012)
NFIB	K3	Promote latency	(C.-C. Lu *et al.*, 2010)
NFKB1A	K1	Promote latency	(Lei *et al.*, 2010)
RBL2	K4-5p	de-repression of DNA methyl transferase	(F. Lu *et al.*, 2010)
SH3BGR	K6-3p	Promote angiogenesis	(Li, Yan, *et al.*, 2016)
SMAD5	K11	Favour cell growth	(Liu *et al.*, 2012)
TGFBRII	K10	Favour cell growth	(Lei *et al.*, 2012)
THBS1	K1, K3-3p, K6-3p, K11	Pro-angiogenic	(Samols *et al.*, 2007)
TWEAKR	K10	Reduced inflammatory response	(Abend *et al.*, 2010)
KSHV RTA	K9-5p, K5, K7	Inhibition of viral lytic cycle	(Bellare and Ganem, 2009)

KSHV miRNAs are fundamental for the viral cycle and required for pathology induction through complex regulation network. Expression of several KSHV latent genes and miRNAs in B cells predisposed humanized mice to lymphomas and hyperplasia (Gottwein *et al.*, 2007; Ballon *et al.*, 2011; Boss *et al.*, 2011). In addition, it appears that a miRNA deletion mutant KSHV failed to transform primary cells (Moody *et al.*, 2013). Here, I want to highlight some of the oncogenic properties of the KSHV miRNAs.

As miRNAs are mainly expressed during the latency, it is thus fundamental for the virus to control and repress lytic gene expression. First, miR-K9-5p and miR-K7 are modulators of RTA, the master regulator of the latent/lytic (Bellare and Ganem, 2009). Other KSHV miRNAs indirectly maintain viral latency such as miR-K4-5p which targets retinoblastoma-like 2 protein (RBL2) inducing elevated methylation of the RTA promotor (F. Lu *et al.*, 2010). In addition, by interfering with the host immune system, KSHV's miRNAs also contribute to the maintenance of the latency. miR-K7 directly target MICB mRNA, a stress-induced ligand of the natural killer. This inhibition greatly reduces the destruction of the infected cell and favours dysplasia apparition (Nachmani *et al.*, 2009).

Other KSHV's miRNAs display oncogenic properties such as angiogenesis induction, apoptosis inhibition and host cell cycle control. Thrombospondin1 (THBS1), an anti-angiogenic and antiproliferative protein is down-regulated by multiple KSHV miRNAs (miR-K1, -K3-3p, -K6-3p and -K11) that greatly contribute to angiogenesis (Samols *et al.*, 2007). Multiple cell signalling pathway, JAK-STAT or AKT-PI3K, involving in cell growth are stimulated indirectly by the viral miRNAs, respectively miR-K6-3p (Li, Yan, *et al.*, 2016) and miR-K3 (Li, Jia, *et al.*, 2016), to induce cell proliferation. Cell survival is increased by miR-K1 through down regulation of p21, which is a key inducer of the cell cycle arrest (Gottwein and Cullen, 2010). Apoptosis is also inhibited by the via CASP3 downregulation mediated by miR-K1, -K3 and -K4-3p (Suffert *et al.*, 2011). Finally, even the host cell metabolism is modulated by the KSHV miRNAs. Metabolic properties of KSHV-infected cell look like metabolic hallmarks of cancer. Viral miRNAs favour Warbug effect, which corresponds to an increase in metabolism based on glycolysis metabolism instead of oxidative phosphorylation leading to an increased cell proliferation (Yogev *et al.*, 2014).

It is now accepted that the complex regulation network set up by the KSHV through its miRNAs play significant roles in infection and malignancies. It is now fundamental to address the individual miRNA expression regulation and its dynamic changes to better understand KSHV pathological induction. Concerning my project, the most important thing is not the miRNA targetome but rather their mature form differential accumulation. Indeed, KSHV mature miRNAs, even though present on the same pri-miRNA, are expressed at different level between them and also depending on the cell line (Umbach and Cullen, 2010; Contrant *et al.*, 2014). Being able to express each of its miRNAs at the required level depending on the host cell could provide fundamental advantages for the virus. As mentioned in the biogenesis part (see I.A.2), structural and sequence features of miRNA precursors correspond to the first regulation layer of miRNA precursor maturation. In the case of the KSHV, there is an absence of correlation between structure and sequence features for some of the miRNA precursors and their mature form abundance (Contrant *et al.*, 2014). Thus, it is likely that additional maturation cofactors can be involved such as the ones discussed in part I.A.2. Finally, since miRNA biogenesis is a co-transcriptional process (Morlando *et al.*, 2008), kinetic and ordered miRNA

maturation can occur. Taken together, specific cofactors and ordered maturation add another layer of post-transcriptional regulation to explain the absence of correlation between the miRNA precursors features and their corresponding abundance.

I.D. RNA binding proteins

By definition, RBPs are able to dynamically interact with RNA molecules to form RNPs (ribonucleoprotein). RNA molecules are rarely devoid of protein partners. RNPs form as soon as an RNA oligonucleotide exit the polymerase. RNPs are critical for all cellular processes and participate actively to RNA processing, transport, localization, decay, storage, folding and more importantly post-transcriptional gene regulation. Mammalian cells express more than a thousand of RBPs (Castello *et al.*, 2016). Through a combination of RNA binding domain (RBD), RBPs are able to interact with a large diversity of RNA sequence and structure (Lunde *et al.*, 2007). To add even more complexity, more than half of the mammalians RBPs do not possess canonical RBDs (Castello *et al.*, 2016; Perez-Perri *et al.*, 2018). In addition, RNPs can form membrane-less granules microscopically visible in different cell lines and subcellular localization such as Cajal bodies, paraspeckles, p-bodies and stress granules (Van Treeck and Parker, 2018). All these studies highlight a previously “hidden” RNPs complexity. It is now accepted that RNA and RBP form a functional crosstalk in which both molecules can lead the fate of the other (Hentze *et al.*, 2018).

I.D.1 RNA binding domains: sequence, structure and context preferences

Recent studies clearly demonstrate that the number of RBDs was under-estimated (Calabretta and Richard, 2015; Järvelin *et al.*, 2016). Few RBDs have been well described such as RNA recognition motif (RRM), K homology domain (KH), double stranded RNA binding domain (dsRBD), zinc finger (ZnF), S1 domain, Piwi/Argonaute/Zwille (PAZ), PIWI domain, Pumilio domain and pentatricopeptide repeat (PPR). They all display some specificity concerning their RNA binding properties, but they usually display RNA sequence or structural preferences. For example, RNA-RBD interaction is mediated through stacking electrostatics and hydrogen bonding (RRM), through hydrophobic interactions between non-aromatic residues and the bases (KH), through shape-specific recognition (dsRBD), through electrostatic interactions between amino acid side chain and RNA backbone (ZnF), through stacking interaction between bases and aromatic residues (S1), via ssRNA 3’ overhang recognition (PAZ).

However, the great diversity of RBP functions do not seem to correlate with an incredible diversity of RBDs. RBPs generally possess multiple RBDs. Combining identical or different RBDs can provide a great versatility for the RBPs (Lunde *et al.*, 2007). Indeed, each individual domain is as important as the way they are arranged to each other to provide versatile function to the RBP. For example, multiple domain can provide a given RBP the ability to bind distant RNA sequence within the same transcript, or two different transcripts. Those different interaction could impact RNA and protein fold as well as increasing or decreasing RNPs function and diversity. Interestingly, different RBPs possessing different RBDs are able to bind similar RNA motifs which are generally composed of one or two different RNA bases (e.g. A-U) of 5-6mer long (Dominguez *et al.*, 2018). In addition, RBPs with similar RBDs organization

often binds different transcript location, suggesting the importance of the upstream and downstream structural (Zhang *et al.*, 2017) or sequence context (Damianov *et al.*, 2016) to drive the RNPs formation (Dominguez *et al.*, 2018). All of these diverse, dynamic and complex interactions between RNA and RBPs can lead to the formation of microscopic complexes grouped under the generic term: granules.

I.D.2 RNPs & membrane-less organelles

All biological processes are generally compartmentalized to favour the spatiotemporal regulation and efficiency of biological processes. This compartmentalization, fundamental for all biological mechanisms, can occur with phospholipid membrane delimitation (nucleus, vesicles, reticulum…) or with membrane-less droplet like structure called granules (stress granules, paraspeckles, p-bodies…). Those granules are composed of RNA and proteins and possess various biological roles. They display complex multi-layered dynamic organization that can be defined as "oil droplets in water" structure or microscopic RNPs complex or more precisely as "liquid-like states of intracellular matter" (Shin and Brangwynne, 2017). Liquid-liquid phase transition based on chemicals and physical molecules properties (hydrophobia, polarity, concentration, temperature, pH, thermodynamic parameters…) are the driving force explaining the dynamism and generation of these liquid-like droplets (Alberti, 2017). Membrane-less organelles formation require multivalent protein-protein, protein-RNA and RNA-RNA interaction. Protein low sequence complexity or intrinsically disordered protein regions are typically enriched in those liquid-like droplets. An important number of RBPs could be potentially involved in liquid-droplet formation. ROA1 for example, has been characterized to be part of the stress granules which correspond to mRNPs stalled in translation initiation (Buchan, 2014). Mutations in the low complexity region of ROA1 as well as associated RNA molecules (Van Treeck and Parker, 2018), involved in the liquid-like droplet formation, are responsible of multiple degenerative disorder such as amyotrophic lateral sclerosis (Molliex *et al.*, 2015). Those liquid-like droplets act as organizational hub, molecule sequestration and reaction platform (Shin and Brangwynne, 2017). Recently, key component of nuclear granules paraspeckles, such as the NEAT1 lncRNA together with NONO and PSF RBPs, have been reported to recruit multiple RBPs (ILF2, ROA1…) as well as the microprocessor complex (Jiang *et al.*, 2017). This suggest the potential implication of liquid-like droplet in the global or specific regulation of miRNA biogenesis.

II. THESIS OBJECTIVES

Clustered viral miRNA precursors represent a good model to study the dynamic of an RNA molecule biogenesis. First, due to its dual tropism, KSHV has to face different cellular proteomes in endothelial and lymphocyte cells. Second, lymphocytes face multiple extracellular signalling from inflammation to cell division. Post-transcriptional regulation of specific miRNAs from a clustered pri-miRNA transcript allows KSHV to modulate several cellular pathways with different efficiency in order to increase its fitness depending of the microenvironment. The use of a miRNAs cluster therefore provides an interesting advantage for the virus, because it does not need to express proteins to act on cellular reprogramming, thus remaining hidden from viral sensors. As mentioned earlier, miRNA biogenesis is a complex and dynamic process involving multiple factors. However, core enzymatic complexes, such as the Microprocessor and DICER/TRBP, together with miRNA precursors structural features are not sufficient to explain differential maturation of specific miRNAs within a cluster. Notwithstanding transcriptional regulation and mature miRNA degradation, it seems that the first miRNA maturation step occurring in the nucleus is the main and limiting process responsible for differential abundance of mature miRNAs. Consequently, the main purpose of my project is to identify proteins interacting with specific miRNA precursors within KSHV pri-miRNA cluster that could explain the differential accumulation of its mature miRNAs.

To achieve this objective, I performed RNA affinity chromatography using different approaches for ten of the KSHV miRNA precursors (described in part III.A). Once potential candidates were identified, I started to validate their implication in miRNA biogenesis using either western blot analysis (part III.B.1.a) luciferase-based screen assay (part III.B.1.b), or northern blot and qRT-PCR experiment (part III.B.1.c). Then I proposed different tools and approaches to assess the functional implication of the candidates and continue the work on the produced candidates list (see part III.B.2.b).

The long-term objective of this project is to understand how these KSHV miRNAs are precisely expressed and if any one of them play a leading role in the associated pathology. The ultimate goal would be to identify potential therapeutic approaches based on KSHV miRNA control.

III. RESULTS

Results will be presented into two main sections which will be discussed in part IV. The first part III.A concerns the identification of individual KSHV miRNA precursors interactants. In this project, miRNA precursor interactants were identified with different RNA chromatography approaches which will be exposed, explained, explored and criticized. The second part III.B concerns the biological validation of the previously identified miRNA precursor interactants. Several tools have been set up to decipher the impact of these interactants on the specific maturation of miRNA precursors. Luciferase reporters, northern blot and reverse transcription followed by quantitative polymerase chain reaction (qRT-PCR) assays will also be exposed.

III.A. RNA affinity chromatography for identification of factors binding specific KSHV MIRNA

RNA chromatography assays that were performed in this work are based on different matrixes (agarose or magnetic streptavidin beads), different baits (stem-loop miRNA or chimeric stem-loop miRNA) and different elution methods (heat or DNase elution). In addition, RNA pulldowns were performed using nuclear extracts of different cell lines obtained by hypotonic shock and cytoplasmic membrane disruption. We used nuclear extracts to increase the proportion of proteins involved in the first miRNA maturation step that occur in the nucleus. As mentioned, three different B lymphocyte (BC-3, BCBL-1, DG75) and one endothelial (EaHy926) cell lines have been used. BC-3 and BCBL-1 cell lines are B lymphocytes naturally infected by the KSHV. They were both established from a human HIV and EBV negative patient with primary effusion lymphoma. The DG75 cell is also a B lymphocyte line non infected by the KSHV established from a human EBV negative patient with pleural effusion lymphoma. EaHy926 is an endothelial cell line established by fusing primary human umbilical vein cells with a thioguanine-resistant clone of A549 by exposure to polyethylene glycol (PEG). Since these cell lines are naturally infected with KSHV, it makes sense to perform RNA chromatography within their nuclear extracts as they represent relevant protein extract for the virus. In addition, these cell lines have been used to study the relative abundance of the different KSHV mature miRNAs. miRNA libraries sequencing revealed that KSHV mature miRNA display dramatic accumulation differences between cell lines (Contrant *et al.*, 2014). We also used heat or DNase treatment to elute the RNA after chromatography in order to determine the more specific approach. Finally, proteins bound during the chromatography were identified by liquid chromatography followed by tandem mass spectrometry (LC-MSMS) using spectral count or label free quantification technology. The RNA baits that were used correspond to individual stem-loop miRNA precursor (SL-miRNA), mimicking a DROSHA-DGCR8 substrate. More precisely, *in vitro* transcribed SL-miRNA resemble a canonical pre-miRNA with 20 nt flanking sequence. SL-miRNAs mimic ten out of twelve of the KSHV miRNA precursors present on the primary transcript (from kshv-miR-K12-1 to kshv-miR-K12-9 plus kshv-miR-K11). To facilitate the nomenclature, I used throughout the manuscript the following naming: SL-miRNA kshv-miR-K12-X will be referred to as SL-miRNA-KX or alternatively -KX. Similarly, the cellular miRNA hsa-let-7a-1 and hsa-miR-155 are respectively referred to as SL-miRNA-let7 or -let7 and SL-miRNA-155 or -155. It is essential to mention here that the

KSHV SL-miRNAs are synthetized individually and out of their cluster context thus inducing a potential misfolding and/or maturation defect. In addition, it is important that each chromatography possesses negative and positive control. Negative control will depend on the matrix and bait used and generally correspond to the bait without the SL-miRNA. It will serve to characterize the background noise of the chromatography and will allow the calculation of enrichment ratio (SL-miRNA bait / negative bait). The positive control always corresponds to the miRNA precursor hsa-let-7a-1 for which specific maturation cofactors such as LN28B, ROA1, FUBP2 (Heo *et al.*, 2008; Michlewski and Cáceres, 2010) have been extensively described and studied. Thus, if the chromatography worked, we should identify at least some of these proteins. In addition, all the produced SL-miRNA, mimicking miRNA precursors, are also supposed to interact with the microprocessor complex as well with its general cofactors: ILF2, ILF3, DHX9 (Gregory *et al.*, 2004; Sakamoto *et al.*, 2009). Consequently, mass spectrometry data will only be considered successful if these general and specific co-factors are enriched by all the SL-miRNAs and the SL-miRNA-let7 baits.

III.A.1 Agarose RNA-pulldown

The first RNA chromatography approach used was based on agarose beads covalently coupled with the SL-miRNA, heat eluted and analysed by spectral count after mass spectrometry. With this approach, SL-miRNA-let7, -K4 and -K11 have been used to realize the pulldown in triplicate using 4 different nuclear extracts coming from the three lymphocyte lines: BCBL-1, DG75, BC-3 and the endothelial line EaHy926. The purpose of this experiment is to identify different proteins binding SL-miRNAs depending on the cell line.

III.A.1.a SL-miRNAs production and quality

As in all affinity-chromatography based approaches, RNA pulldown efficiency depends on the bait quality. *In vitro* transcribed and gel purified SL-miRNAs were systematically visualized on a 12 % acrylamide 8 M urea gel (Figure 5.A) to check their integrity before their binding to beads. As can be seen in Figure 5.A SL-miRNAs display the expected length and almost no degradation products. SL-miRNA correct folding after denaturation and renaturation has not been assessed using chemical probing or RNase treatment. However, identification of dsRNA binding proteins (E2AK2, DSRAD, ILF2, ILF3) as well as ssRNA binding proteins (hnRNPs) following RNA-chromatography will strongly suggest a correct stem loop structure fold of the SL-miRNA bait. This will be the main purpose of the positive control SL-miRNA-let7.

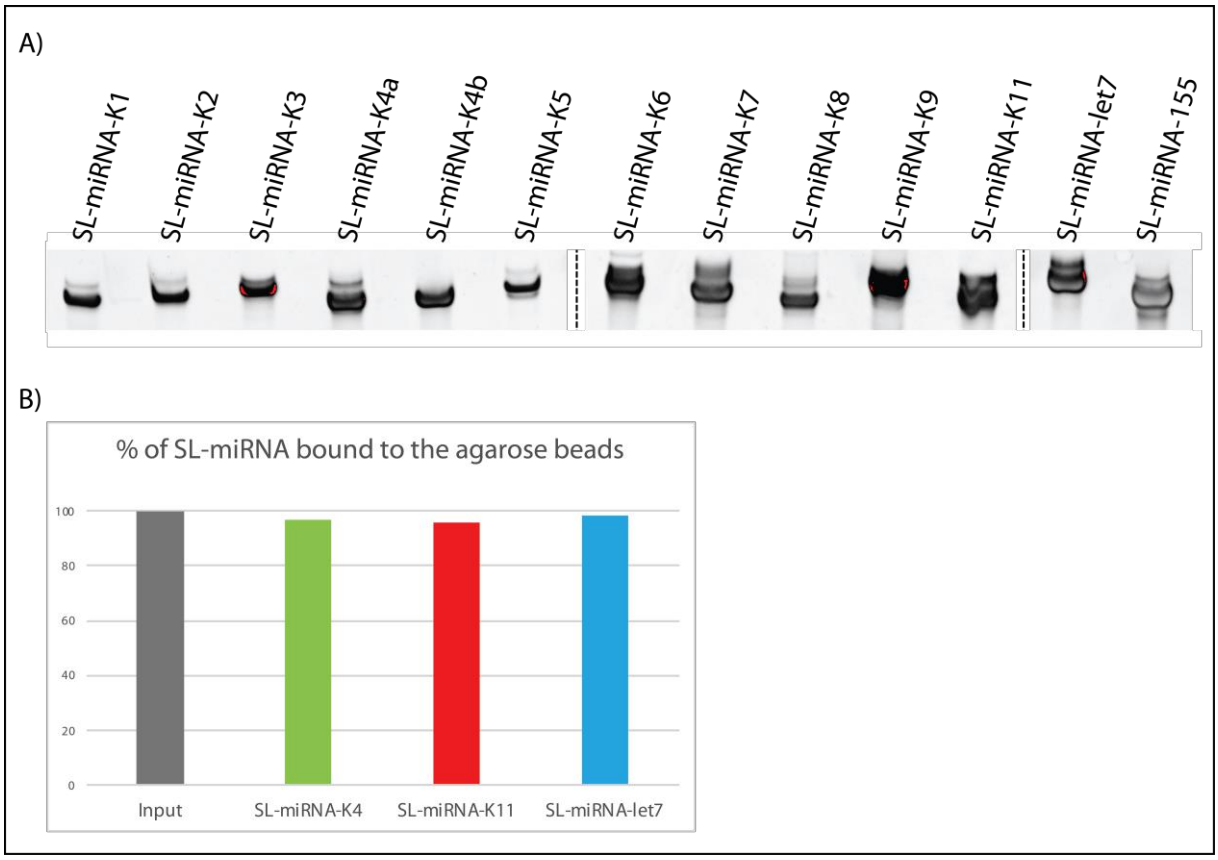


Figure 5. SL-miRNA production and beads attachment. A) Ethidium bromide staining of SL-miRNA in a 12 % acrylamide 8 M gel. Each band correspond to a specific SL-miRNA transcript. Average length of the SL-miRNA is around 100 nt. Black dashed lines represent different gels. B) Assessment of the binding efficiency of the SL-miRNAs on agarose beads using radioactive labelling. After a 5' radiolabelling and a 3' sugar oxidation steps, 30,000 cpm of hot SL-miRNAs are mixed with 500 pmol of cold SL-miRNAs to be incubated with the agarose beads. Radioactive signal within the non-retained fraction right after the binding incubation is measured by Cerenkov counting.

III.A.1.b SL-miRNAs beads binding

After denaturation and renaturation, SL-miRNAs are covalently linked to NH_2 groups present on the beads via their oxidized 3' ribose extremity. Overnight binding efficiency of the SL-miRNAs on the agarose beads have been characterized after 5' radioactive labelling. The 5' tri-phosphate of the *in vitro* SL-miRNAs have been dephosphorylated with an alkaline phosphatase treatment to allow a simple radiolabelling with a [γ-32P]dATP and a T4 polynucleotide kinase. Radiolabelled SL-miRNAs are then gel purified and passively eluted from the gel. Radioactive [γ-32P] signals are measured by Cerenkov counting. Here, 500 pmol of cold SL-miRNA were mixed with 30,000 cpm of radiolabelled SL-miRNA to perform the binding step with the agarose beads during an overnight incubation at 4°C. After the binding step, radioactive signal of the non-retained fraction is measured to determine the binding efficiency, which was above 95 % for all three SL-miRNAs (Figure 5.B). Once the baits have been generated and controlled, nuclear extract need to be produced.

III.A.1.c Nuclear extract quality

As this project only focuses on the nuclear processing step mediated by DROSHA and DGCR8, we set up a nuclear extract preparation protocol. We used hypotonic shock for the nuclear extraction and monitored the cell phenotype at each step by microscopy (Figure 6.A.B.C). Here, one representative example of nuclear extract production is shown. All nuclear extracts were verified in the same way. DAPI staining was used to visualize nucleus integrity until their final lysis. B lymphocytes possess a diameter of about 18 µm and a nuclear-cytoplasmic ratio of 0.7. Hypotonic shock swells the cells without lysis until a diameter of ~21 µm, inducing drop of the nuclear-cytoplasmic ratio around 0.6. Following a mechanical cytoplasmic membrane disruption with a Douncer homogenizer, nucleus are purified and lysed by sonication in an ice water bath. We then determined the quality of cell fractionation by western blot analysis of RPS21, a ribosomal protein located in the cytoplasm, and Histone H3, a core component protein of nucleosome (Figure 6.D). Overall, the results indicate that the cell fractionation protocol works fine. Now that the baits and the nuclear extract are available, the affinity chromatography can be performed.

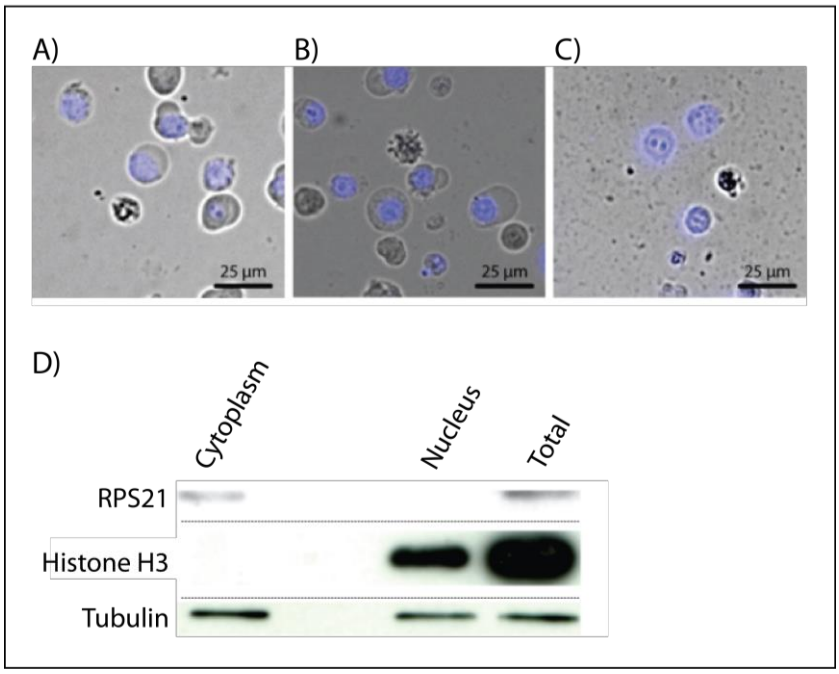


Figure 6. Nuclear extract preparation. A, B, C) Microscopic observation of B cells upon nuclear extraction stained with DAPI. A) Lymphocytes prior to the hypotonic shock. B) Lymphocytes after incubation within a hypotonic buffer. C) Nucleus after the cytoplasmic membrane disruption. D) Western blot analysis of the nuclear extract. Tubulin signal act as a loading control.

III.A.1.d Agarose RNA-pulldown

RNA pulldowns were performed with the SL-miRNA-K4, -K11, -let7 and one negative control corresponding to the agarose beads only. All baits were heat eluted after performing RNA-pulldowns with the aforementioned 4 different nuclear extracts. Briefly, RNA pulldowns are carried out with a nuclear extract of 50.10^6 cells during 45 min at room temperature followed

by a heat elution at 95°C during 5 min in Laemmli buffer. Agarose-based RNA chromatography were performed in the early stages of the project. During this period, I had a lot of troubles to obtain samples allowing a good mass spectrometry analysis. Indeed, agarose beads are pretty fragile and tend to be degraded upon temperature variation and heat elution. Agarose degradation products are deleterious for each mass spectrometry step, from HPLC to peptide (mass on charge) assessment. Even though, I managed to obtain few samples clean enough to perform the identification and the spectral count relative quantification using a TripleTOF mass spectrometer. As specified earlier, beads only baits are used to generate enrichment ratio. It is important to precise that spectral count technology is far from being a relevant quantification method especially since some agarose polymers degradation products are present. Thus, RNA chromatography based on agarose bead and spectral count identification should be considered more a "qualitative data" as you will see in the next section. This is one of the reasons why new RNA chromatography approach based on magnetic beads, DNase elution and label free quantification (Cox *et al.*, 2014) have been set up in order to generate better candidates list in terms of experimental specificity and sensibility (see III.A.3).

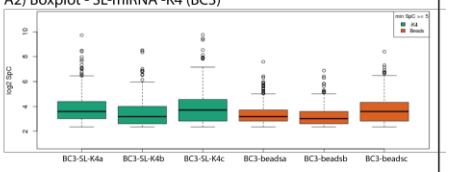
III.A.1.e Mass spectrometry data exploration

As mentioned above, statistical differential analysis is difficult to perform since agarose degradation product, which greatly limit the spectral counting step and overall eluates comparison, are present and not homogeneously distributed among all the agarose pulldown eluates. Nevertheless, I will describe the mass spectrometry data using descriptive and differential analysis before looking at the positive bait (SL-miRNA-let7) proteins identification. The obtained results are presented for each SL-miRNA. First, I will briefly present and discuss overall dataset features using summary table, boxplot, density-plot and principal component analysis (PCA) (Mcdonald; Abdi and Williams, 2010) (SL-miRNA-K4: Figure 8, Figure 9; SL-miRNA-K11: Figure 10, Figure 11; SL-miRNA-let7: Figure 12, Figure 13). Then, I will present and discuss the differential analysis, based on the negative-binomial model (Li *et al.*, 2018). This model allows for data overdispersion and could be indicated where a significant source of biological variability is expected. It also requires the estimation of two parameters, but here its implementation was entirely based on the solution provided by the edgeR package (Chen *et al.*), which includes empirical Bayes method to share information among features, and thus may be employed even when the number of replicates is as low as two. The negative binomial is downward limited, when no overdispersion is observed, by the Poisson distribution. The p-values are adjusted with false discovery rate (FDR) control by the Benjamini-Hochberg method (Burger, 2018). All the enriched datasets will be displayed as a table and Venn diagram. The results indicate that, overall, the quality of agarose beads and heat eluted chromatography data was low.

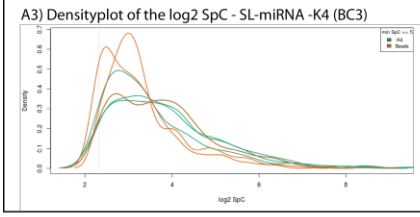
A1) Protein and SpC summary - SL-miRNA -K4 (BC3)

	Nb proteins	Total SpC	max SpC count
BC3b_SL-K4	671	10226	845
BC3c_SL-K4	478	6250	365
BC3d_SL-K4	1219	16805	865
BC3b_beads	500	4051	192
BC3c_beads	379	2737	118
BC3d_beads	1178	11719	339

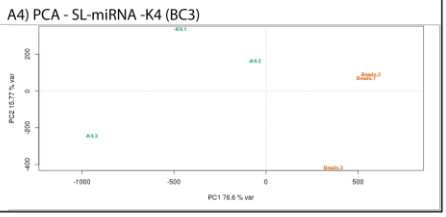
A2) Boxplot - SL-miRNA -K4 (BC3)



A3) Densityplot of the log2 SpC - SL-miRNA -K4 (BC3)



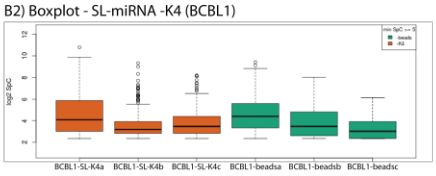
A4) PCA - SL-miRNA -K4 (BC3)



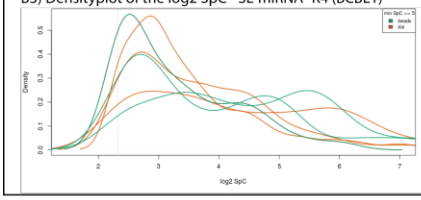
B1) Protein and SpC summary - SL-miRNA -K4 (BCBL1)

	Nb proteins	Total SpC	max SpC count
BCBL1a_SL-K4	815	24360	1783
BCBL1b_SL-K4	844	9148	644
BCBL1c_SL-K4	731	7251	297
BCBL1a_beads	640	13692	685
BCBL1b_beads	449	6658	257
BCBL1c_beads	557	3218	70

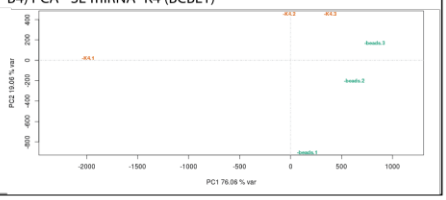
B2) Boxplot - SL-miRNA -K4 (BCBL1)



B3) Densityplot of the log2 SpC - SL-miRNA -K4 (BCBL1)



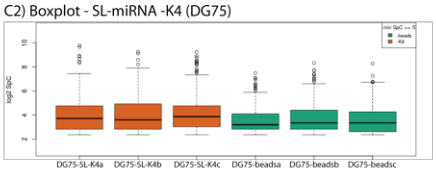
B4) PCA - SL-miRNA -K4 (BCBL1)



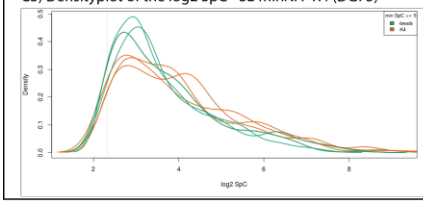
C1) Protein and SpC summary - SL-miRNA -K4 (DG75)

	Nb proteins	Total SpC	max SpC count
DG75a_SL-K4	853	15414	875
DG75b_SL-K4	702	11262	607
DG75c_SL-K4	1210	22428	597
DG75a_beads	613	5989	179
DG75b_beads	551	6948	319
DG75c_beads	708	6691	306

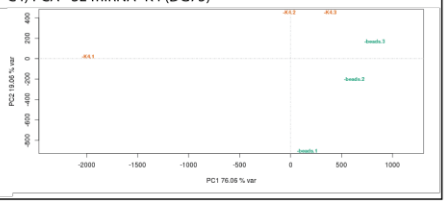
C2) Boxplot - SL-miRNA -K4 (DG75)



C3) Densityplot of the log2 SpC - SL-miRNA -K4 (DG75)



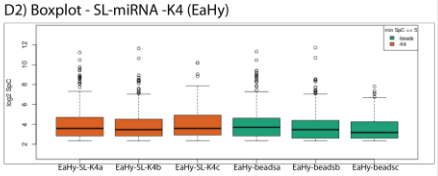
C4) PCA - SL-miRNA -K4 (DG75)



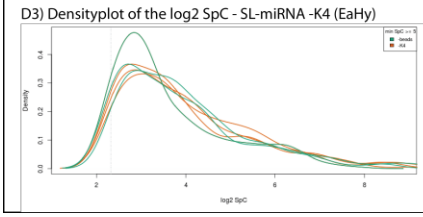
D1) Protein and SpC summary - SL-miRNA -K4 (EaHy)

	Nb proteins	Total SpC	max SpC count
EaHya_SL-K4	922	22978	2374
EaHyb_SL-K4	639	16035	3173
EaHyc_SL-K4	818	15119	1168
EaHya_beads	862	21568	2541
EaHyb_beads	638	15515	3333
EaHyc_beads	622	6508	225

D2) Boxplot - SL-miRNA -K4 (EaHy)



D3) Densityplot of the log2 SpC - SL-miRNA -K4 (EaHy)



D4) PCA - SL-miRNA -K4 (EaHy)

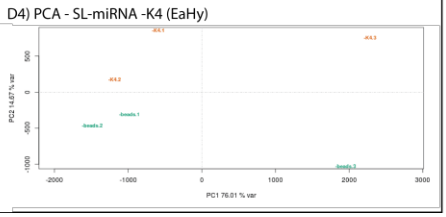


Figure 7. Descriptive analysis of the four chromatography datasets with the SL-miRNA-K4 bait. Table summary (1), boxplot (2), density-plot (3) and PCA (4) are displayed for all the nuclear extract used. (A) BC-3 nuclear extract -K4 chromatography. B) BCBL-1 nuclear extract -K4 chromatography. C) DG75 nuclear extract -K4 chromatography. D) EaHy926 nuclear extract -K4 chromatography.

A)

IDs	SL-K4_BC3	SL-K4_BCBL1	SL-K4_DG75	SL-K4_EaHy
ABCF1_HUMAN	Enriched	NA	NA	NA
CIRBP_HUMAN	NA	NA	NA	Enriched
DDX1_HUMAN	NA	NA	Enriched	NA
DHX9_HUMAN	Enriched	NA	NA	Enriched
DRG1_HUMAN	Enriched	Enriched	Enriched	NA
DSRAD_HUMAN	NA	Enriched	Enriched	Enriched
E2AK2_HUMAN	Enriched	Enriched	Enriched	Enriched
EIF2A_HUMAN	Enriched	NA	NA	NA
FUBP3_HUMAN	NA	NA	NA	Enriched
GNL3_HUMAN	NA	NA	NA	Enriched
GTPB8_HUMAN	Enriched	Enriched	NA	Enriched
HNRL1_HUMAN	NA	Enriched	NA	Enriched
IF2B_HUMAN	Enriched	NA	NA	NA
IF2P_HUMAN	NA	NA	Enriched	NA
ILF2_HUMAN	Enriched	NA	NA	Enriched
ILF3_HUMAN	Enriched	Enriched	Enriched	Enriched
JUNB_HUMAN	Enriched	NA	NA	NA
MCES_HUMAN	Enriched	NA	NA	NA
NSUN6_HUMAN	NA	NA	NA	Enriched
NTPCR_HUMAN	NA	NA	NA	Enriched
PA2G4_HUMAN	Enriched	Enriched	Enriched	NA
PRKRA_HUMAN	NA	Enriched	NA	Enriched
RAN_HUMAN	Enriched	Enriched	Enriched	Enriched
RANG_HUMAN	Enriched	Enriched	Enriched	Enriched
RBMS2_HUMAN	NA	NA	NA	Enriched
RBP2_HUMAN	NA	NA	Enriched	Enriched
RED1_HUMAN	NA	NA	NA	Enriched
RS20_HUMAN	Enriched	NA	NA	NA
RS7_HUMAN	NA	NA	Enriched	NA
STAU1_HUMAN	Enriched	Enriched	NA	Enriched
STAU2_HUMAN	NA	NA	NA	Enriched
STRBP_HUMAN	Enriched	Enriched	Enriched	Enriched
THYN1_HUMAN	Enriched	NA	NA	NA
TIA1_HUMAN	NA	NA	Enriched	NA
TIAR_HUMAN	NA	NA	Enriched	NA
WIBG_HUMAN	Enriched	Enriched	Enriched	Enriched
ZC3HF_HUMAN	Enriched	Enriched	NA	NA
ZCCHV_HUMAN	Enriched	Enriched	NA	Enriched
ZFR_HUMAN	Enriched	Enriched	Enriched	Enriched
ZMAT3_HUMAN	NA	NA	NA	Enriched
ZN346_HUMAN	Enriched	Enriched	Enriched	Enriched

B)

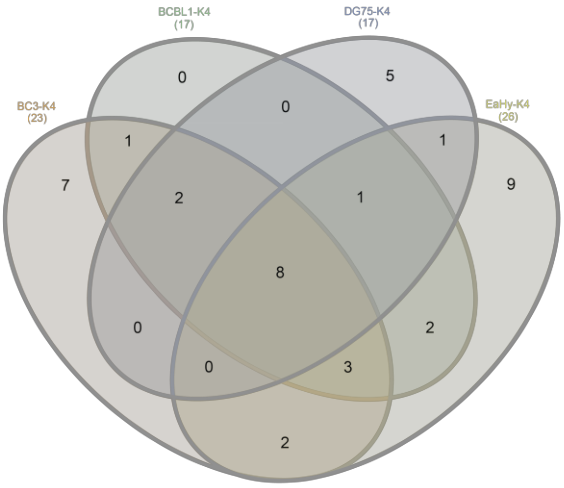
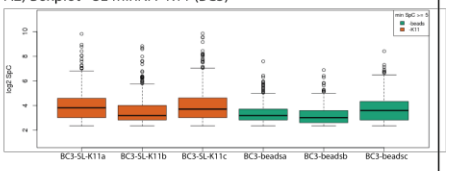


Figure 8. SL-miRNA-K4 significantly (p.value<0.05) enriched dataset by the 4 different cell lines. A) Merged list of the significantly enriched proteins. B) Venn diagram comparison of the significantly enriched proteins.

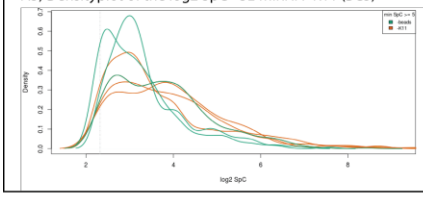
A1) Protein and SpC summary - SL-miRNA -K11 (BC3)

	Nb proteins	Total SpC	max SpC count
BC3b_SL-K11	731	12124	897
BC3c_SL-K11	531	7538	453
BC3d_SL-K11	1349	21987	919
BC3b_beads	500	4051	192
BC3c_beads	379	2737	118
BC3d_beads	1178	11719	339

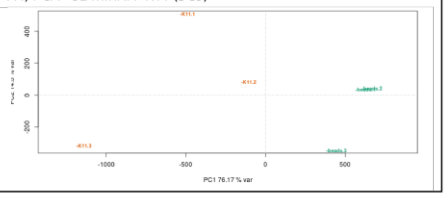
A2) Boxplot - SL-miRNA -K11 (BC3)



A3) Densityplot of the log2 SpC - SL-miRNA -K11 (BC3)



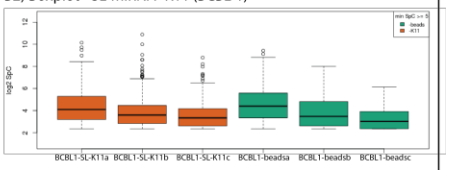
A4) PCA - SL-miRNA -K11 (BC3)



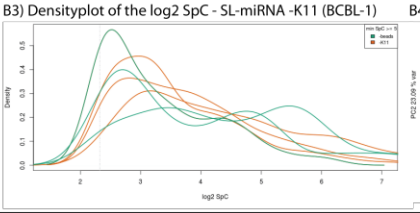
B1) Protein and SpC summary - SL-miRNA -K11 (BCBL-1)

	Nb proteins	Total SpC	max SpC count
BCBL1a_SL-K11	810	22310	1131
BCBL1b_SL-K11	1249	19539	1890
BCBL1c_SL-K11	867	8783	442
BCBL1a_beads	640	13692	685
BCBL1b_beads	449	6658	257
BCBL1c_beads	557	3218	70

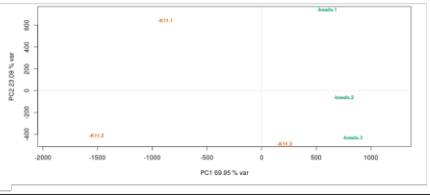
B2) Boxplot - SL-miRNA -K11 (BCBL-1)



B3) Densityplot of the log2 SpC - SL-miRNA -K11 (BCBL-1)



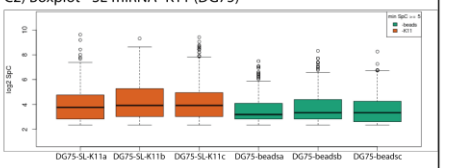
B4) PCA - SL-miRNA -K11 (BCBL-1)



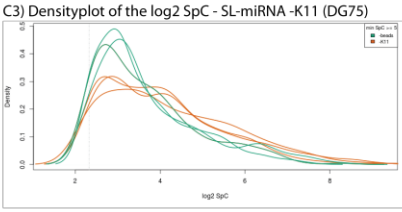
C1) Protein and SpC summary - SL-miRNA -K11 (DG75)

	Nb proteins	Total SpC	max SpC count
DG75a_SL-K11	850	13459	789
DG75b_SL-K11	603	11403	635
DG75c_SL-K11	1137	23073	683
DG75a_beads	613	5989	179
DG75b_beads	551	6948	319
DG75c_beads	708	6691	306

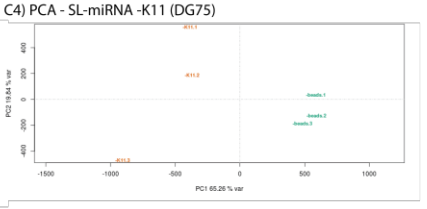
C2) Boxplot - SL-miRNA -K11 (DG75)



C3) Densityplot of the log2 SpC - SL-miRNA -K11 (DG75)



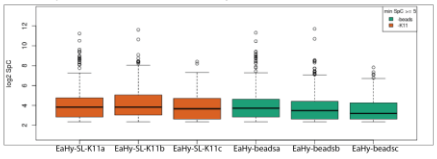
C4) PCA - SL-miRNA -K11 (DG75)



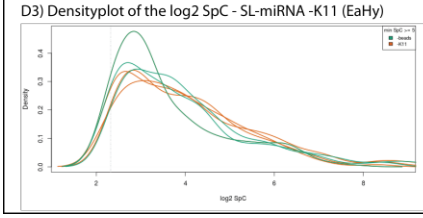
D1) Protein and SpC summary - SL-miRNA -K11 (EaHy)

	Nb proteins	Total SpC	max SpC count
EaHya_SL-K11	942	23130	2410
EaHyb_SL-K11	703	19239	3132
EaHyc_SL-K11	601	7043	339
EaHya_beads	862	21568	2541
EaHyb_beads	638	15515	3333
EaHyc_beads	622	6508	225

D2) Boxplot - SL-miRNA -K11 (EaHy)



D3) Densityplot of the log2 SpC - SL-miRNA -K11 (EaHy)



D4) PCA - SL-miRNA -K11 (EaHy)

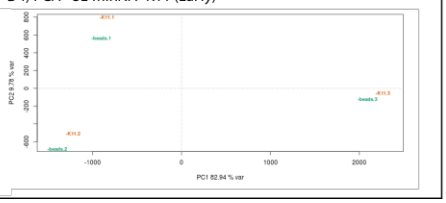


Figure 9. Descriptive analysis of the four chromatography datasets with the SL-miRNA-K11 bait. Table summary (1), boxplot (2), density-plot (3) and PCA (4) are displayed for all the nuclear extract used. (A) BC-3 nuclear extract -K11 chromatography. B) BCBL-1 nuclear extract -K11 chromatography. C) DG75 nuclear extract -K11 chromatography. D) EaHy926 nuclear extract -K11 chromatography.

A)

IDs	SL-K11_BC3	SL-K11_BCBL1	SL-K11_DG75	SL-K11_EaHy
CELF1_HUMAN	NA	Enriched	NA	Enriched
CIRBP_HUMAN	NA	NA	Enriched	Enriched
CPSF5_HUMAN	NA	Enriched	Enriched	Enriched
CPSF6_HUMAN	NA	Enriched	NA	NA
CPSF7_HUMAN	NA	Enriched	NA	NA
DAZP1_HUMAN	NA	Enriched	NA	NA
DDX1_HUMAN	NA	NA	Enriched	NA
DHX30_HUMAN	NA	Enriched	NA	NA
DHX36_HUMAN	NA	NA	NA	Enriched
DHX9_HUMAN	NA	Enriched	Enriched	Enriched
DRG1_HUMAN	NA	Enriched	Enriched	NA
DSRAD_HUMAN	NA	NA	Enriched	Enriched
E2AK2_HUMAN	Enriched	Enriched	Enriched	Enriched
EWS_HUMAN	NA	NA	NA	Enriched
FA98B_HUMAN	NA	NA	Enriched	NA
FUBP3_HUMAN	NA	Enriched	NA	Enriched
GNL3_HUMAN	NA	NA	Enriched	Enriched
GRSF1_HUMAN	NA	Enriched	Enriched	Enriched
HBS1L_HUMAN	NA	NA	Enriched	Enriched
HNRL1_HUMAN	NA	Enriched	Enriched	Enriched
IF2G_HUMAN	NA	NA	Enriched	NA
IF2GL_HUMAN	NA	NA	Enriched	NA
ILF2_HUMAN	Enriched	NA	NA	Enriched
ILF3_HUMAN	Enriched	Enriched	Enriched	Enriched
LA_HUMAN	NA	Enriched	NA	Enriched
LN28B_HUMAN	NA	Enriched	NA	NA
MBNL1_HUMAN	NA	Enriched	NA	NA
MCES_HUMAN	NA	Enriched	NA	NA
MET16_HUMAN	NA	NA	NA	Enriched
PRKRA_HUMAN	NA	Enriched	NA	NA
RAN_HUMAN	NA	NA	Enriched	NA
RANG_HUMAN	NA	NA	Enriched	Enriched
RBMS1_HUMAN	NA	NA	NA	Enriched
RBMS2_HUMAN	NA	NA	NA	Enriched
RBMS3_HUMAN	NA	NA	NA	Enriched
RBP56_HUMAN	NA	NA	Enriched	Enriched
RECQ1_HUMAN	NA	Enriched	NA	NA
RED1_HUMAN	NA	NA	NA	Enriched
RENT1_HUMAN	NA	NA	NA	Enriched
RL22L_HUMAN	NA	NA	Enriched	NA
RS17_HUMAN	NA	NA	Enriched	NA
RS17L_HUMAN	NA	NA	Enriched	NA
RS20_HUMAN	NA	NA	Enriched	NA
RS7_HUMAN	NA	NA	Enriched	NA
RTCB_HUMAN	NA	NA	Enriched	NA
STAU1_HUMAN	NA	Enriched	NA	Enriched
STAU2_HUMAN	NA	NA	NA	Enriched
STRBP_HUMAN	Enriched	Enriched	Enriched	Enriched
SYHC_HUMAN	NA	Enriched	NA	NA
SYHM_HUMAN	NA	Enriched	NA	NA
SYNC_HUMAN	NA	NA	Enriched	Enriched
SYTM_HUMAN	NA	Enriched	NA	NA
TFB2M_HUMAN	NA	NA	NA	Enriched
TIA1_HUMAN	NA	Enriched	Enriched	NA
TIAR_HUMAN	NA	Enriched	Enriched	NA
TRI25_HUMAN	NA	Enriched	Enriched	Enriched
WIBG_HUMAN	NA	NA	Enriched	NA
ZCCHV_HUMAN	NA	Enriched	Enriched	Enriched
ZFR_HUMAN	Enriched	Enriched	Enriched	Enriched
ZN346_HUMAN	NA	Enriched	Enriched	Enriched

B)

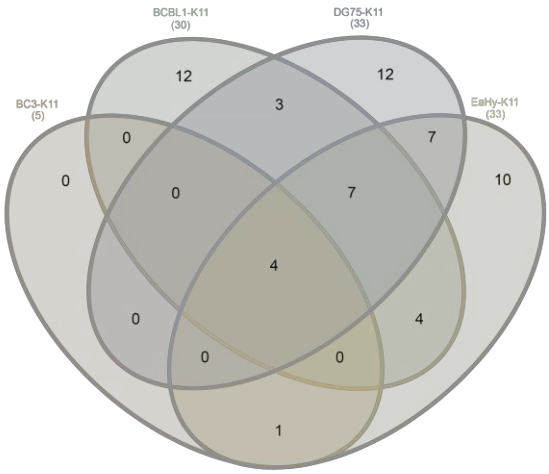
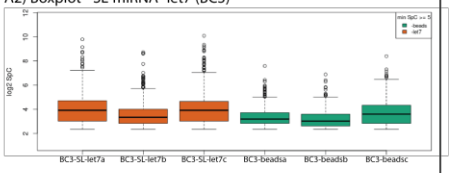


Figure 10. SL-miRNA-K11 significantly (p.value<0.05) enriched dataset in the 4 different cell lines. A) Merged list of the significantly enriched proteins. B) Venn diagram comparison of the significantly enriched proteins.

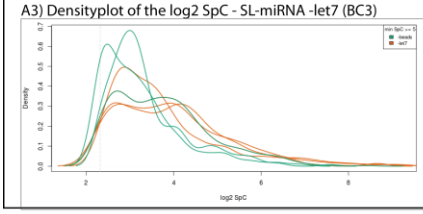
A1) Protein and SpC summary - SL-miRNA -let7 (BC3)

	Nb proteins	Total SpC	max SpC count
BC3b_SL-let7	813	14440	891
BC3c_SL-let7	567	7770	424
BC3d_SL-let7	1475	25181	1092
BC3b_beads	500	4051	192
BC3c_beads	379	2737	118
BC3d_beads	1178	11719	339

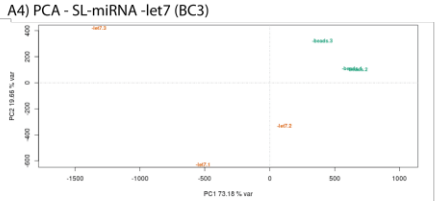
A2) Boxplot - SL-miRNA -let7 (BC3)



A3) Densityplot of the log2 SpC - SL-miRNA -let7 (BC3)



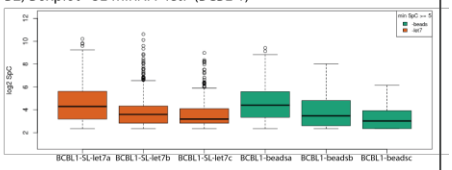
A4) PCA - SL-miRNA -let7 (BC3)



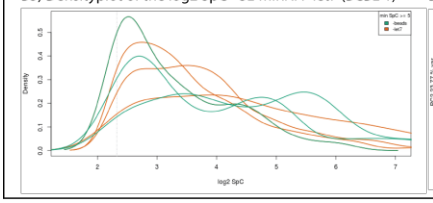
B1) Protein and SpC summary - SL-miRNA -let7 (BCBL-1)

	Nb proteins	Total SpC	max SpC count
BCBL1a_SL-let7	725	22940	1186
BCBL1b_SL-let7	1345	22002	1570
BCBL1c_SL-let7	943	10547	505
BCBL1a_beads	640	13692	685
BCBL1b_beads	449	6658	257
BCBL1c_beads	557	3218	70

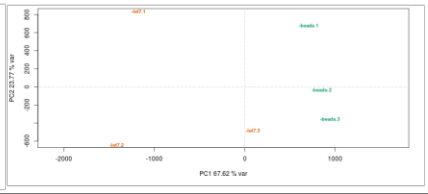
B2) Boxplot - SL-miRNA -let7 (BCBL-1)



B3) Densityplot of the log2 SpC - SL-miRNA -let7 (BCBL-1)



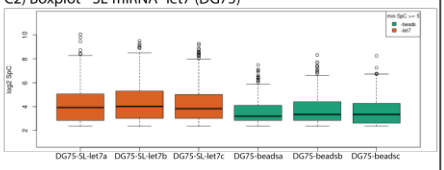
B4) PCA - SL-miRNA -let7 (BCBL-1)



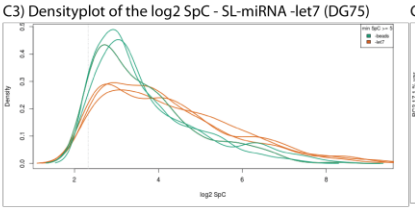
C1) Protein and SpC summary - SL-miRNA -let7 (DG75)

	Nb proteins	Total SpC	max SpC count
DG75a_SL-let7	974	21838	1068
DG75b_SL-let7	833	19647	731
DG75c_SL-let7	1012	20978	619
DG75a_beads	613	5989	179
DG75b_beads	551	6948	319
DG75c_beads	708	6691	306

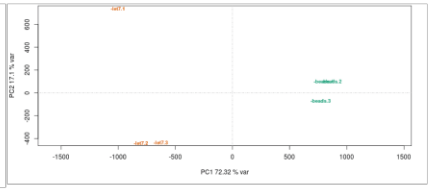
C2) Boxplot - SL-miRNA -let7 (DG75)



C3) Densityplot of the log2 SpC - SL-miRNA -let7 (DG75)



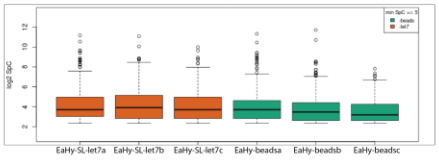
C4) PCA - SL-miRNA -let7 (DG75)



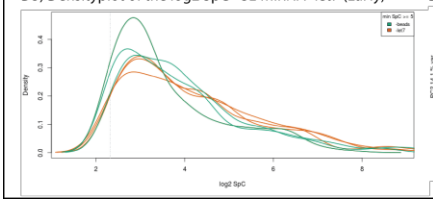
D1) Protein and SpC summary - SL-miRNA -let7 (EaHy)

	Nb proteins	Total SpC	max SpC count
EaHya_SL-let7	935	24303	2310
EaHyb_SL-let7	766	19858	2154
EaHyc_SL-let7	794	16803	1019
EaHya_beads	862	21568	2541
EaHyb_beads	638	15515	3333
EaHyc_beads	622	6508	225

D2) Boxplot - SL-miRNA -let7 (EaHy)



D3) Densityplot of the log2 SpC - SL-miRNA -let7 (EaHy)



D4) PCA - SL-miRNA -let7 (EaHy)

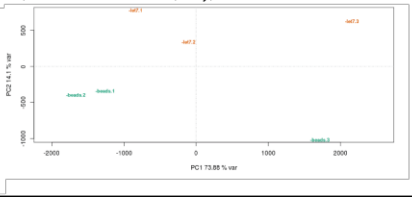


Figure 11. Descriptive analysis of the four chromatography datasets with the SL-miRNA-let7 bait. Table summary (1), boxplot (2), density-plot (3) and PCA (4) are displayed for all the nuclear extract used. (A) BC-3 nuclear extract -let7 chromatography. B) BCBL-1 nuclear extract -let7 chromatography. C) DG75 nuclear extract -let7 chromatography. D) EaHy926 nuclear extract -let7 chromatography.

A)

IDs	SL-let7_BC3	SL-let7_BCBL1	SL-let7_DG75	SL-let7_EaHy
ABCE1_HUMAN	NA	NA	Enriched	NA
ABCF1_HUMAN	Enriched	NA	Enriched	NA
ACTBL_HUMAN	NA	NA	NA	Enriched
AKAP8_HUMAN	NA	NA	Enriched	NA
AMPM2_HUMAN	NA	NA	Enriched	NA
APEX1_HUMAN	NA	NA	Enriched	Enriched
C2AIL_HUMAN	NA	NA	NA	Enriched
CARF_HUMAN	NA	NA	NA	Enriched
CELF1_HUMAN	NA	NA	NA	Enriched
CIRBP_HUMAN	Enriched	NA	Enriched	Enriched
CN166_HUMAN	NA	NA	Enriched	NA
CSTF3_HUMAN	NA	NA	Enriched	NA
DAZP1_HUMAN	NA	Enriched	NA	NA
DBPA_HUMAN	NA	NA	NA	Enriched
DDX1_HUMAN	NA	NA	Enriched	NA
DDX58_HUMAN	NA	NA	NA	Enriched
DHX15_HUMAN	NA	NA	NA	Enriched
DHX36_HUMAN	NA	NA	NA	Enriched
DHX9_HUMAN	NA	Enriched	Enriched	Enriched
DICER_HUMAN	NA	NA	NA	Enriched
DRG1_HUMAN	Enriched	Enriched	Enriched	Enriched
DSRAD_HUMAN	NA	Enriched	Enriched	Enriched
E2AK2_HUMAN	Enriched	Enriched	Enriched	Enriched
EIF2A_HUMAN	NA	NA	Enriched	NA
EWS_HUMAN	NA	NA	NA	Enriched
FA98B_HUMAN	NA	NA	Enriched	NA
FAK1_HUMAN	NA	NA	NA	Enriched
FEN1_HUMAN	Enriched	NA	NA	NA
FUBP3_HUMAN	NA	NA	NA	Enriched
G3BP1_HUMAN	NA	NA	Enriched	NA
GNL3_HUMAN	NA	NA	Enriched	Enriched
GRSF1_HUMAN	NA	Enriched	Enriched	Enriched
GTPB1_HUMAN	NA	NA	Enriched	NA
GTPB8_HUMAN	NA	Enriched	NA	NA
HBS1L_HUMAN	NA	Enriched	Enriched	Enriched
HNRL1_HUMAN	NA	Enriched	Enriched	Enriched
HNRLL_HUMAN	NA	NA	Enriched	Enriched
HNRPD_HUMAN	NA	NA	NA	Enriched
IF2A_HUMAN	NA	NA	Enriched	NA
IF2B_HUMAN	Enriched	NA	Enriched	NA
IF2B1_HUMAN	NA	NA	Enriched	Enriched
IF2B2_HUMAN	NA	NA	NA	Enriched
IF2B3_HUMAN	NA	NA	NA	Enriched
IF2G_HUMAN	NA	NA	Enriched	NA
IF2P_HUMAN	NA	NA	Enriched	NA
ILF2_HUMAN	Enriched	NA	Enriched	Enriched
ILF3_HUMAN	Enriched	Enriched	Enriched	Enriched
IQGA1_HUMAN	NA	NA	Enriched	NA
JUN_HUMAN	NA	NA	NA	Enriched
LA_HUMAN	NA	Enriched	NA	Enriched
LN28B_HUMAN	Enriched	Enriched	Enriched	NA
LPPRC_HUMAN	NA	NA	Enriched	Enriched
LSG1_HUMAN	NA	NA	Enriched	NA

B)

IDs	SL-let7_BC3	SL-let7_BCBL1	SL-let7_DG75	SL-let7_EaHy
MCES_HUMAN	NA	NA	Enriched	NA
MET16_HUMAN	NA	NA	Enriched	Enriched
MSI2H_HUMAN	NA	Enriched	Enriched	NA
NMT1_HUMAN	NA	NA	Enriched	NA
NUCL_HUMAN	Enriched	NA	NA	NA
PA2G4_HUMAN	NA	Enriched	Enriched	NA
PABP4_HUMAN	NA	NA	NA	Enriched
PAIRB_HUMAN	NA	NA	Enriched	NA
PCF11_HUMAN	NA	NA	Enriched	NA
PPIB_HUMAN	NA	NA	Enriched	NA
PRKRA_HUMAN	NA	Enriched	NA	Enriched
PTBP2_HUMAN	NA	NA	Enriched	NA
PXK_HUMAN	NA	NA	Enriched	NA
RAGP1_HUMAN	NA	NA	Enriched	Enriched
RAN_HUMAN	NA	Enriched	Enriched	Enriched
RANG_HUMAN	NA	Enriched	Enriched	Enriched
RBM3_HUMAN	NA	NA	Enriched	Enriched
RBMS2_HUMAN	NA	NA	NA	Enriched
RBP2_HUMAN	NA	NA	Enriched	Enriched
RBP56_HUMAN	NA	NA	Enriched	Enriched
RCC1_HUMAN	NA	NA	Enriched	NA
RED1_HUMAN	NA	NA	NA	Enriched
RENT1_HUMAN	NA	NA	Enriched	Enriched
RGPD1_HUMAN	NA	NA	Enriched	NA
RGPD2_HUMAN	NA	NA	Enriched	NA
RGPD5_HUMAN	NA	NA	Enriched	NA
RGPD8_HUMAN	NA	NA	Enriched	NA
RL22_HUMAN	NA	NA	Enriched	NA
RL22L_HUMAN	NA	NA	Enriched	NA
ROA0_HUMAN	NA	Enriched	Enriched	Enriched
RS17_HUMAN	NA	NA	Enriched	NA
RS17L_HUMAN	NA	NA	Enriched	NA
RS7_HUMAN	NA	NA	Enriched	NA
RTCA_HUMAN	NA	NA	Enriched	NA
RTCB_HUMAN	NA	NA	Enriched	NA
SLIRP_HUMAN	NA	Enriched	NA	Enriched
SNRPA_HUMAN	NA	NA	Enriched	Enriched
SRP09_HUMAN	NA	NA	Enriched	NA
SRSF9_HUMAN	NA	Enriched	Enriched	Enriched
STAU1_HUMAN	NA	Enriched	NA	Enriched
STAU2_HUMAN	NA	NA	NA	Enriched
STRBP_HUMAN	Enriched	Enriched	Enriched	Enriched
SYNC_HUMAN	NA	NA	Enriched	NA
TFB2M_HUMAN	NA	NA	NA	Enriched
TIA1_HUMAN	NA	NA	Enriched	NA
TIAR_HUMAN	NA	Enriched	Enriched	NA
TRI25_HUMAN	NA	Enriched	Enriched	Enriched
WIBG_HUMAN	NA	NA	Enriched	NA
XRN2_HUMAN	NA	NA	Enriched	Enriched
YBOX2_HUMAN	NA	NA	NA	Enriched
YLPM1_HUMAN	NA	NA	Enriched	NA
ZC3HF_HUMAN	NA	NA	Enriched	NA
ZFR_HUMAN	Enriched	Enriched	Enriched	Enriched
ZN346_HUMAN	Enriched	Enriched	Enriched	Enriched

C)

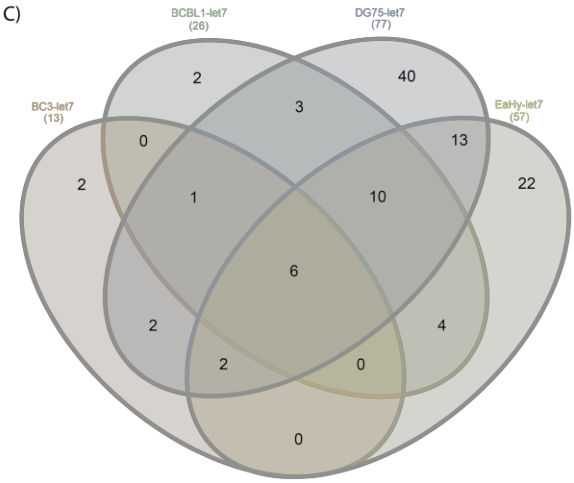


Figure 12. SL-miRNA-let7 significantly (p.value<0.05) enriched dataset by the 4 different cell lines. A) and B) Merge list of the significantly enriched proteins. C) Venn diagram comparison of the significantly enriched proteins.

Knowing the number of identified proteins and spectral count is fundamental to know if the eluates are overall similar in term of distribution and homoscedasticity. For all the SL-miRNA chromatographies and the corresponding control conditions, we can see an important variation of the identified protein number, total spectral count and maximum spectral count obtained for a given chromatography (Figure 7, Figure 9, Figure 11). In addition, boxplot chart and density-plot reveal that spectral count median and dispersion vary a lot among samples. This disparity means that the different replicates, for a given condition, display important variation which can be visualized by the PCA (Figure 7, Figure 9, Figure 11). As we can see, replicates tend to be separated on the PCA chart suggesting that more than 2 factors could explain replicate differences. Here, the main factor that is supposed to explain sample difference is the bait identity. All these charts, for all the agarose chromatography datasets, suggest that the differential analysis will be very difficult to perform. Indeed, without taking into account the variance of the spectral count number for a given protein among a specific triplicate condition, boxplot quartile and density-plot clearly emphasize that each sample spectral count does not seem to follow a similar symmetric normal distribution. Overall descriptive analysis clearly suggests important chromatography issues. Thus, I expect that proteins identified by both the SL-miRNA and the control baits certainly will not be considered as significantly enriched during the statistical analysis, even though the proteins could display a good enrichment ratio.

Differential analysis is performed for each SL-miRNA and cell lines and displayed as a merged list of enriched proteins. In these lists, the identified proteins are labelled with their uniprot entry names and sorted by alphabetic order. Proteins that are common or specific to each cell lines are visualized using a Venn diagram (-K4: Figure 8; -K11: Figure 10; -let7: Figure 12). The only proteins considered as significantly enriched by the SL-miRNA are the proteins not identified or very weakly identified by the negative control bait. This means that the differential analysis only highlights proteins present in one condition (SL-miRNA) and not in the other (control beads) which correspond to the proteins “missing on an entire condition” (MEC). Nonetheless, the differential analysis managed to display several proteins considered as significantly enriched on the SL-miRNA (Figure 8, Figure 10, Figure 12). Among these significantly enriched proteins, we find a number of known cofactors such as the microprocessor partners ILF2, ILF3, DHX15 (Gregory *et al.*, 2004). In addition, E2AK2 (PKR) and DSRAD (ADAR), two dsRNA-binding proteins, are also significantly enriched (Bou-Nader *et al.*, 2019; Herbert, 2019). Taken together, these results suggest that SL-miRNA baits fold as expected. Concerning the positive control bait, SL-miRNA-let7 does allow significant enrichment over the negative control of the well-known Let-7 interactor, LN28B (Heo *et al.*, 2008). However, it is not the case for another known co-factor, ROA1 (hnRNP-A1) (Michlewski and Cáceres, 2010). Although this indicates that the agarose/heat elution approach is not optimal for SL-miRNA co-factors identification, the generated datasets still contain some useful information for future projects in the laboratory. In addition, the candidates list can also be generated with a more qualitative approach, meaning without statistical approach. For example, for a given protein, enrichment ratios can be simply calculated for the replicates and transformed as a qualitative dataset (enriched – non-enriched) which will greatly increase the number of protein candidates.

III.A.2 Magnetics RNA-pulldown for sensitivity and specificity increase beads

As the descriptive analysis of the agarose chromatography suggested, this approach is not enough sensible, specific and reproductible. This is why, I decided to set up new RNA chromatography approach. To perform the pulldown experiment for all the different KSHV's SL-miRNAs, I decided to use magnetic beads instead of agarose beads. It will be easier to manipulate, and it won't produce deleterious agarose degradation products. In addition to the magnetic beads, a "specific" DNase elution has been set up to decrease the background noise as well as limiting the streptavidin sample contamination. To do so, a 5' phosphate-ssDNA-3'TEG-Biotin oligo (ssDNA) is ligated to the 3' end of the SL-miRNA. Resulting chimeric molecule can afterward be linked to streptavidin magnetic beads and eluted from the baits by the DNase RQ1 treatment after the pulldown. Here, bait generation, elution and mass spectrometry analysis only differ from the previous chromatography. Now, I will describe the different steps require to validate this new approach. In the following sections, I will expose the chimeric bait formation and control (see part III.A.2.b and III.A.4.a). Then, I'll assess the relevance of the specific DNase elution by comparing it to a classical heat elution (see part III.A.4). Finally, I will describe the data mining and exploration procedure (see part III.A.4.b) for the chromatography performed with all the KSHV SL-miRNA.

III.A.2.a Chromatography set up

The same *in vitro* transcribed SL-miRNA transcripts as described in part III.A.1.a were used to generate the chimeric bait. To form the chimera between the SL-miRNA and a modified ssDNA oligo, we used RNA ligation. To avoid SL-miRNA self-ligation, a mutated T4 DNA ligase called T4 Rnl2 (1–249) is used. Unlike the wildtype ligase, T4 Rnl2 (1–249) cannot ligate a mono-phosphorylated 5´ end of an RNA or DNA oligo to the 3´OH end of an RNA oligo due to the T4 Rnl2 (1-249) inability to pre-adenylate the RNA, which is absolutely required for the ligation. Therefore, T4 Rnl2 (1–249) specifically ligates pre-adenylated 5' end of DNA or RNA to an available 3' hydroxyl end. This ligase is commonly used to generate miRNA library (Hafner *et al.*, 2011). Here, the 3' end of the ssDNA is protected by a biotin molecule, and the 5' end of the ssDNA will be pre-adenylated. Thus, only the SL-miRNA-ssDNA-TEG-Biotin will be ligated together. These different steps are exposed below.

III.A.2.b Pre-adenylation step

Pre-adenylation of the ssDNA is realized by the Mth RNA ligase (New England Biolabs). This enzyme generates the conversion of a 5'P-DNA to a 5'App-DNA (Zhelkovsky and McReynolds, 2011). Pre-adenylation is controlled by ethidium bromide staining or northern blot analysis on a 17.5 % acrylamide 8 M urea gel (Figure 13.A.B) prior to any downstream enzymatic reaction. Pre-adenylation step has been tested with the recommend amount of enzyme or two time less. Ethidium bromide staining as well as northern blot demonstrate that this reaction is a highly efficient process, as we cannot see any signal at 21 nt after the enzymatic treatment, thus eliminating the need for App-DNA gel purification. Once the ssDNA oligo is pre-adenylated, chimeric formation conditions are tested.

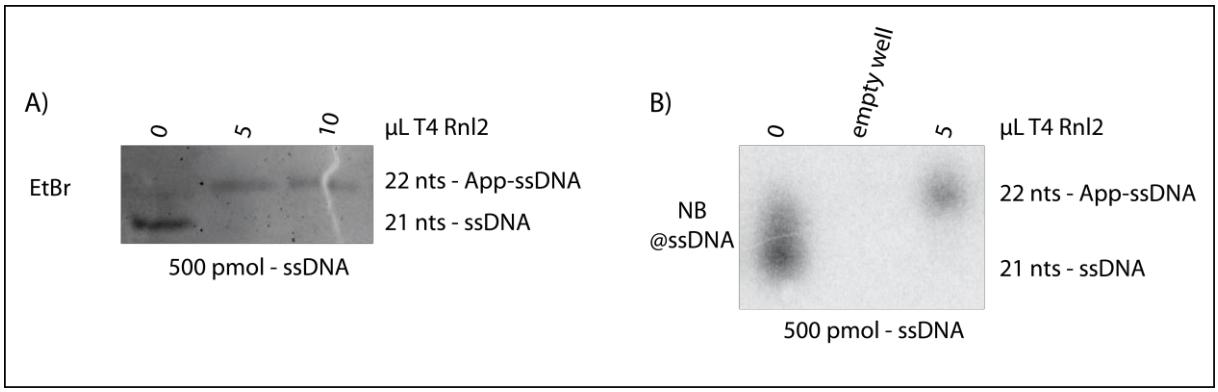


Figure 13. Pre-adenylation control. Both sample assays are migrated on a 17.5 % acrylamide 8 M urea gel. A) Ethidium bromide staining of 500 pmol of 5'phosphate-ssDNA-TEG-Biotin treated with the recommended amount of enzyme (10 µL) or 2 times less. B) Northern blot detection of the 5'phosphate-ssDNA-TEG-Biotin and its adenylated counterpart.

III.A.2.c SL-miRNA and ssDNA ligation

It is important to generate a substantial amount of chimeric molecule since the ssDNA bearing the biotin group will be required for the binding to the streptavidin beads. Therefore, it is important to select a condition in which all the pre-adenylated ssDNA will be transformed into the chimeric molecule. This step needs to be as efficient as possible to avoid any attachment of non-ligated ssDNA to the beads. Indeed, the negative control will be ssDNA attached to beads, which will allow the generation of an enrichment ratio for specific protein binding SL-miRNA. To this end, ligation procedure was tested with different excess of SL-miRNA during 2 h in the presence of T4 Rnl2 (1-249) ligase. The resulting molecules are analysed after the ligation step by northern blot analysis (Figure 14). A probe directed against the SL-miRNA allows visualization of the non-ligated SL-miRNA, the chimeric molecule as well as chimera degradation products (Figure 14.C). A probe directed against the ssDNA allow visualization of the non-pre-adenylated ssDNA, the pre-adenylated ssDNA, the chimera as well as its degradation products (Figure 14.A.B). We can see that using a stochiometric amount of SL-miRNA and ssDNA is not sufficient to transform all the ssDNA into chimera. Using a three- or four-time excess seems more suitable to make sure that the entirety of the ssDNA is transformed into chimera. Surprisingly, in this assay as well as in further experiments, non-pre-adenylated ssDNA is detected after ligation even though the previous adenylation step was highly efficient. This observation suggests that the chimeric formation step can somehow be aborted, inducing a loss of the pre-adenylation modification on the ssDNA. Actually, ligase enzymes excise the adenine bi-phosphate at the 5' end to obtain energy (Ho *et al.*, 2004) to perform the ligation, partially explaining the resurgence of the 21 nt signal. We then checked if increasing the reaction time could improve the ligation efficiency but saw that it only increased the degradation of both SL-miRNA and chimera. Thus, 1 h of ligation seems as efficient as 2 h or overnight and displays less degradation product (Figure 14.C). For the following step, we therefore settled with the following conditions: 5 µL Mth for 500 pmol of ssDNA; 3-time excess of SL-miRNA compared to the App-ssDNA, 1 h of ligation incubation with 2,000 U of T4 Rnl2 (1-249). Once the chimeric molecules are formed, the chimeric bead binding step has to be assessed.

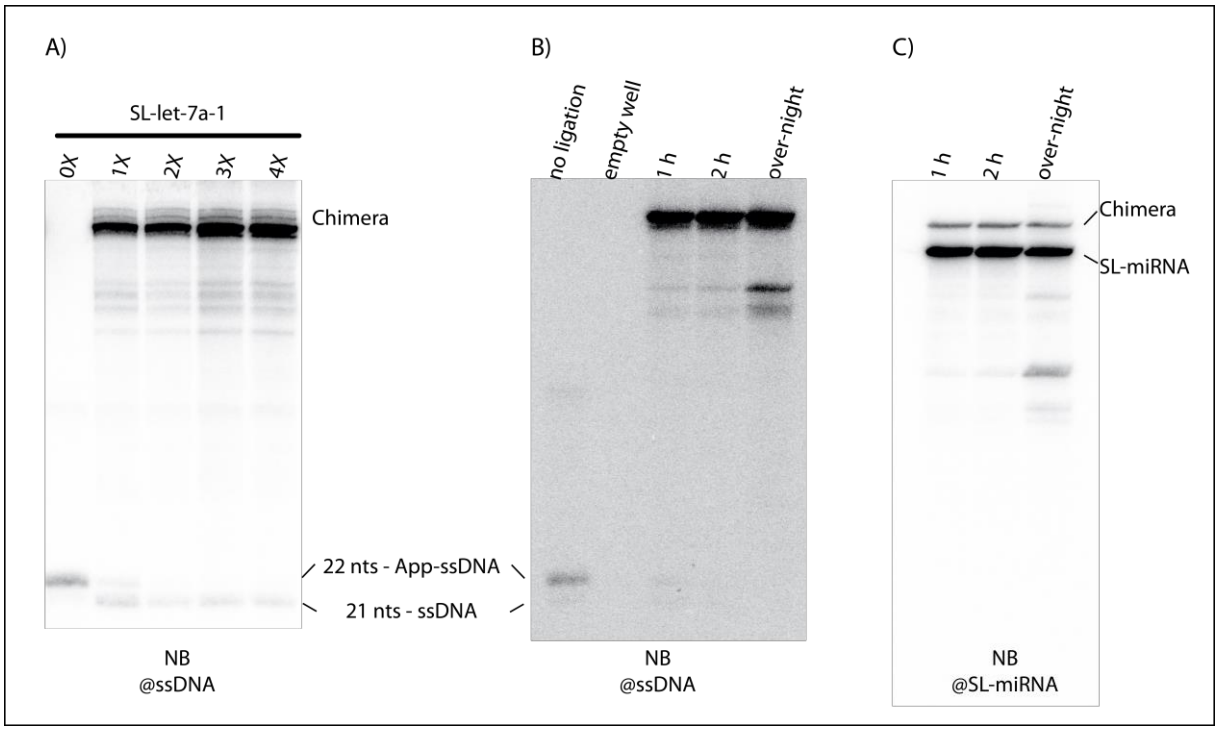


Figure 14. Chimera formation. A) Northern blot analysis of ligation test of 5'phosphate-ssDNA-TEG-Biotin with different excess of SL-miRNA. 500 pmol of the ssDNA has been pre-adenylated and used to generate chimeric molecules with 500 pmol (1X) or different excess of SL-miRNA (up to 4 times). After the 2 h of ligation incubation samples are loaded in a 17.5 % acrylamide 8 M urea gel, transferred, and analysed with a radioactive probe matching to the ssDNA. B) and C) Northern blot analysis of different chimeric ligation times. Ligation products are analysed by northern blot analysis following a 17.5 % acrylamide 8 M urea gel migration with either a probe matching the ssDNA (A) or the SL-miRNA (B). Different ligations times have been tested: 1 h, 2 h and overnight with a three-time excess of SL-miRNA compare to the App-ssDNA.

III.A.2.d Magnetic beads binding, washing and elution

Binding efficiency and elution have been assessed with a radiolabelled SL-miRNA following the same conditions as the agarose beads binding test (see part III.A.1.b). One-hour binding of the biotinylated chimera with the magnetic streptavidin beads has been tested, as well as 3 different RQ1 DNase treatment: 2 h with 10 U; 2 times 2 h with 5 U; 4 h with 10 U (Figure 15.A.B.C). Chimera binding on the magnetic beads during 1 h at room temperature seems potent (>95 %) and reproductible. The strong interaction between the biotin and the streptavidin is responsible of this binding efficacy reproducibility. The different elution conditions do not display significant differences and never reach more than 48.7 % of SL-miRNA eluted from the beads (Figure 15.A.B.C). Following this experiment, binding and elution conditions have been set up at 1 h room temperature and 2 h 37°C with 10 U of RQ1 DNase respectively (Figure 15.A.B.C). Elution efficiency will certainly decrease within the pulldown context due to steric hinderance with cellular factors, but I chose not to increase the elution time any further to avoid potential sample degradation or contamination. Now that the bait formation and elution have been set up, we need to determine the washing conditions. The

same washing condition as the agarose protocol (200 mM of KCl, inspired from (Michlewski and Caceres, 2010)) was first used. Several methods can eventually be used to increase specific protein SL-miRNA interaction during pulldown experiments. For example, baits can be blocked with BSA (1 µg/mL) and total yeast tRNA (50 µg/mL) or the cellular extract can be pre-cleared by the negative bait control. Here, I only tried the pulldown experiment with either non-blocked beads or blocked beads, followed by 5 washing steps (200 mM KCl) and the previously set up DNase treatment (2 h at 37°C with 10 U of RQ1 DNase). Pre-clearing the sample has not been tested since it would consume a lot of expensive magnetic beads. I then tested the whole procedure followed by the loading of an acrylamide gel and silver staining (Figure 15.D.E). This was done in duplicate for the SL-let-7a-1-beads (Chimera-1 and -2) and ssDNA-beads (ssDNA-1 and -2) in order to test the reproducibility of the experiment. As can be seen in Figure 15.D.E, proteins still come off after the 5th washing step suggesting that either the wash stringency is too high or that multiple washing steps simply disturbed baits-proteins complexes. The blocking step does not seem to make any difference, as far as the silver nitrate stained gels indicate, it will therefore not be used in the procedure. The different bait conditions within one gel (blocked or not) display a closely related profile. However, silver nitrate staining is far less sensible than the mass spectrometry identification. Similar bands between conditions can correspond to the global background assay. Now that all the pulldown steps are set up, I can compare the DNase elution method with the heat elution one.

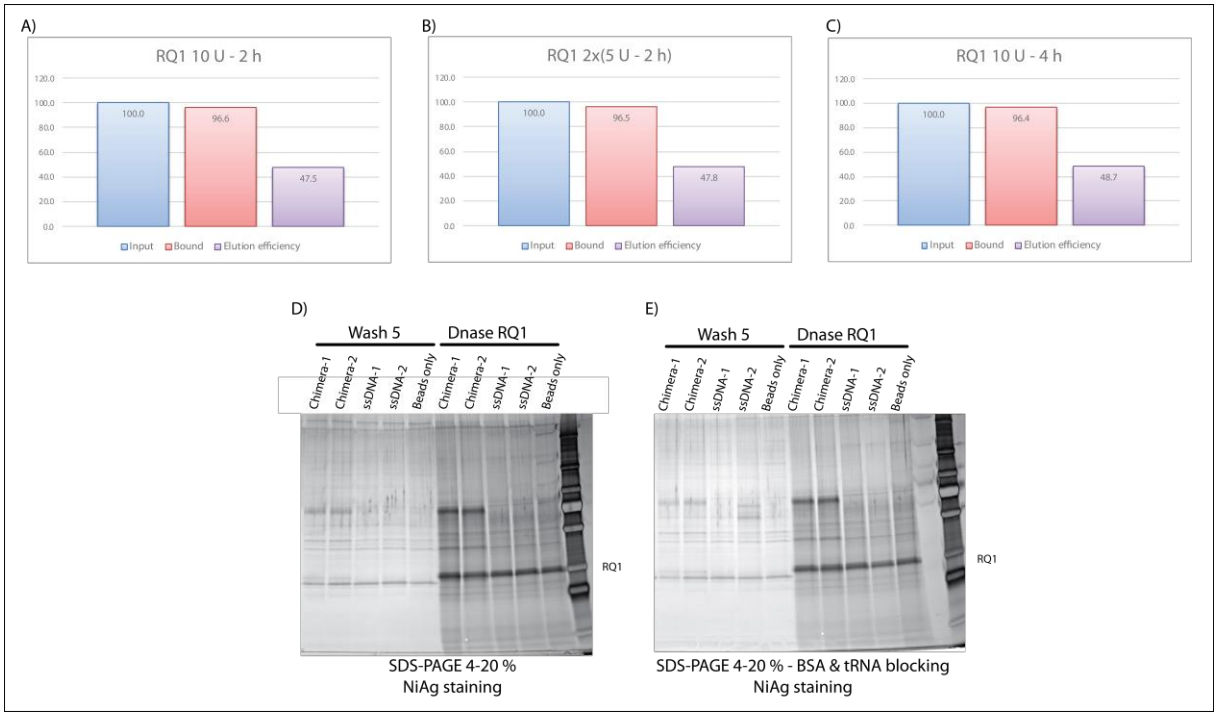


Figure 15. Chimera binding, washing and elution efficiency. A), B), C), % of the chimera signal is compare to the input after chimera radiolabelling and Cerenkov counting. Elution with DNase RQ1 occurred either with 10 U during 2 h (A), 2 times 5 U during 2 h (B) or 10 U during 4 h (C). D) and E) Silver nitrate staining of RNA pulldown. RNA pulldown is performed with either non-blocked beads (D) or blocked beads (1 µg/mL BSA; 50 µg/mL total yeast tRNA) (E) following 5 washing steps at 200 mM KCl within the same nuclear extract. Chimera-1 and Chimera-2 correspond to SL-let-7a-1 linked to the ssDNA oligo and the beads. ssDNA-1 &

ssDNA-2 correspond to the ssDNA bound to the beads without any SL-miRNAs. Beads only correspond to the beads without ssDNA, or chimera. Samples are loaded in a 4-20 % SDS-PAGE and stained with silver nitrate kit (SilverQuest, Invitrogen).

III.A.2.e Heat vs DNase treatment elution

We directly compared DNase and heat elution in order to control that the DNase elution limits the number of non-specific protein elution from the baits. In Figure 16.A we can see that a similar profile after the pulldown experiments is obtained between the chimera and the ssDNA for both elution methods. However, it is obvious that heat elution displays a greater intensity and a far greater number of proteins bands. At this time, it is therefore evident that the DNase treatment uncouples less proteins from the baits than the heat elution. The absence of visible differences between the chimera and the control ssDNA upon DNase elution can come from a low protein quantity for which silver nitrate staining is not sensible enough. Northern blot assay shows both SL-miRNA and chimera molecules release from the beads after, respectively, the DNase and heat elution. As expected, the DNase treatment only releases the SL-miRNA compare to the heat treatment as we can see a shift on the northern blot (Figure 16.B). In addition, the quantity of chimera released by heat seems lower than the DNase-released SL-miRNA suggesting either a potential chimera degradation upon heat elution or simply a weaker elution efficiency of the biotin/streptavidin bound chimera from the beads. However, during the heat elution, proteins bound to SL-miRNA are denatured and lose their nucleic acid binding activity as we can clearly see a great quantity of protein stained by silver nitrate. SDS-PAGE staining is not sufficient to really assess the utility of the DNase treatment. This is why pulldown experiments are analysed by mass spectrometry in order to better compare the two elution methods.

The detailed comparison between heat and DNase elution has been first assembled into a manuscript that we wanted to submit for publication. However, a wrong dataset had been used for the draft of the manuscript, which we therefore could not submit. In the following part, we used the correct dataset to perform this comparison (see disclaimer below).

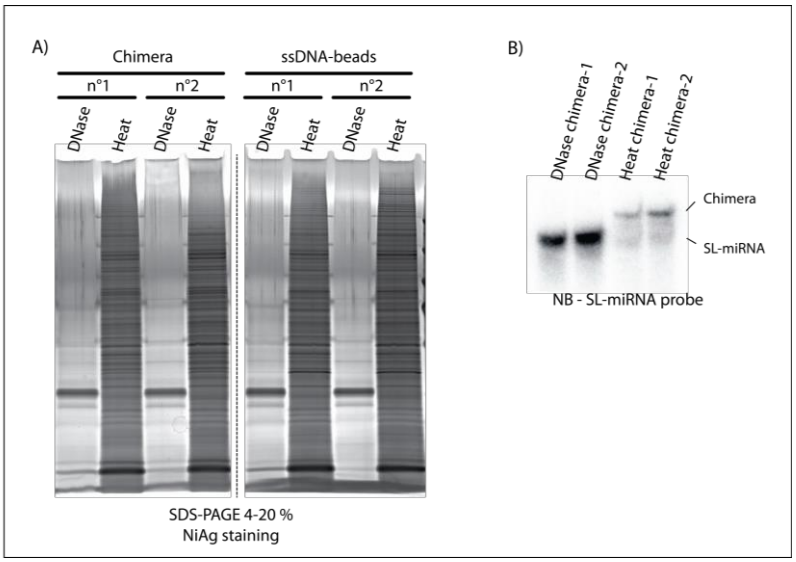


Figure 16. DNase and heat elution comparison. Each bait (chimera or ssDNA) are either eluted using DNase or heat treatment. A) Pulldown samples are loaded in an SDS-PAGE and stained with NiAg. B) Northern blot analysis is performed to detect either the SL-miRNA or the chimera after, respectively, the DNase or heat elution. Black dashed lines represent a cut on the

raw data for the sake of the figure understanding.

III.A.3 Identification of microRNA precursor binding proteins by DNase-assisted RNA chromatography coupled with mass spectrometry

DISCLAIMER: Initially a wrong dataset has been used to perform the DNase and Heat elution comparison. When we looked at the correct dataset, we realized that there was one problematic replicate for one of the samples. We figured out that this problem came from a corrupt file that was generated during the MS analysis. Everything was tried to see whether it was possible to extract any data from this corrupt file, and the MS facility has even been in touch with the engineers at Thermo, but unfortunately without success. So, what we are going to do now is to use the available data, in which one condition (DNase-let7_replicate3) will exist with one duplicated replicate (DNase-let7_replicate2) instead or a real triplicate, in order to perform the comparison.

RNA chromatography using heat elution followed by mass spectrometry is a commonly used approach to identify specific RNA binding proteins. In order to lower the inherent limitations of this approach such as the important background noise, we set up an approach that specifically elutes the RNA bait from the matrix using DNase treatment, thereby increasing the enrichment of specific binders and decreasing non-specific contaminants. This method relies first on the formation of a chimeric molecule composed of the RNA of interest ligated to a biotinylated single stranded oligodesoxyribonucleotide and second on its specific elution from the matrix by RQ1 DNase treatment. Here, we present a comparison of our DNase-assisted chromatography approach to a classic heat eluted chromatography using pri-miRNA hsa-let-7a-1 as a bait. We also describe how we processed and analyzed mass spectrometry data with the help of the user-friendy software ProStaR.

III.A.3.a Introduction to the DNase-assisted RNA chromatography experiment

Similarly to any RNA molecules, as soon as pri-miRNAs are transcribed, they interact with a great number of protein partners to form ribonucleoprotein complexes that will ensure their correct maturation, structural arrangement (Contrant *et al.*, 2014; Ganser *et al.*, 2019) and cell specificity (Landgraf *et al.*, 2007; Juzenas *et al.*, 2017; Alberti *et al.*, 2018), which are absolutely fundamental to achieve their proper function in the cell. In order to be able to respond to internal and external stimuli, RNPs have to be highly dynamic and tightly regulated (Hentze *et al.*, 2018). The comprehensive identification of specific RBPs involved in these processes are essential to decipher the myriad of functions associated with RNAs. One of the most commonly used methods to discover such co-factors, is based on RNA chromatography followed by mass spectrometry analysis. Although this approach has been subjected to several improvements to decrease the background noise and increase the specificity such as playing with the stringency of the washing steps, bead blocking, or cell extract pre-clearing, it is still a big challenge to discriminate between true and false positive binders.

Here, we adapted an RNA chromatography previously described to purify ribosomal particles (Prongidi-Fix *et al.*, 2013), in order to discover new co-factors of pri-miRNA

maturation step. The ingenuity of this method relies on the formation of a chimeric molecule composed of the pri-miRNA of interest and a DNA oligonucleotide coupled to beads, such as specific pri-miRNPs are selectively eluted by DNase digestion of the DNA moiety. We also propose an improved protocol to produce the chimera baits, using a modified RNA ligase. To highlight the advantages of our approach, we compared two identical RNA chromatographies using the same chimeric baits but eluted either by heat or by DNase treatment. The pri-miRNA hsa-let-7a-1 was chosen since specific RBPs have already been described, such as ROA1 (HNRNPA1) or LN28B (LIN28B) (Heo *et al.*, 2008; Michlewski and Cáceres, 2010). Here, we describe in details the different steps involved in the two RNA chromatographies in solution as well as we propose a hand-to-hand guidance to analyze mass spectrometry data using the ProStaR software (Wieczorek *et al.*, 2017, 2019) to identify specific protein binders.

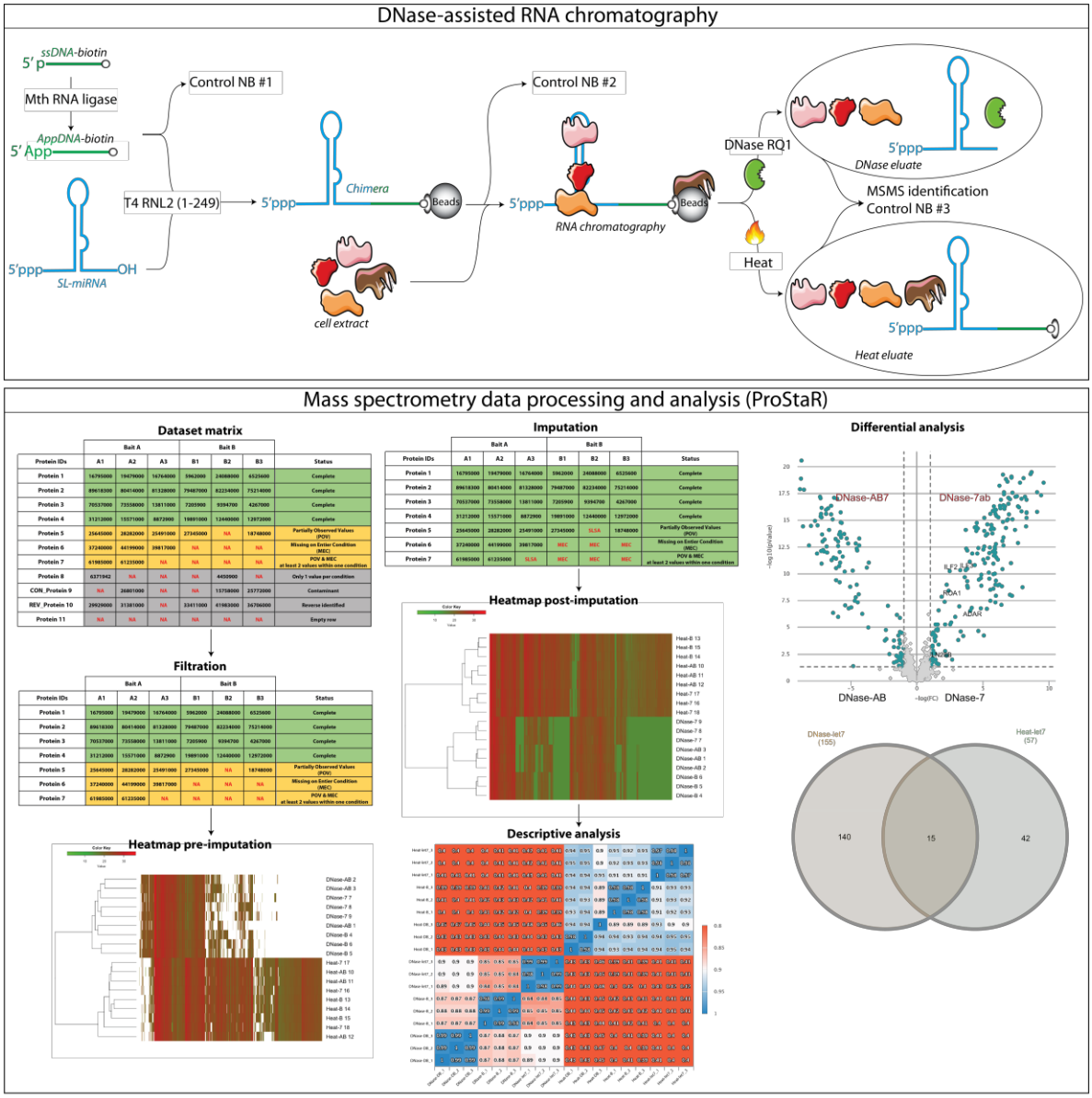


Figure 17. Graphical abstract of the DNase-assisted RNA chromatography and the mass spectrometry data analysis pipeline. Briefly, chromatography relies on a chimeric bait molecule composed of an RNA of interest and a ssDNA oligo allowing a specific elution of the RNA from the matrix using a DNase treatment. NB #1 correspond to the first northern blot

requires to control the pre-adenylation step. NB #2 correspond to the chimera northern blot control. NB #3 correspond to the elution northern blot control. Mass spectrometry dataset, manipulated with the ProStaR software, is filtered and missing values are imputed for the partially observed values (POV) or the missing on entire condition (MEC) values. Correlation matrix and principal component (descriptive analysis) are generated to get insight on the data quality prior to any differential analysis.

III.A.3.b DNase-assisted RNA chromatography results

Our DNase-assisted RNA chromatography in solution relies on a classical chromatography approach (Figure 17). Briefly, a chimeric molecule composed of a stem loop miRNA precursor (SL-miRNA) and a DNA oligonucleotide was generated in vitro. First, a single stranded biotinylated DNA oligonucleotide (ssDNA-biotin) is pre-adenylated (AppDNA-biotin) with the Mth RNA ligase. AppDNA-biotin and SL-miRNA are ligated to each other with the T4 RNL2 ligase and attached to magnetic beads to construct the baits. These baits were then incubated with a BC-3 lymphocytes cell extract and washed several times. Finally, the proteins retained were released either by heat treatment (Heat-) or by cleavage of the DNA moiety of the chimeric molecule (DNase-) and characterized by a label free mass spectrometry analysis. Resulting dataset matrix is then cautiously explored and filtrated with the help of ProStar software. Missing values are imputed prior to any descriptive analysis. Finally, the differential analysis is performed. To compare both elution methods, we performed the chromatography in triplicate from the same BC-3 lymphocytes nuclear extract and using 3 different baits: beads alone (B_); ssDNA-biotin oligonucleotide coupled to the beads (DB_); pri-miRNA/DNA chimera coupled to the beads (let7_). Table 2 lists the names of the different samples used for the mass spectrometry data processing and analysis. Here, it's important to precise again that the DNase-let7_3 sample is a duplicate of the sample DNase-let7_2.

Bait / Elution	DNase	Heat
Beads	DNase-B_1	Heat-B_1
	DNase-B_2	Heat-B_2
	DNase-B_3	Heat-B_3
ssDNA-beads	DNase-DB_1	Heat-DB_1
	DNase-DB_2	Heat-DB_2
	DNase-DB_3	Heat-DB_3
SL-let-7a-1-ssDNA-beads	DNase-let7_1	Heat-let7_1
	DNase-let7_2	Heat-let7_2
	DNase-let7_3	Heat-let7_3

Table 2. MSMS sample labelling during the data analysis procedure. Due to an informatic issue, the DNase-let7_3 sample (highlighted in yellow) is a duplicate of the sample DNase-let7_2.

Chimeric molecules are formed with the let-7a-1 precursor which mimic the pri-let-7a-1 substrate of the microprocessor to generate the bait (-let7). ssDNA-biotin oligonucleotide

coupled to the beads (DB_) control bait act as the negative control for the chimera bait condition to generate fold enrichment of protein binding SL-miRNA. Beads alone (B_) control bait can act as a negative control for the ssDNA-biotin oligonucleotide coupled to the beads (DB_) to generate fold enrichment of protein binding ssDNA. Pre-adenylation step, chimera formation and the elution are controlled by northern blot analysis. Protein identification and quantification was done using MaxQuant (Cox *et al.*, 2014) and ProStaR. In the following part, the affinity chromatography is described from the bait construction to the elution control. Then, the mass spectrometry data processing and analysis is exposed.

III.A.3.c DNase-assisted RNA chromatography assay and controls

Stem loop miRNA precursors (SL-miRNAs), corresponding to the hsa-let-7a-1 pre-miRNA with 20 nt of flanking sequences required for efficient recognition by DROSHA/DGCR8, were in vitro transcribed and purified on a 12 % acrylamide/bisacrylamide urea gel. A synthetic 5' phosphate-ssDNA-3'TEG-Biotin oligonucleotide (ssDNA-biotin) of 21 nt (IDT) was then ligated to the SL-miRNA to form the chimeric molecule. To avoid SL-miRNA self-ligation during the chimera formation, the truncated T4 RNA ligase RNL2 (1–249) was used. This enzyme specifically ligates a 5' adenylated DNA or RNA to the 3' hydroxyl of another RNA (Viollet *et al.*, 2011). Here, the 3' end of the ssDNA is protected by a biotin molecule, and the 5' end of the ssDNA was pre-adenylated by the Mth RNA ligase (New England Biolabs). This enzyme catalyzes the conversion of a 5' mono phosphate DNA to a pre-adenylated DNA (AppDNA-biotin). Thus, only the SL-miRNA-ssDNA-TEG-Biotin will be generated during the following RNA ligation step.

Each step of the bait formation procedure was followed by northern blot analysis (Figure 18). As seen on this figure, the pre-adenylation reaction was highly efficient, as no signal remained at 21 nt after the enzymatic treatment, thus eliminating the need for gel purification of the AppDNA-biotin oligonucleotide (Figure 18A). Once the ssDNA-biotin oligo was pre-adenylated, chimeric molecules were formed. In order to maximize the quantity of chimeric molecules, we selected a condition in which almost all the pre-adenylated ssDNA-biotin was ligated to the RNA. This step needs to be as efficient as possible because ssDNA-biotin bound to the beads will be used as a negative control during the chromatography to calculate the enrichment ratio of proteins binding specifically to the SL-miRNA. For this reason, the ligation procedure was performed with a 3 times excess of SL-miRNA compared to the AppDNA-biotin (Figure 18.B). The formation of the chimeric molecules induced a signal shift, from 22 nt to ~120 nt or from 100 nt to ~120 nt, depending on the probe used to reveal the northern blot (i.e. specific to the ssDNA-biotin or to the SL-miRNA respectively) (Figure 18A and B). After the ligation reaction, a 21 nt band, which most probably corresponds to an incomplete by-product due to the excision of the adenine bi-phosphate by the RNA ligase, was still detectable (Figure 18.A).

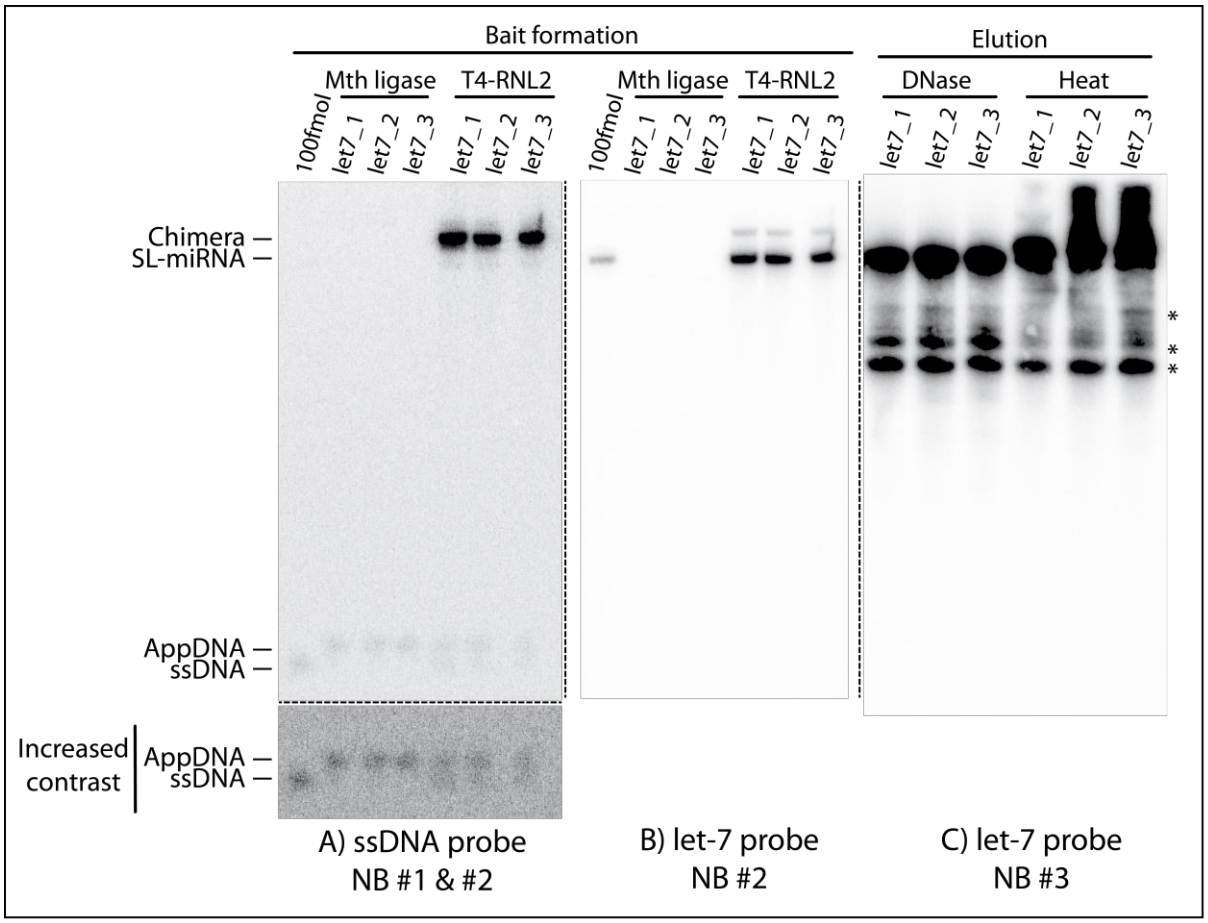


*Figure 18. Bait formation and elution controls. First panel highlights the pre-adenylation procedure done with the Mth RNA ligase (A) as well as the chimera formation done with the T4 RNL2 (A & B) and elution either done with Heat or DNase treatment (C) using northern blot analysis and a specific probe for the ssDNA-biotin or the SL-miRNA-let7. * represents SL-miRNA degradation by-products. NB #1 correspond to the first northern blot requires to control the pre-adenylation step. NB #2 correspond to the chimera northern blot control. NB #3 correspond to the elution northern blot control.*

Once formed, the chimeric RNA/DNA molecules were incubated with magnetic streptavidin beads. Chimera binding to the beads is an efficient process due to the high affinity between the biotinylated chimera and the streptavidin beads. We used a radiolabeled SL-miRNA to assess the binding efficiency using Cerenkov counting. In our hands, more than 95 % of the produced chimeras were retained on the beads. RNA chromatography was then performed with a nuclear extract during 45 min at room temperature followed by either a heat elution at 95°C during 5 min or a RQ1 DNase elution with 10 U of enzyme during 2 h. One percent (2.5 µL) of the eluates were analyzed by northern blot to validate the release of the SL-miRNA or the chimera from the beads with both approaches (Figure 18.C). The radiolabeled chimera mentioned earlier allowed us to determine the elution efficiency that reaches 45 %. Since the SL-miRNA or the chimera were uncoupled from the bait during the elution step (Figure 18.C), it seems that, at least, the elution worked. Therefore, the remaining part of the eluates can be analyzed by mass spectrometry.

III.A.3.d Mass spectrometry data analysis

A unique normalized quantitative dataset has been obtained for all the conditions analysed. Thus, if a given protein is identified by only one sample, it generates a line in the dataset table in which not available (NA) values will be added for all the other samples. These NA values represent the main issue for mass spectrometry data analysis. They can come from the absence of a given protein identification within one or all replicate of a given condition, respectively called partially observed values (POV) and missing on entire condition values (MEC). NA values can also be produced by technical issues, such as peptide ionization trouble, HPLC separation trouble or identification defect. POV are problematic for the statistical relevance of a protein identification and MEC values, for their part, do simply not allow condition comparison. Thus, NA values will be imputed by an arbitrarily assigned quantitative value to perform a complete data analysis.

III.A.3.d.1.1 Filtration and imputation

Prior to missing value imputation, it is important to clean-up the dataset to avoid analysis of non-relevant identification. Therefore, contaminant proteins, reverse identified proteins and empty lanes commonly found in raw mass spectrometry data are removed from the dataset. In our analysis, the only proteins that are kept must respect one criterion: a given protein (line) has to possess a quantitative value in 2 out of 3 replicates of a given condition to be maintained as a line in the dataset (POV filtration). Indeed, we decided that only one quantitative value among the triplicate is not sufficient to be considered as a relevant identification. After applying this filter, kept lines should possess a maximum of 16 missing values in our dataset of 18 samples. Before starting to impute the NA values, descriptive dataset figures are generated to take a step back on the dataset and get some highlights on the chromatography success and sample differences. In our experiment, DNase and heat elution allowed the identification of respectively 896 and 1741 proteins.

Figure 19.A shows that missing values, notably the MEC values, are over-represented in the DNase conditions as suspected. Indeed, heat treatment elutes a greater number of proteins than the DNase elution, which results in missing values within the DNase conditions. This was also suggested at the end of the pulldown because the DNase sample protein concentration was lower than the heat one (data not shown). Nonetheless, it is important to note that heat eluted samples also contain missing values. Since the heat elution is supposed to detach all proteins from the bait, it might seem surprising to have missing values. The greater heterogeneity of the heat samples can be responsible for those missing values because low abundant peptides can be hidden by other highly abundant peptides during the mass spectrometry analysis. All those differences between conditions are also highlighted by the clustered heatmap (Figure 19.B) that reveals a great number of NA values within the DNase conditions (in white). Moreover, Euclidian clustering clearly shows that heat and DNase samples are different since samples from each elution condition cluster together. We can also see that prior to NA value imputation and within elution condition, the different baits tend to cluster as well, suggesting a correct pulldown efficacy. At this point, bait controls and the mass spectrometry dataset display good features: correct bait formation and elution, more than 800 proteins identified by samples, well balanced POV and satisfying Euclidian clustering before the imputation step.

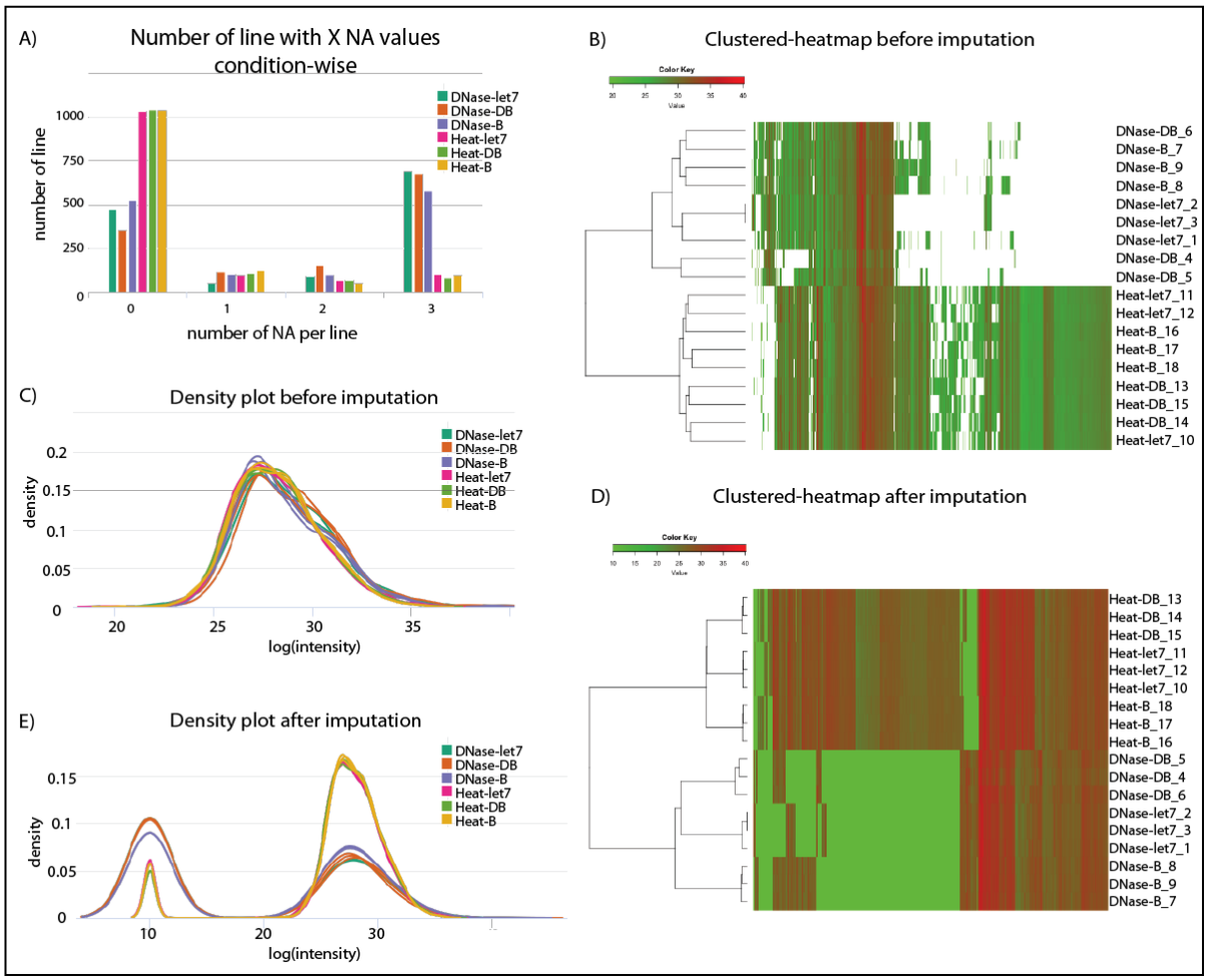


Figure 19. Overview of the matrix dataset before and after the imputation procedure. Prior to any data imputation: A) displays an overview of the not-available values per condition among the triplicate. B) represent an Euclidian clustered heatmap of the different bait conditions. White areas show missing values. C) displays the density plot of the global protein intensity among conditions. After the imputation of the partially observed values (POV) and the missing on entire condition values (MEC): D) represent an Euclidian clustered heatmap of the different bait conditions. E) represents the density plot of the global protein intensity among conditions after the complete imputation step.

Now that the matrix has been cleaned-up, data imputation can occur. ProStaR offers several methods to impute the POV depending on the generated raw dataset. Here, we used the "Structured Least Squares Algorithm" (SLSA) method, an adapted version of the LS impute algorithm (Bo, 2004). This function imputes the NA in a condition-wise manner only for the POV using the present values in a given condition as a template and a statistical normal distribution. For the MEC NA values, we chose to impute a fixed value below the lower log(intensity) present in the matrix. Here, the density plot on the Figure 19.C shows that first the dataset is normalized, as the log(intensity) of the different samples stack with each other and secondly that all samples display a log(intensity) above 15. Thus, we chose to impute a log(intensity) of 10 for the MEC NA-values (Figure 19.E). This allows us to obtain an important enrichment ratio if we compare a protein identified in a condition and not in another. It is

important to note that the significance of the mean difference between conditions calculated afterward will be impacted by this assigned MEC imputation. Once the whole matrix is filled, a new clustered heatmap is generated (Figure 19.D). Again, as on the density plot, we can clearly see that DNase samples had a large number of MEC values imputed (Figure 19.D, in green, and E). Surprisingly, the DNase-DB samples show more imputed values than the DNase-let7 and DNase-B ones (Figure 19.E). We would expect to see more MEC imputations within the DNase-B than the DNase-DB and finally the DNase-let7 samples since the baits share constituents and become more complex, thus supposedly interact with more factors. This result could be explained by the steric hindrance of the ssDNA or the SL-miRNA on the beads limiting the binding of bead specific proteins upon the chromatography. Concerning the Euclidian clustering, we can now see that each triplicate condition (Heat-let7 for example) groups perfectly compared to the non-imputed dataset (compare Figure 19.B and D). We can also see that heat samples form a more Euclidian clustered group than the DNase one, as the branch lengths of the clustered network are smaller, suggesting that the three conditions are less distinct. All conditions cluster better after the imputation which is expected since the SLSA method generates values in close proximity of the "template" value thus increasing triplicate correlation. This is why data imputation should be considered with caution as it smoothens out sample differences. At this point, the descriptive analysis on the imputed dataset can be done to compare chromatography approaches prior to explore the data in detail. Here, it's important to precise again that the samples DNase-let7_2 and DNase-let7_3 are identical due to a data duplication. This duplication is visualized by the absence of branch on the Euclidian cluster.

III.A.3.d.1.2 Descriptive analysis

Before going into a detailed analysis of the proteins identified in each condition, we look for a correct correlation among replicates as well as a decrease of correlation between bait conditions. In addition, we look for potential factors explaining or not these correlations in order to critically assess the pulldown elution approaches. To do so, we generated a correlation matrix and performed a principal component analysis (PCA). First, the correlation matrix obtained after imputation (Figure 20.A) reveals that all triplicate correlations within each heat-eluted bait condition seem correct (~1). Due to important differences between the DNase and Heat samples (correlation ~0.12), a "perfect" correlation is obtained for a given replicate conditions. Nevertheless, correlation between the different -DB, -B or -let7 bait conditions for the DNase and heat elution seems similar. In addition, we can see that among DNase samples, DNase-let7 replicates are more correlated between themselves (~1) than with -DB (~0.83) and with -B (~0.81). As mentioned earlier, this result makes sense because SL-miRNA conditions are closer to the -DB conditions due to the shared ssDNA oligo.

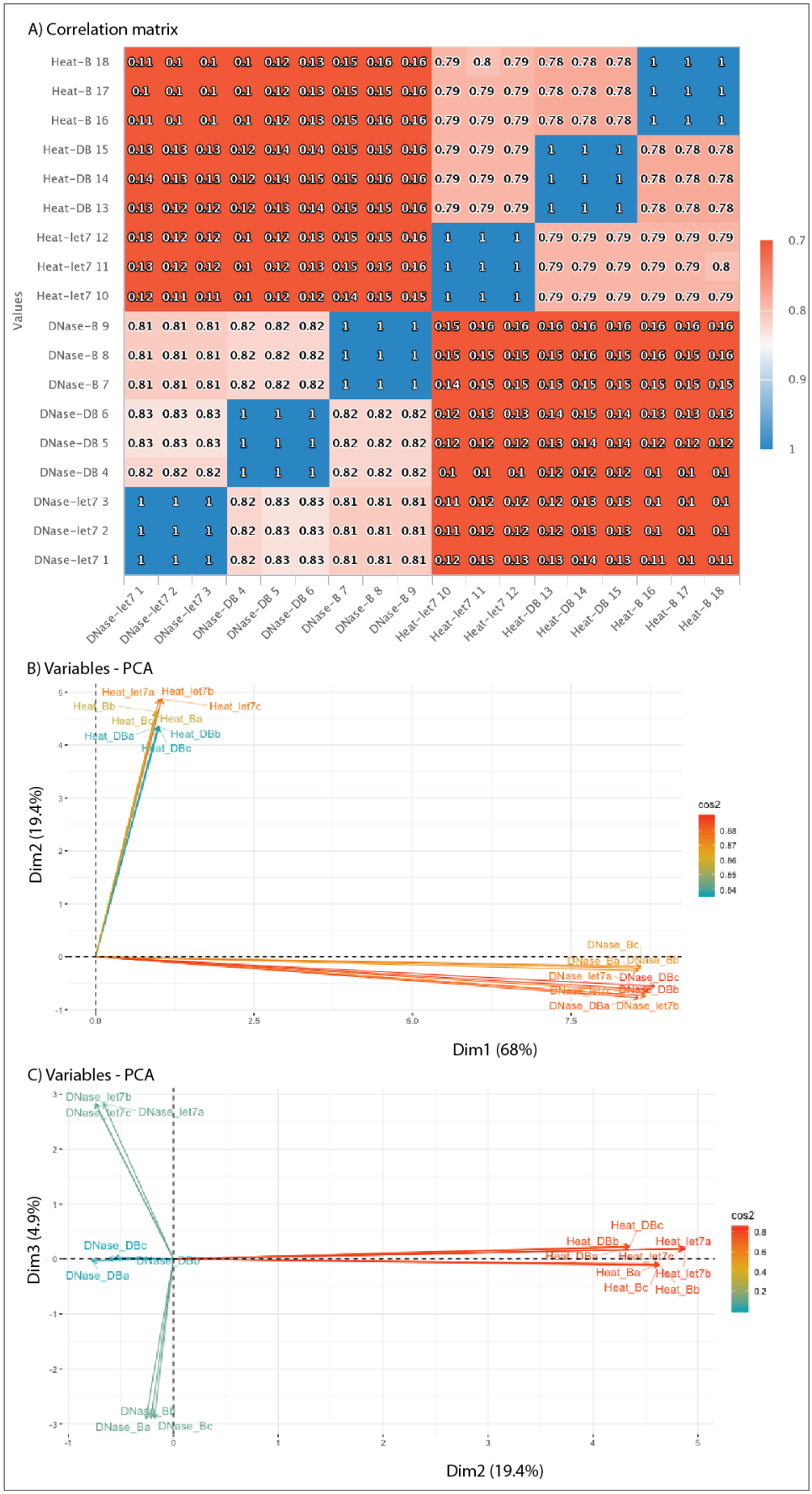


Figure 20. Descriptive analysis of the dataset. (A) Correlation matrix of the chromatography baits. The two-colors gradient represents the correlation value (B) Principal component analysis (PCA) with the first two dimensions. (C) Principal component analysis with the second and third dimensions.

Another way to assess the correlation or differences between samples, is to perform a PCA. PCA converts a set of observations supposedly correlated into a set of values of linearly uncorrelated variables called principal components. The purpose is to reveal data-driven factor that best explains data correlation. Here, PCA using two dimensions (Figure 20.B) shows that heat samples group together and are different from the DNase ones, indicating that the elution method is the main factor explaining the differences between samples. If we look on the second and third dimension (Figure 20.C), we can see that within the DNase samples, the bait conditions can be a factor explaining the observed differences, whereas the heat samples stay grouped regardless of the bait. In addition, within the DNase samples, the bead-baits (-B) seem more distant from the ssDNA-bait (-DB) and the let-7-bait (-let7), as the clustered-heatmap and correlation matrix suggested. This result, together with the clustered heatmap, clearly emphasizes that the DNase elution is more appropriate, or at least better, to generate differences between bait condition eluates than the heat one. Therefore, in this assay, DNase elution displays a better profile than the heat elution in terms of sample correlation and background noise. In addition, upon DNase treatment, bait conditions are considered as an important data-driven factor explaining the sample correlation.

At this point though, we do not know yet if the DNase and the heat treatments are able to retrieve potential good candidate proteins. To answer this question, we have to perform a differential analysis, first by comparing baits within a given elution condition, then similar baits upon different elution approaches to finally produce differentially enriched protein datasets depending on the bait and the elution. We will then be able to compare these enriched datasets to each other to determine if the DNase treatment is more appropriate than the heat treatment to identify potential good SL-miRNA binding proteins. Again, since one DNase-let7 sample has been duplicated, the interpretation of all statistical tests will not be relevant and the obtained results are only indicative.

III.A.3.d.1.3 Differential analysis

In order to perform the statistical analysis of the differential enrichment between samples, we chose the Limma test provided by the ProStaR software (Ritchie *et al.*, 2015). The Limma test allows to make the same interpretation as an ordinary t-test with the exception that standard errors have been moderated across lines (intensity of protein identified), which effectively takes into account information from the ensemble of lines to aid with inference about each individual protein. The p-value is obtained from the moderated Limma test, usually after adjustment for multiple testing. Here, we chose the most standard form of adjustment: False Discovery Rate (FDR) which is based on the Benjamini and Hochberg's method (Burger, 2018). Once statistical tests have been calculated, we generated volcano plots to represent the differences between conditions with a fold change threshold of 2 (LogFC>1) and a p-value threshold of 5 % (Figure 21.A, B and C).

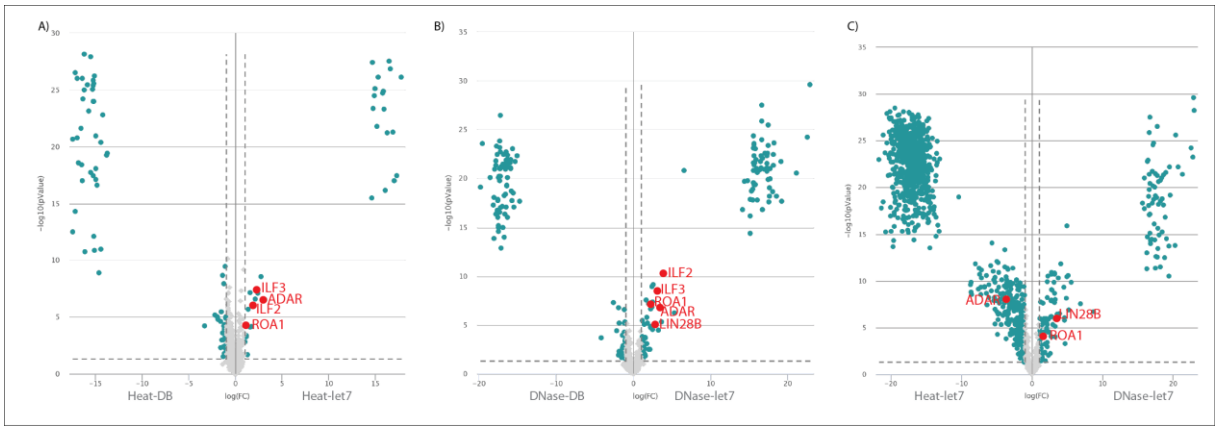


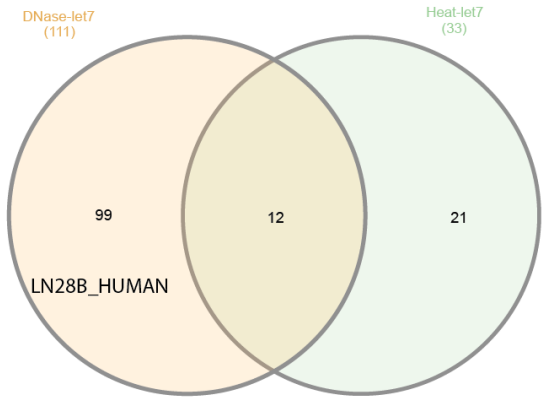
Figure 21. Differential analyses performed on the dataset. Volcano plots of the Heat-let7 vs. Heat-DB (A), DNase-let7 vs. DNase-DB (B) and DNase-let7 vs. Heat-let7 (C) comparisons using the Limma test and the Benjamini and Hochberg's method for p-value adjustment provided by ProStaR. Proteins differentially expressed (LogFC>1 and p-value < 0.05) are highlighted in blue.

For simplicity reason, we will only present here the differential analysis of the condition -let7 (SL-let-7a-1-ssDNA-bead) to the condition -DB (ssDNA-bead) (Figure 21.A, B, and C). We found that 111 proteins were specifically enriched in the DNase-let7 condition compared to the DNase-DB condition, whereas this number dropped to 33 in the heat-eluted -let7 vs. -DB samples. As expected, and suggested by the correlation matrix and the PCA, this number is lower due to the higher background noise shared among heat sample conditions. This does not mean that all those proteins are relevant potential SL-miRNA binders, but it surely provides a bigger candidate list. The maximal logFC enrichment observed, in the DNase-let7 vs. DNase-DB is 22.9, while it only reaches 17.7 with the heat samples, due to MEC imputation (Figure 21.A and B).

Finally, we can explore the list of identified proteins in each condition. Using the SL-let-7a-1 bait, we can identify a total of fifteen proteins that are shared between conditions (Figure 22.A and B). Among them, as expected, we find known microprocessor cofactors such as ILF2, ILF3 and DHX9 (Ha and Kim, 2014), which indicates that our in vitro SL-miRNA adopted a good pre-miRNA foldback structure. The proteins E2AK2 (PKR) (Clemens, 1997) and DSRAD (ADAR) (Knight and Bass, 2002; Chawla and Sokol, 2014), known to interact with dsRNA structures, are also enriched with both methods. It is important to specify that DROSHA and DGCR8 have never been identified in our raw dataset, potentially due to the processive characteristic of the enzyme, that rapidly detach from the SL-miRNA after the cleavage step. Other proteins such as SRSF9 (Treiber *et al.*, 2017), and MSI2H (Choudhury *et al.*, 2013), which are already known to be implicated or potentially involved in specific miRNA biogenesis, are also found in both conditions. Finally, we retrieved the already characterized pri-let-7a-1 binding proteins ROA1 (Michlewski and Cáceres, 2010) and DHX9 (Kawai and Amano, 2012) both in the DNase- and heat-eluted samples. However, another let-7 known interactant, LN28B, was only identified in the DNase treated let-7 samples (Figure 21.B). In addition, all the proteins are more enriched upon DNase elution compared to heat elution (Figure 22.B in green). Additional potential SL-miRNA binding proteins such as DHX28, DHX30, ROA3, ROAO, PTBP1, HNRH3, HNRL1, ZCCHV, NXF1, ALKB5, and HUR are

found among the other DNase-specific 111 proteins. We can therefore say that the number of enriched proteins and the greater protein signal over background noise for the specific SL-miRNA binders indicate that the DNase elution performs better than the heat elution for the discovery of proteins binding SL-miRNAs.

A)



B)

	LogFC		
Entry name	DNase-let7	Heat-let7	DNase/Heat
STRBP_HUMAN	21.16	2.70	7.83
E2AK2_HUMAN	19.24	1.68	11.46
CFA58_HUMAN	19.12	1.54	12.43
ZFR_HUMAN	18.53	1.03	17.97
MSI2H_HUMAN	17.36	15.13	1.15
ILF2_HUMAN	3.90	1.84	2.12
HNRH3_HUMAN	3.69	1.32	2.79
DSRAD_HUMAN	3.34	3.18	1.05
ILF3_HUMAN	3.18	2.41	1.32
SRSF9_HUMAN	2.54	1.33	1.91
DHX9_HUMAN	2.48	2.15	1.15
ROA1_HUMAN	2.41	1.13	2.13

Figure 22. Enriched protein dataset comparison. (A) Venn diagram of the significantly enriched proteins in the DNase-let7 and Heat-let7 datasets. (B) List of enriched proteins identified by both elution methods. Ratio in favor of DNase-let7 are highlighted in green.

III.A.3.d.1.4 Discussion and conclusion

RNA chromatography is a standard approach to discover RBPs interacting with a given RNA molecule. However, its inherent limitations require important downstream biological validations. Bait and RNP formation have to be performed in vitro with the risk that the folding will not be exactly similar to the in vivo one. In addition, post transcriptional modifications will not be present. It is also important to keep in mind that RNA molecules are very dynamic, and that various RNA processing steps occur simultaneously such as the DROSHA pri-miRNA cleavage which can happen concomitantly with transcription (Morlando *et al.*, 2008). Therefore, the chosen RNA bait will only represent a snapshot of a biological relevant RNA during its biogenesis. Because the chromatography in solution increases the non-biologically relevant RNP formation as well as the retention of protein binding to the chromatography matrix, it is important to work as much as possible with standardized cellular extracts to limit biological variations, which could have impact on bait correlations. Here, we propose an optimized approach of the chromatography elution as well as a data analysis procedure using ProStaR. The advantages of our method are i) the possibility to use different RNA baits thanks to the use of robust enzymatic reactions (Prongidi-Fix *et al.*, 2013), ii) controlled chimeric bait formation and elution steps, and iii) an elution method enabling the specific release of proteins interacting with the RNA of interest and at the same time decreasing the background noise. Although it is difficult and time-consuming to systematically assess the correct folding of the bait prior to the chromatography, we can confidently estimate the correct generation of the chimera and the correct elution of the RNA moiety by northern blot analysis. It is difficult to come up with positive controls allowing to estimate the quality of the pulldown before performing the mass spectrometry analysis. Increasing the washing stringency could help to

reveal such differences. However, it requires a time-consuming set up that we chose to avoid thanks to the DNase specific elution. For this reason, it is only after the mass spectrometry data generation that we can estimate whether it was successful. In this respect, the ProStaR software is of great help for the wet scientists to explore the sample profile and correlation as well as to perform differential analyses.

Descriptive analysis of the different bait conditions allowed us to really assess the advantages of the DNase treatment compared to the heat elution. In our dataset, all the DNase baits display a better profile in term of correlation between replicates and bait protein identification differences highlighted by the PCA. Thus, without looking at the identified proteins, it strongly emphasizes that the DNase treatment provides an increase of the protein identification specificity. More surprisingly, a greater number of enriched proteins were retrieved by the DNase elution compared to the heat one suggesting an increase of chromatography sensitivity. Concerning the identified proteins, our enriched datasets reveal a correct identification of global and specific miRNA cofactors whatever the elution method used. Nonetheless, heat elution was not able to identify LN28B protein. It is known that LN28B protein is not really abundant within lymphocyte cells (Yuan and Muljo, 2013). The ability of the DNase elution to identify this protein pinpoints the increased sensibility of the approach.

As the let-7a pri-miRNA has been studied a lot, we can compare our enriched DNase-let7 dataset with other pri-let-7a-1 pulldowns already published. As discussed above, it is complicated to compare different RNA chromatographies if the cell extract, the bait and the elution are different. For example, the Kim laboratory published a pri-let-7a-1 chromatography in which they identified LN28B, ILF3, DHX9, HNRPL, HRNPF, HNRH1, HNRH3, ROA0 and RPS2 (Heo *et al.*, 2008). Our DNase approach managed to identify all of them as significantly enriched. Another example comes from the Meister laboratory, where they chose to distinguish background binders and specific proteins by calculating a score correlated to the enrichment of a given protein associated to a specific SL-miRNA compared to all other SL-miRNAs used (Treiber *et al.*, 2017). With this approach they identified, among others, LN28B, hnRNPs, helicases and TUTases. Our DNase approach managed to also identify most of these proteins, with the exception of the TUTase proteins.

In conclusion, we present here an improved RNA chromatography using a DNase elution treatment that is more sensitive and specific than a classical heat elution approach. The improved specificity was shown using descriptive analysis with clustered-heatmap, correlation matrix and PCA. The improved sensibility was determined by comparing enrichment ratio of specific binders first within each elution method and then by comparing them. In addition, we proposed a mass spectrometry data analysis pipeline, using the user-friendly ProStaR software, allowing a powerful dataset management, as well as descriptive and differential analyses. Our approach can be improved in the future by the assessment of a better saturation of the beads with the chimera. Indeed, a weak or strong steric hindrance induced by the chimera on the beads can either increase or decrease both the analytical specificity and sensibility. Different ssDNA sequences or lengths as well as DNase enzymes could potentially improve the elution efficiency thereby increasing even further our method's specificity and sensibility. Our method is working with miRNA precursors and can surely be adapted to other RNA molecules.

III.A.4 DNase assisted RNA chromatography with KSHV SL-miRNA

Now that the DNase assisted RNA chromatography is set up, I can use this approach with all the different KSHV SL-miRNAs. As always, positive (SL-miRNA-let7) and negative controls are included. Here, the negative controls correspond to the ssDNA fixed to the beads, labelled -AB, and the beads only, labelled -B. Enrichment ratio will be calculated by comparing a given SL-miRNA bait with the -AB negative control. Due to the large number of conditions, we were not able to perform the pulldown in triplicate with the 10 different KSHV SL-miRNAs (from -K1 to -K9 plus -K11) and the two cellular SL-miRNAs (-let7, -155) from the same nuclear extract. Indeed, we could not perform pulldowns with more than 30 baits at once to ensure that the experiment could be handled correctly in terms of incubation, wash and elution times. To overcome this issue, we split the number of samples in four groups that could each be performed during one day. The pulldowns were performed from BC-3 cells nuclear extracts, labelled as Day1 to Day3. In addition, baits -K11, -155 as well as the controls were also used to do a pulldown from a DG75 cells nuclear extract. Since we used BC-3 nuclear extracts prepared independently for each day, it will be difficult to compare the protein identification coming from different days, due to the inherent variability of cell culture and extract preparation. Nevertheless, to ensure a good reproducibility, identical baits are pooled together and split prior to the chromatography to guarantee a good bait homogeneity among triplicate. Bait formation and elution are controlled by northern blot analysis (Figure 23). Protein identification was based on Mascot-LFQ technology (Cox *et al.*, 2014) and done by the mass spectrometry platform as well as the samples normalization. Then, I took care of the data mining with either the help of the laboratory bioinformatician, B. Chane-Woon-Ming, or the ProStaR mass spectrometry data manager software (Wieczorek *et al.*, 2017, 2019) using a determined sample labelling displayed in Table 3.

Table 3. RNA chromatography sample labelling. RNA chromatography is performed with each bait in triplicate.

	Baits						
Day1	-AB	-B	-let7	-K1	-K2	-K3	-K5
Day2	-AB	-B	-let7	-K6	-K7	-K8	-K9
Day3	-AB	-B	-let7	-K4	-K11	-155	
DG75	-AB	-B	-let7	-K11	-155		

In the following sections, I will present all the bait formations and elution controls (part III.A.4.a). Then I will present the data mining procedure done with the bio-informatician (part III.A.4.b.1.1) for each pulldown day. Finally, I will present another procedure of data mining performed with the ProStaR software (part III.A.4.b.1.2).

III.A.4.a Bait formation and elution control

As mentioned earlier, controlling the elution step is fundamental to validate the chromatography approach. This was done by northern blot analysis as previously mentioned. As can be seen in Figure 23, it appears that there were reveals some potential issues during the pulldown assays. First, chimera molecules are not always detected, which could be due to a

detection issue during the northern blot procedure. It is not necessarily worrying if the SL-miRNA can be detected in the DNase lane. Another issue was that some of the baits (-K5, -K9) were poorly or simply not eluted. Finally, the first replicates of the -K6, -K11 and -155 baits do not give any signal (Figure 23.B.C), which could be a gel loading issue. These problems will have to be kept in mind for the downstream mass spectrometry analysis, which was performed for all samples.

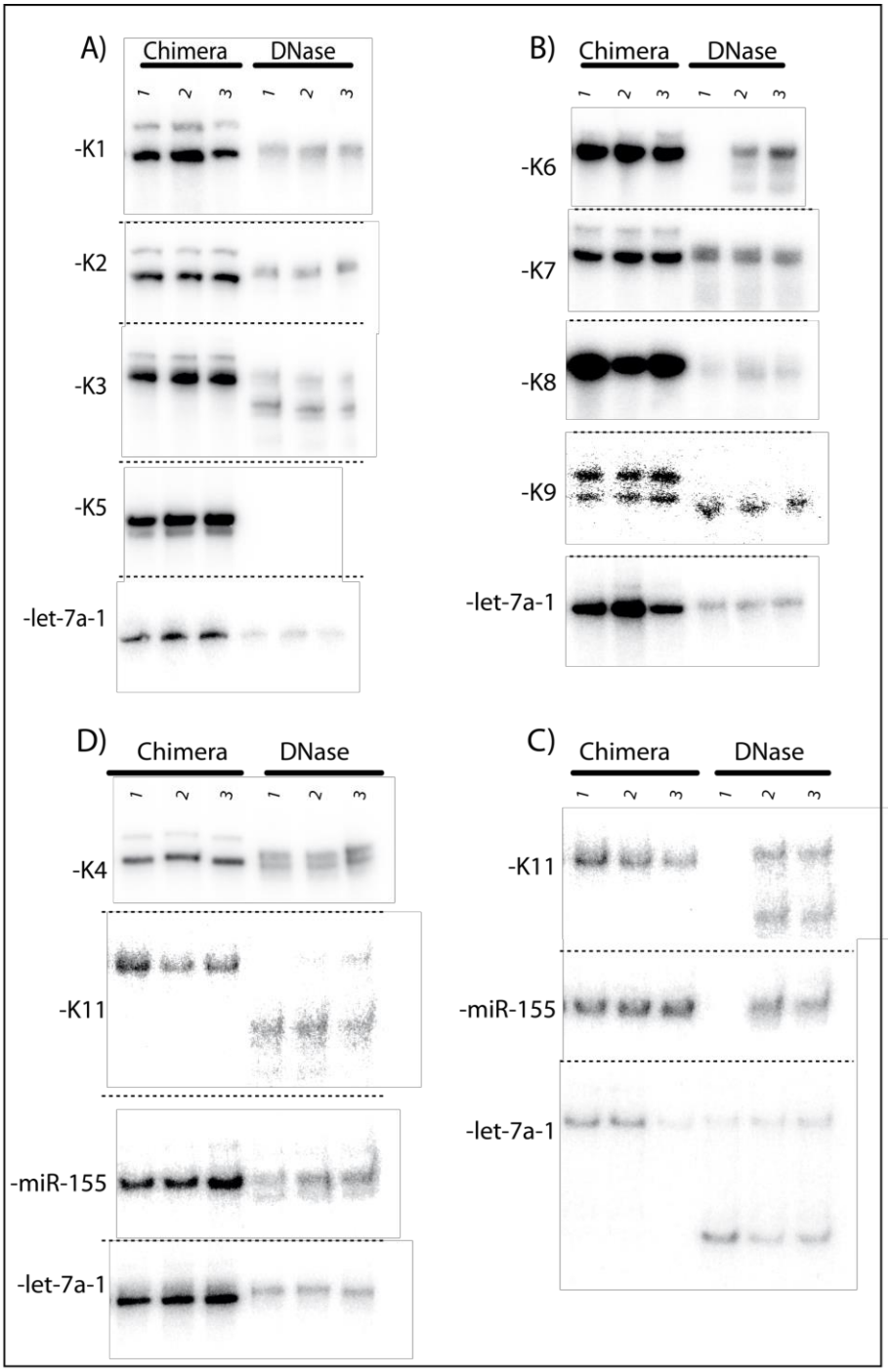


Figure 23. Bait formation and elution controls. (A), (B), (C) and (D) represent northern blot analysis of triplicate bait formation and elution. A) Bait used for the first chromatography within BC3 nuclear extract. B) Bait used for the second chromatography within BC3 nuclear extract. C) Bait used for the third chromatography within BC3 nuclear extract. D) Bait used for the fourth chromatography within DG75 nuclear extract.

III.A.4.b MSMS data exploration

The first data mining procedure was performed with a tool generated in-house by B. Chane-Woon-Ming (B.CWM_MSMS). This program follows a similar workflow as the ProStaR software, the main difference between the data mining procedures concerns the imputation method for the partially observed values (POV). In this B.CWM_MSMS "home-made" analysis, the only imputation performed on the matrix was for the POV for which average signal based on the available values were imputed. For the missing on an entire condition values (MEC), all the "0" values were simply kept. I will briefly expose this first data mining approach before going into the more detailed ProStaR analysis, as it was useful to get an idea of the quality of our data and helped us to get a first list of potential candidates.

III.A.4.b.1.1 B.CWM_MSMS "home-made" analysis.

Data filtering was the first step of the in-house data mining. Contaminant and reversed identified protein were excluded from the matrix. Then, no imputation of the MEC values was performed, meaning that for each bait triplicate the "0" value present within the dataset were kept. For the POV, no statistical approaches such as SLSA or KNN, were used but instead an average signal based on the two available replicate data was imputed. To produce the differential analysis, several statistical tests have been used. First ANOVA test are done (McHugh, 2011), and on the generated ANOVA residues the normal distribution of each protein is assessed by a Shapiro test and the homoscedasticity by a Bartlett test (Mcdonald). All proteins were kept for the differential analysis, performed by a Dunnett test, whatever the validity of the Shapiro and Bartlett test. Information concerning the proteins that are normally distributed and homoscedastic or not are used after the differential analysis to criticize the relevance of the obtained mean difference between bait identity. To performed the differential analysis, we select a multiple comparison test: the Dunnett test (Shun *et al.*, 2003). Dunnett test, corrected by the Benjamini-Hochberg method (FDR), tells us if two different samples show a significative mean difference (Burger, 2018). Finally, enrichment ratios are generated for each comparison. All these tests allow us to produce heatmaps highlighting both enrichment ratio and differential p-values to explore the mass spectrometry results. As mentioned, normal distribution and homoscedasticity are then check for the most interesting candidates. It is important to precise that the different statistical tests (ANOVA, Dunnett, Benjamini Hochberg) were performed on each dataset corresponding to the cell line used as well as the experiment (Day1 or Day2…).

With these p-values heatmap representations (Figure 24), we can have a global overview on the dataset in terms of enrichment, mean significative difference and global or specific protein identification. Here, each p-values heatmap have been sorted for all the corresponding SL-miRNA baits. For example, the first pulldown (Day1: Figure 24.A) clearly shows a very weak protein enrichment with no significant data for the -K5 conditions. It was expected since the northern blot for the -K5 bait displayed no elution signal. However, the cell extract seems good because in all the other Day1 conditions, baits were able to enrich protein from the background noise with significant mean differences. Heatmaps corresponding to the second pulldown (Day2: Figure 24.B) displays lower enrichment and almost no significant mean differences even though northern blot (Figure 23) shows a correct DNase elution. A nuclear extract issue can be responsible of this disappointing result, thus increasing POV. For the Day3

(Day3: Figure 24.C) and DG-75 (DG75: Figure 24.D) pulldowns, heatmaps display important enrichment ratio as well as important mean significative difference.

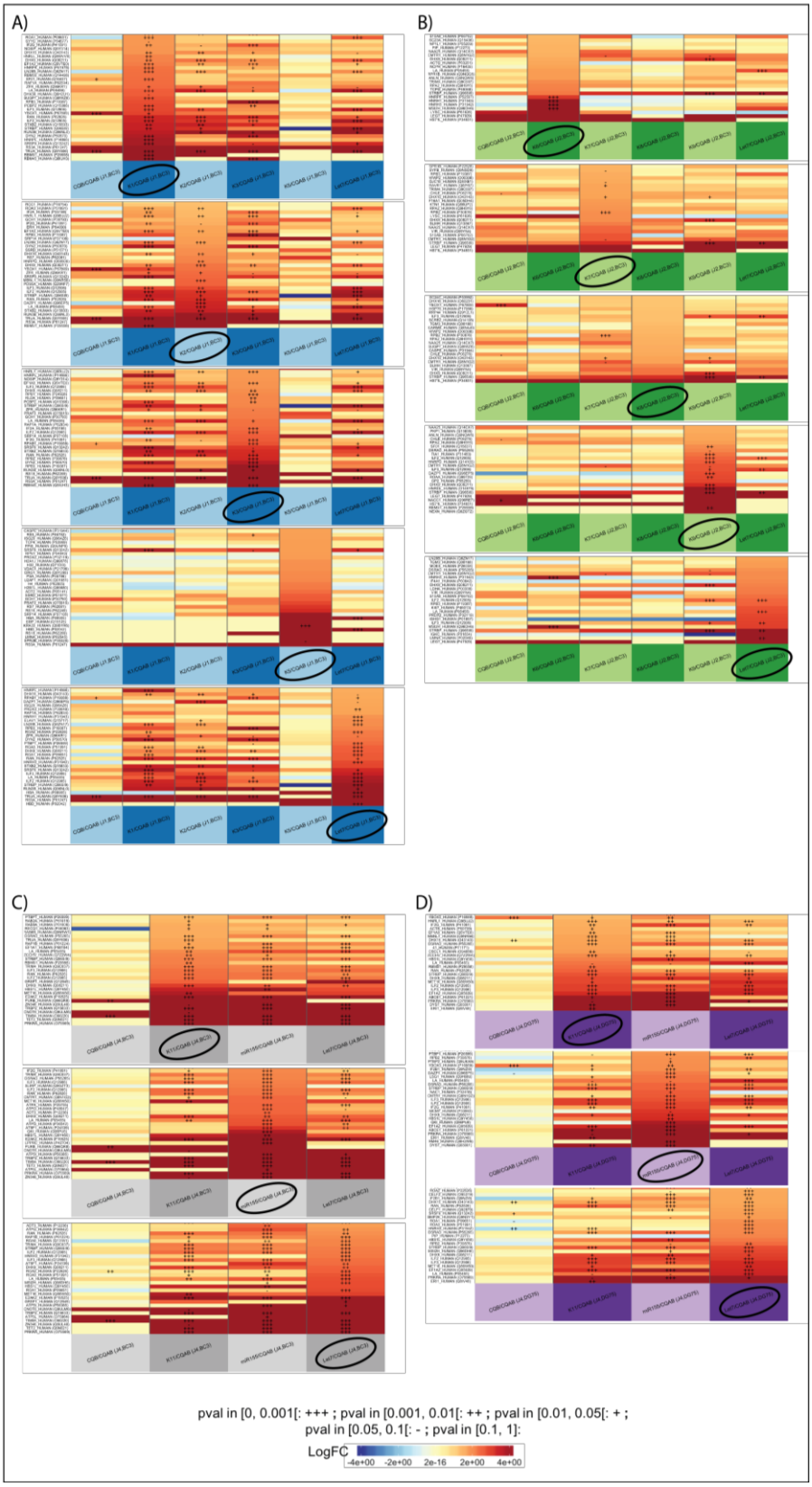


Figure 24. p-values-heatmap of the different RNA chromatography. Here, each heatmap is crop and sort for each bait used. Sorted condition is surrounded by a black line. Heatmap colour gradient represent the fold change enrichment of a given protein (line) by a given bait (column) compare to its control - AB. p-values are represented by + and – signs as followed: pval in [0, 0.001[: +++ ; pval in [0.001, 0.01[: ++ ; pval in [0.01, 0.05[: + ; pval in [0.05, 0.1[: - ; pval in [0.1, 1]:. A) p-values heatmap of the Day1 chromatography. B) p-values heatmap of the Day2 chromatography. C) p-values heatmap of the Day3 chromatography. D) p-values heatmap of the DG75 chromatography.

All the chromatographies are validated by the positive control -let7 except Day2. Indeed, microprocessor's cofactors ILF2 and ILF3 as well as the Let-7a specific cofactors LIN28B and ROA1 are significantly enriched and are part of the top hit list. For Day2, only the ILF2 and ILF3 proteins are significantly enriched. The other proteins, LIN28B and ROA1 display a good enrichment but with a p-values coming from the Dunnett-FDR test above 10%. Our data mining approach was overall good and allowed us to obtain interesting enriched dataset. However, with this approach descriptive analysis statistical POV imputation and MEC imputation are absent. This is why I decided to re-explore these datasets with the newly available version of the ProStaR software (ProStaR version 1.16.6; DAPAR version 1.16.5). The analysis with ProStaR allowed me to generate more comparison, imputation and descriptive statistical analysis as described afterward. With this ProStaR software, POV are imputed using a statistical approach (SLSA) and a fixed value was selected to impute the MEC.

III.A.4.b.1.2 ProStaR data mining

In this part, I will first discuss the handling of missing values as well as their distribution. Then Euclidian clustering heatmap will be generated before and after the imputation procedure. Finally, after the imputation, descriptive analysis using correlation matrix and PCA will be exposed. Once the descriptive analysis validates the chromatography, differential analysis is performed using a Limma test corrected with the Benjamini-Hochberg method in order to generate significantly enriched protein datasets (Ritchie *et al.*, 2015). Since mass spectrometry analysis always generate missing values, I performed the same filtration step for all the datasets. In order to be more exhaustive, I decided to only work with proteins that have been identified at least 2 times in a given triplicate. The data will be presented separately for each pulldown day.

III.A.4.b.1.3 Day1: SL-miRNA-K1, -K2, -K3, -K5, -let7

For the Day1 RNA chromatography, we used 7 different conditions in triplicate. As mentioned, positive (-let7) and negative (-AB, -B) baits are used. For this experiment, 849 different proteins have been identified (Table 4). The number of identified proteins is lower than the DNase samples (>1200 proteins) used to validate the DNase elution method (see part III.A.3).

Table 4. Mass spectrometry protein identification overview of the first chromatography performed with BC3 nuclear extract (Day1). Seven baits condition have been used (-B, -AB, -let7, -K1, -K2, -K3, -K5).

Definition	Value
Number of samples	21
Number of conditions	7
Number of lines	849
Number of missing values	2635
% of missing values	14.78
Number of empty lines	14

Obtaining a short list of proteins could be either a good or a bad sign. Low number of proteins can come from a protein sample with a low background noise, which is good, or from a non-efficient pulldown elution which is not good at all. However, missing values are well distributed among the conditions, suggesting a similar protein identification efficiency among conditions (Figure 25.A). The mean density of the log(intensity) for these samples is around 25 (Figure 25.C) and display a symmetric normal distribution. We can see on the density-plot that our dataset has been normalized since all the signal density conditions show similar profiles. Prior to the imputation methods, Euclidian clustered heatmap reveals disparity between triplicate condition sample, suggesting that the NA values greatly impact triplicate clustering (Figure 25.B). Since technical bias can come both from the chromatography and the mass spectrometry identification, it is difficult at this point to say that our experiment is poorly reproducible. As mentioned earlier for the DNase vs. Heat elution comparison, data imputation, no matter how dangerous it is, will smoothen out this suggested weak reproducibility.

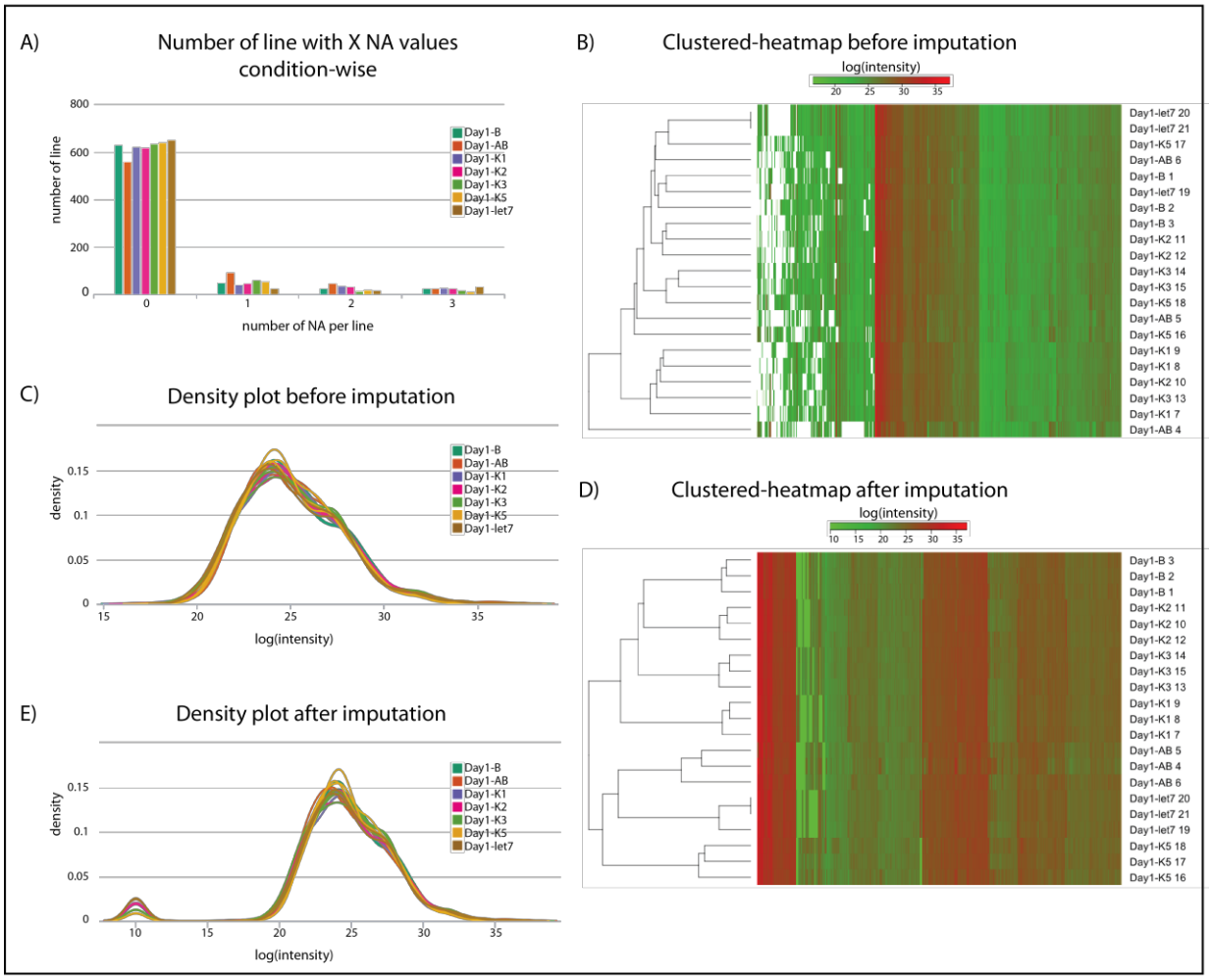


Figure 25. Overview of the Day1 matrix dataset before and after imputation procedure. Prior to any data imputation, (A) displays the non-available values among the matrix, (B) represent the Euclidian clustered heatmap, (C) shows the density plot of the global protein intensity among conditions. After the imputation of the partially observed values (POV) and the missing on entire condition values (MEC), (D) displays an Euclidian clustered heatmap, (E) shows the density plot of the global protein intensity among conditions.

As previously, I imputed the partially observed value using the SLSA algorithm. Then, I imputed MEC proteins using a fixed value of 10, corresponding to a value below the minimal log(intensity) globally observed on the matrix (Figure 25.E). Once the imputation is done, we can observe a better clustering of the triplicates that group with each other (Figure 25.D). Indeed, clustering display a relatively small dendrogram branches, highlighting the imputation of SLSA algorithm in close proximity to the template value among a condition We should note here that the DNase-let7 sample 20 and 21 are identical, these is due to the artificial duplication of sample 20 because the file for sample 21 was corrupted as mentioned earlier in the manuscript. Thus, correlation matrix and PCA analysis should display a perfect correlation between the DNase-let7 sample 20 and 21.

The other triplicate conditions show a very good similarity (Figure 26.A). The minimal correlation between conditions is around 0.7 which is expected since all the baits should possess a similar background noise and different protein interactants depending on the SL-miRNA. PCAs, for their part, clearly demonstrate that the differences between conditions come from the bait identity (Figure 26.B.C). However, it is clear that the -K5 baits pulldown did not work correctly, as expected from the northern blot analysis. Indeed, it is very close to the -AB condition on the PCA. I want to highlight the fact that the -K1 and -K2 baits display the better profile in this experiment and this result also seems correlated to the northern blot analysis. It is therefore clear that northern blot control seems to give a good insight on the pulldown efficiency. Overall descriptive analysis of this experiment is good to perform the differential analysis.

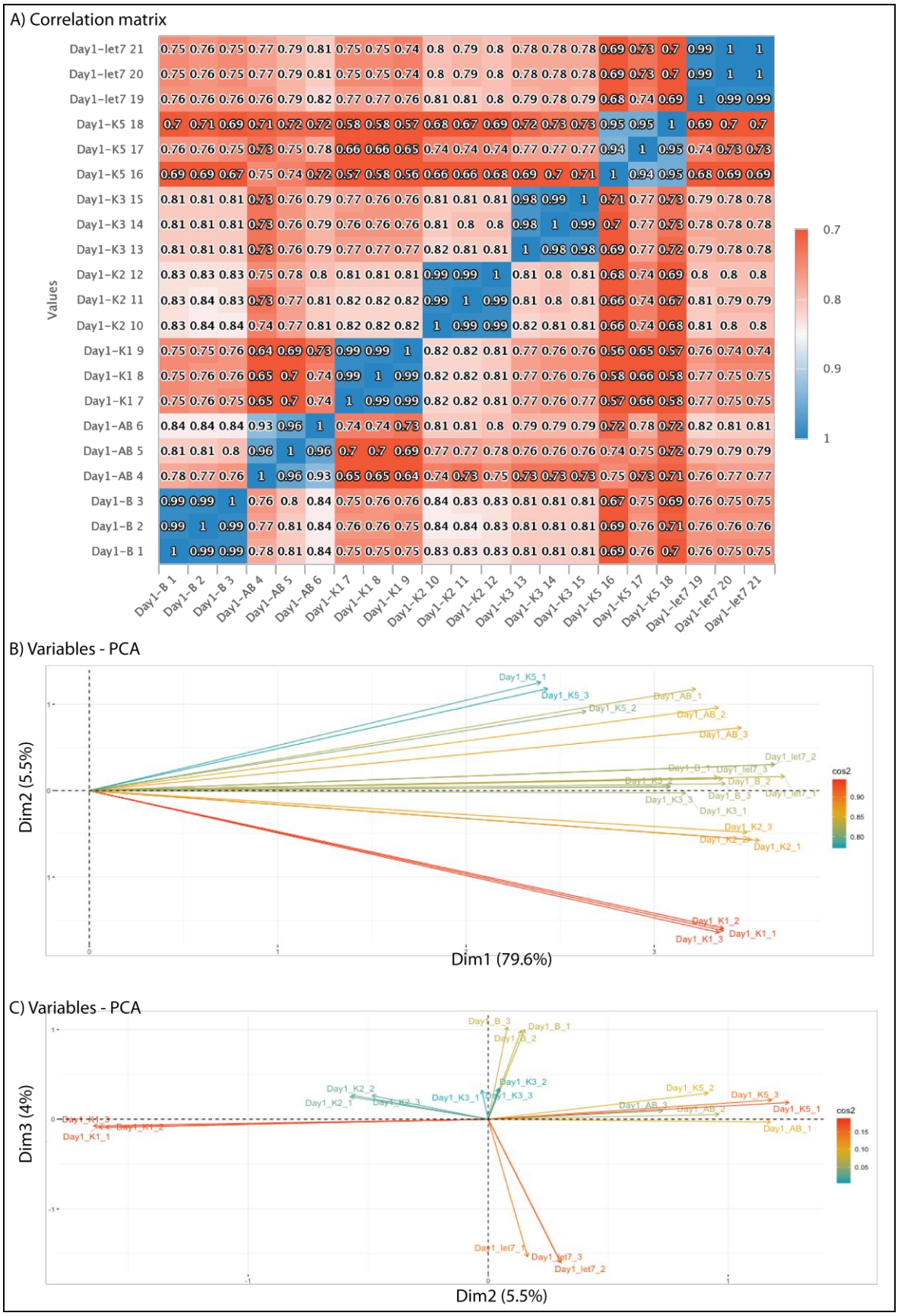


Figure 26. Descriptive analysis of the Day1 chromatography dataset. (A) Correlation matrix of the chromatography baits. (B) Principal component analysis (PCA) with the first and second dimensions. (C) Principal component analysis with the second and third dimensions.

I performed the same differential analysis as in the DNase vs -Heat part, namely: Limma statistical test corrected with Benjamini and Hochberg's method. Once statistical test is performed, volcano plots are generated with a fold change threshold of 2 and a p-value threshold of 5 %. Enriched datasets are generated for all the SL-miRNA bait condition compared to the control ssDNA-bead (-AB). As always, let-7a-1 bait (Day1-let7) act as a positive control in which we should identify at least LN28B and ROA1. In our enriched dataset, let-7a-1 bait managed to significantly enrich LN28B, ROA1 and FUBP2, thus together with the descriptive analysis strongly validating the pulldown. Finally, each SL-miRNA enriched datasets from the same chromatography are compared using Venn diagram. Global lists of enriched protein are also generated (Figure 27.A.B). To facilitate the analysis, specific or shared proteins between baits can be explored by applying filter on the matrix. Here, it is clear that the -K5 chromatography was the worst condition in this assay. Indeed, for this SL-miRNA, even the cofactors ILF2 and ILF3 are missing and an important number of ribosomal proteins are present.

Now, we can take a closer look on the cell specific and significantly enriched proteins by a given SL-miRNA (Figure 27.A.B). Here, I will only focus on proteins that are clearly known to be involved in RNA interaction or processing, which could represent relevant candidates.

-**SL-miRNA-K1** managed to specifically enrich, among others, the helicase DHX15 (Treiber *et al.*, 2017), a component of the tri-snRNP complex LSM2 (Catala *et al.*, 2019), a major nucleolar protein involves in pre-rRNA maturation NUCL (Woo *et al.*, 2017), a splicing protein SUGP1 (Treiber *et al.*, 2017), and TADBP (Bhardwaj *et al.*, 2013) a protein involves in various RNA biogenesis.

-**SL-miRNA-K2** managed to specifically enrich, among others, an AU-rich element binder HNRPD (Lee *et al.*, 2014), a splicing factor MBNL1 (Rau *et al.*, 2011) and an interactant of the HEXIM1-DNA-PK-paraspeckle components-ribonucleoprotein complex: XRCC6 (Treiber *et al.*, 2017).

-**SL-miRNA-K3** managed to specifically enrich, among others, three splicing factor MGN2, RU2B and SMD1 (Treiber *et al.*, 2017).

-Commonly identified by the different SL-miRNA: Among them we can cite FUBP2 (Michlewski and Cáceres, 2010), DHX15, DHX9 (Robb and Rana, 2007), YBOX1 (X.-J. Yang *et al.*, 2019) which have already been described to be involved in miRNA biogenesis (Figure 27.D).

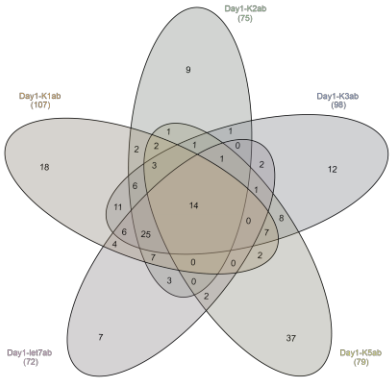
A)

SL-K1	SL-K2	SL-K3	SL-K5	SL-let7	Entry name
Enriched	Enriched	Enriched	NA	NA	3BP1_HUMAN
NA	Enriched	NA	Enriched	NA	ABC3A_HUMAN
NA	NA	NA	Enriched	NA	ABC3C_HUMAN
Enriched	NA	Enriched	NA	NA	ABCE1_HUMAN
NA	NA	NA	Enriched	NA	ADT2_HUMAN
NA	NA	NA	Enriched	NA	ANXA1_HUMAN
Enriched	NA	NA	NA	NA	AP2B1_HUMAN
NA	NA	NA	NA	Enriched	AP3B1_HUMAN
NA	NA	Enriched	Enriched	NA	AP3S1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	BAG2_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	CAH1_HUMAN
Enriched	NA	Enriched	NA	NA	CBX3_HUMAN
Enriched	NA	Enriched	NA	Enriched	CFA20_HUMAN
Enriched	NA	NA	NA	Enriched	CFA58_HUMAN
Enriched	NA	Enriched	NA	Enriched	CIRBP_HUMAN
Enriched	NA	Enriched	NA	NA	CMTR1_HUMAN
NA	NA	Enriched	NA	NA	COF1_HUMAN
NA	NA	Enriched	NA	NA	CPSF4_HUMAN
NA	NA	Enriched	NA	NA	CPSF5_HUMAN
NA	NA	NA	Enriched	NA	CREB1_HUMAN
NA	NA	NA	Enriched	NA	CSK23_HUMAN
NA	NA	NA	Enriched	NA	CTNB1_HUMAN
Enriched	Enriched	Enriched	NA	NA	CWC27_HUMAN
NA	Enriched	NA	NA	Enriched	DAZP1_HUMAN
NA	NA	NA	Enriched	NA	DDX5_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	DHX15_HUMAN
Enriched	NA	NA	NA	NA	DHX36_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	DHX9_HUMAN
Enriched	NA	NA	NA	NA	DJC10_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	DRG1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	DSRAD_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	DYN2_HUMAN
Enriched	NA	Enriched	NA	Enriched	E2AK2_HUMAN
NA	NA	NA	Enriched	NA	EBP_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	EF1A3_HUMAN
Enriched	NA	Enriched	NA	Enriched	EI2BB_HUMAN
NA	Enriched	NA	NA	Enriched	ELAV1_HUMAN
NA	Enriched	Enriched	NA	NA	ELMO1_HUMAN
NA	NA	Enriched	NA	Enriched	ELOB_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	ERH_HUMAN
NA	NA	Enriched	NA	Enriched	ERP29_HUMAN
Enriched	NA	NA	NA	NA	F16A1_HUMAN
NA	NA	NA	Enriched	NA	FABP5_HUMAN
NA	NA	Enriched	Enriched	NA	FEN1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	FKB11_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	FUBP2_HUMAN
NA	NA	NA	Enriched	NA	G3P_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	GCH1_HUMAN
NA	NA	NA	Enriched	NA	H2A1J_HUMAN
NA	NA	NA	Enriched	NA	H2B1L_HUMAN
NA	NA	NA	Enriched	NA	H32_HUMAN
NA	NA	NA	Enriched	NA	H4_HUMAN
NA	NA	NA	Enriched	Enriched	HBA_HUMAN
NA	NA	NA	Enriched	Enriched	HBB_HUMAN
NA	NA	NA	NA	Enriched	HBD_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	HBS1L_HUMAN
Enriched	NA	NA	NA	Enriched	HNRH1_HUMAN
NA	NA	NA	NA	Enriched	HNRH3_HUMAN
Enriched	Enriched	Enriched	NA	NA	HNRL1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	HNRLL_HUMAN
NA	Enriched	NA	NA	NA	HNRPD_HUMAN
Enriched	NA	Enriched	NA	NA	HNRPK_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	HNRPL_HUMAN
Enriched	Enriched	NA	Enriched	NA	HNRPQ_HUMAN
Enriched	Enriched	Enriched	NA	NA	HSP7E_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	IF2A_HUMAN
Enriched	NA	Enriched	Enriched	NA	IF2B_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	IF2G_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	ILF2_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	ILF3_HUMAN
Enriched	NA	Enriched	Enriched	NA	IMA1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	IRF4_HUMAN
Enriched	NA	Enriched	NA	Enriched	ISG20_HUMAN
NA	NA	NA	Enriched	NA	K22O_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	LA_HUMAN
NA	NA	NA	Enriched	NA	LEG7_HUMAN
NA	NA	NA	Enriched	NA	LMNA_HUMAN
Enriched	Enriched	NA	NA	Enriched	LN28B_HUMAN
Enriched	NA	NA	NA	NA	LSM2_HUMAN
NA	Enriched	NA	NA	NA	MAOM_HUMAN
Enriched	NA	NA	NA	NA	MAP2_HUMAN
NA	Enriched	NA	NA	NA	MBNL1_HUMAN
NA	Enriched	NA	NA	NA	MCM3_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	MCTS1_HUMAN
NA	NA	Enriched	NA	NA	MGN2_HUMAN
NA	Enriched	Enriched	Enriched	NA	MGST1_HUMAN
NA	NA	NA	Enriched	NA	MPCP_HUMAN
Enriched	NA	Enriched	Enriched	NA	MRGBP_HUMAN
NA	NA	NA	NA	Enriched	MSI2H_HUMAN
NA	Enriched	NA	NA	NA	NEPR1_HUMAN
Enriched	NA	NA	NA	Enriched	NIPBL_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	NOSIP_HUMAN
Enriched	NA	NA	NA	NA	NUCL_HUMAN
Enriched	Enriched	NA	NA	Enriched	PABP1_HUMAN
Enriched	NA	NA	NA	NA	PACS1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	PARVB_HUMAN

B)

SL-K1	SL-K2	SL-K3	SL-K5	SL-let7	Entry name
Enriched	NA	Enriched	NA	NA	PCBP1_HUMAN
Enriched	NA	Enriched	Enriched	NA	PCBP2_HUMAN
Enriched	NA	NA	NA	NA	PDIA6_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	PHB_HUMAN
Enriched	NA	NA	NA	NA	PMS2_HUMAN
NA	NA	NA	Enriched	NA	PPIE_HUMAN
NA	NA	Enriched	Enriched	NA	PRAF3_HUMAN
NA	NA	NA	NA	Enriched	PRDX3_HUMAN
Enriched	NA	NA	NA	Enriched	PROF1_HUMAN
NA	NA	NA	Enriched	NA	PRP8_HUMAN
NA	NA	NA	NA	Enriched	PRS6A_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	PTBP1_HUMAN
Enriched	NA	Enriched	Enriched	NA	QKI_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RAB21_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RAB2A_HUMAN
NA	NA	Enriched	NA	NA	RAB7A_HUMAN
NA	NA	Enriched	NA	NA	RAGP1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RAN_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RAP1A_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RB11A_HUMAN
NA	NA	Enriched	NA	NA	RBBP6_HUMAN
NA	NA	NA	NA	Enriched	RBM3_HUMAN
Enriched	NA	Enriched	NA	NA	RBM39_HUMAN
Enriched	NA	Enriched	NA	NA	RBM45_HUMAN
Enriched	Enriched	NA	NA	NA	RBMS1_HUMAN
Enriched	Enriched	NA	NA	NA	RBMS2_HUMAN
NA	NA	NA	Enriched	NA	RBMX_HUMAN
Enriched	Enriched	Enriched	NA	NA	RBP2_HUMAN
Enriched	NA	NA	Enriched	NA	RBP56_HUMAN
NA	Enriched	NA	NA	NA	RCC1_HUMAN
Enriched	NA	NA	NA	NA	RECQ1_HUMAN
NA	NA	NA	Enriched	NA	RFA1_HUMAN
NA	NA	NA	Enriched	NA	RFA2_HUMAN
NA	Enriched	NA	NA	Enriched	ROA0_HUMAN
Enriched	Enriched	NA	NA	Enriched	ROA1_HUMAN
Enriched	Enriched	NA	NA	Enriched	ROA2_HUMAN
Enriched	Enriched	NA	NA	Enriched	ROA3_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RPAB1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RPAC1_HUMAN
NA	NA	Enriched	NA	NA	RPB1_HUMAN
NA	NA	Enriched	NA	NA	RPB2_HUMAN
Enriched	NA	Enriched	NA	Enriched	RPB3_HUMAN
NA	NA	Enriched	NA	NA	RPB9_HUMAN
Enriched	NA	Enriched	NA	NA	RPP40_HUMAN
NA	NA	NA	Enriched	NA	RS11_HUMAN
NA	NA	Enriched	Enriched	NA	RS16_HUMAN
NA	NA	Enriched	Enriched	NA	RS18_HUMAN
NA	NA	NA	Enriched	NA	RS2_HUMAN
NA	NA	NA	Enriched	NA	RS23_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	RS3A_HUMAN
Enriched	NA	Enriched	Enriched	NA	RS5_HUMAN
NA	NA	NA	Enriched	NA	RS7_HUMAN
NA	NA	NA	Enriched	NA	RS9_HUMAN
Enriched	Enriched	NA	NA	Enriched	RT27_HUMAN
Enriched	NA	NA	NA	NA	RTF2_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RU2A_HUMAN
NA	NA	Enriched	NA	NA	RU2B_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RUN3B_HUMAN
NA	NA	Enriched	Enriched	NA	RUXE_HUMAN
NA	NA	Enriched	Enriched	NA	RUXF_HUMAN
Enriched	Enriched	NA	NA	Enriched	SCFD1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	SF01_HUMAN
Enriched	NA	Enriched	NA	NA	SF3A2_HUMAN
NA	NA	Enriched	NA	NA	SMD1_HUMAN
NA	NA	NA	Enriched	NA	SPR2B_HUMAN
Enriched	NA	Enriched	Enriched	NA	SRP09_HUMAN
NA	NA	NA	Enriched	NA	SRP14_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	SRSF9_HUMAN
NA	NA	NA	Enriched	NA	SSRD_HUMAN
NA	Enriched	NA	NA	NA	STAG2_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	STRBP_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	STXB2_HUMAN
Enriched	NA	NA	NA	NA	SUGP1_HUMAN
Enriched	Enriched	NA	Enriched	NA	SYDC_HUMAN
Enriched	NA	NA	NA	NA	SYFB_HUMAN
Enriched	NA	NA	NA	NA	SYYC_HUMAN
Enriched	Enriched	Enriched	NA	NA	T2EA_HUMAN
Enriched	NA	NA	NA	NA	TADBP_HUMAN
Enriched	NA	Enriched	NA	NA	TASOR_HUMAN
NA	NA	NA	Enriched	NA	TBA4A_HUMAN
NA	NA	Enriched	Enriched	NA	TCEA1_HUMAN
NA	NA	NA	Enriched	NA	TCP4_HUMAN
NA	NA	NA	Enriched	NA	THOC4_HUMAN
NA	NA	NA	Enriched	NA	THOC6_HUMAN
NA	Enriched	NA	NA	NA	THYN1_HUMAN
Enriched	NA	NA	NA	NA	TLN1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	TLN2_HUMAN
NA	NA	Enriched	Enriched	Enriched	TMEDA_HUMAN
NA	NA	NA	Enriched	NA	U2AF1_HUMAN
Enriched	NA	Enriched	NA	NA	U2AF2_HUMAN
Enriched	NA	NA	Enriched	NA	U520_HUMAN
NA	NA	NA	Enriched	NA	VDAC1_HUMAN
Enriched	NA	NA	NA	NA	WWP2_HUMAN
NA	Enriched	NA	NA	NA	XRCC6_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	YBOX1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	ZFR_HUMAN

C)



D)

Common proteins (minus -K5)		
BAG2_HUMAN	IF2G_HUMAN	RAP1A_HUMAN
DHX15_HUMAN	ILF2_HUMAN	RB11A_HUMAN
DHX9_HUMAN	ILF3_HUMAN	RPAB1_HUMAN
DSRAD_HUMAN	IRF4_HUMAN	RPAC1_HUMAN
DYN2_HUMAN	LA_HUMAN	RU2A_HUMAN
EF1A3_HUMAN	MCTS1_HUMAN	RUN3B_HUMAN
ERH_HUMAN	NOSIP_HUMAN	SF01_HUMAN
FKB11_HUMAN	PARVB_HUMAN	SRSF9_HUMAN
FUBP2_HUMAN	PHB_HUMAN	STRBP_HUMAN
HBS1L_HUMAN	PTBP1_HUMAN	STXB2_HUMAN
HNRLL_HUMAN	RAB21_HUMAN	TLN2_HUMAN
HNRPL_HUMAN	RAB2A_HUMAN	YBOX1_HUMAN
IF2A_HUMAN	RAN_HUMAN	ZFR_HUMAN

Figure 27. Differential analysis of the Day1 dataset. (A) and (B) display the bait enriched protein compare to the control -AB selected from volcano plots with a fold change and p-values threshold of respectively 2 and 5 %. C) Venn diagram of the bait enriched proteins. D) Common protein enriched minus -K5 bait.

III.A.4.b.1.4 Day2: SL-miRNA-K6, -K7, -K8, -K9, -let7

For the second day of RNA pulldown, we used 7 different conditions in triplicate. As always, positive (-let7) and negative (-AB, -B) baits are used. For this experiment, 944 different proteins have been identified containing ~17 % of NA values which is a bit more important than the first day pulldown (~15%) (Table 5).

Table 5. Mass spectrometry protein identification overview of the second chromatography performed with BC3 nuclear extract (Day2). Seven baits condition have been used (-B, -AB, -let7, -K6, -K7, -K8, -K9).

Definition	Value
Number of samples	21
Number of conditions	7
Number of lines	944
Number of missing values	3420
% of missing values	17.25
Number of empty lines	12

This augmentation of the NA values in this dataset comes from the -K6 bait that clearly show fewer line with 0 NA values compare to the other conditions (Figure 28.A).This can come from the absence of the SL-miRNA detection after the elution on the northern blot (Figure 23) which is also highlighted first by an increased number of line possessing one NA values among the -K6 samples and second by the clustered heatmap in which sample Day2-K6 (7) shared a distant node with the other samples (Figure 28.B). The mean density of the log(intensity) for these samples is around 25 (Figure 28.C) which correspond to the previous pulldown experiment suggesting a potential similar cell extract used in terms of protein quantity and quality. Prior to the imputation methods, Euclidian clustered heatmap reveals disparity between triplicate condition sample as in the previous experiment (Figure 28.B). Here it is even more striking for the -K6 bait conditions. As previously, I imputed the POV using the SLSA algorithm, then the MEC using a fixed value of 10, corresponding to a value below the minimal log(intensity) globally observed on the matrix (Figure 28.E). After the imputation (Figure 28.D), we can see a better clustering of the triplicates that group with each other (Figure 28.D). Again, triplicate conditions share the same dendrogram node and display homogeneous dendrogram branch length highlighting the ability of the SLSA method to impute POV values in close proximity to the template value. As previously observed after the imputation step, correlation matrix shows a good similarity between bait conditions (Figure 29.A). Here, the minimal correlation between sample is around 0.8 suggesting a more abundant background noise compared to Day1 chromatography. PCAs, for their parts, demonstrate that these conditions differences come from a predominant factor which can be the bait identity. Here, -K7 and -K8 baits seem closely related as the samples tend to stay together on the PCA whatever the number of dimensions used (Figure 29.B.C).

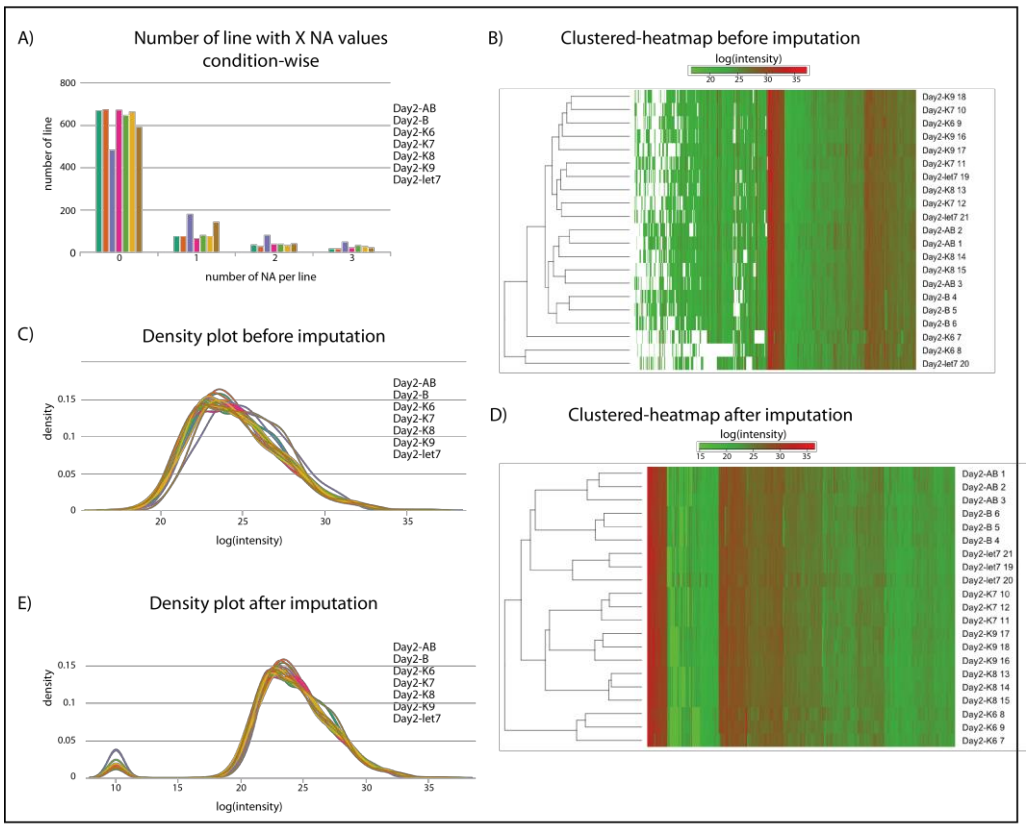


Figure 28. Overview of the matrix dataset before and after imputation procedure of the Day2 RNA chromatography. Prior to any data imputation, (A) displays the non-available values among the matrix, (B) represent the Euclidian clustered heatmap, (C) shows the density plot of the global protein intensity among conditions. After the imputation of the partially observed values (POV) and the missing on entire condition values (MEC), (D) displays an Euclidian clustered heatmap, (E) shows the density plot of the global protein intensity among conditions.

Overall descriptive analysis of this experiment suggest that chromatography experiment is good enough to perform the differential analysis using the limma statistical test corrected with Benjamini and Hochberg's method. Once the differential statistical test has been performed, volcano plots are generated with a fold change threshold of 2 and a p-value threshold of 5 %. Significantly enriched datasets are produced for all the SL-miRNA bait condition when compared to the control ssDNA-bead (-AB). This time let7a-1 bait allows the enrichment of LN28B and ROA1 but not FUBP2 (Figure 30.A.B). The absence FUBP2 enrichment is not problematic as this protein is not as abundant as ROA1 in B lymphocyte lines (gene expression across species: https://www.ebi.ac.uk/gxa/home). Since FUBP2 and ROA1 compete with each other for the let-7a-1 precursor, it could make sense to preferentially identify the most abundant one. Contrary to the previous chromatography, the SL-miRNA -K6, -K7 and -K8 significantly enriched either ILF2 or ILF3, but not both (Figure 30.A.B). Recent study showed that RNA binding capacity of either ILF2 or ILF3 is improve within the ILF2-ILF3 complex (Schmidt *et al.*, 2017). Finally, each SL-miRNA enriched datasets from the Day2 chromatography are compared using Venn diagram. Global lists of enriched protein are also generated (Figure 30.A.B.C.D).

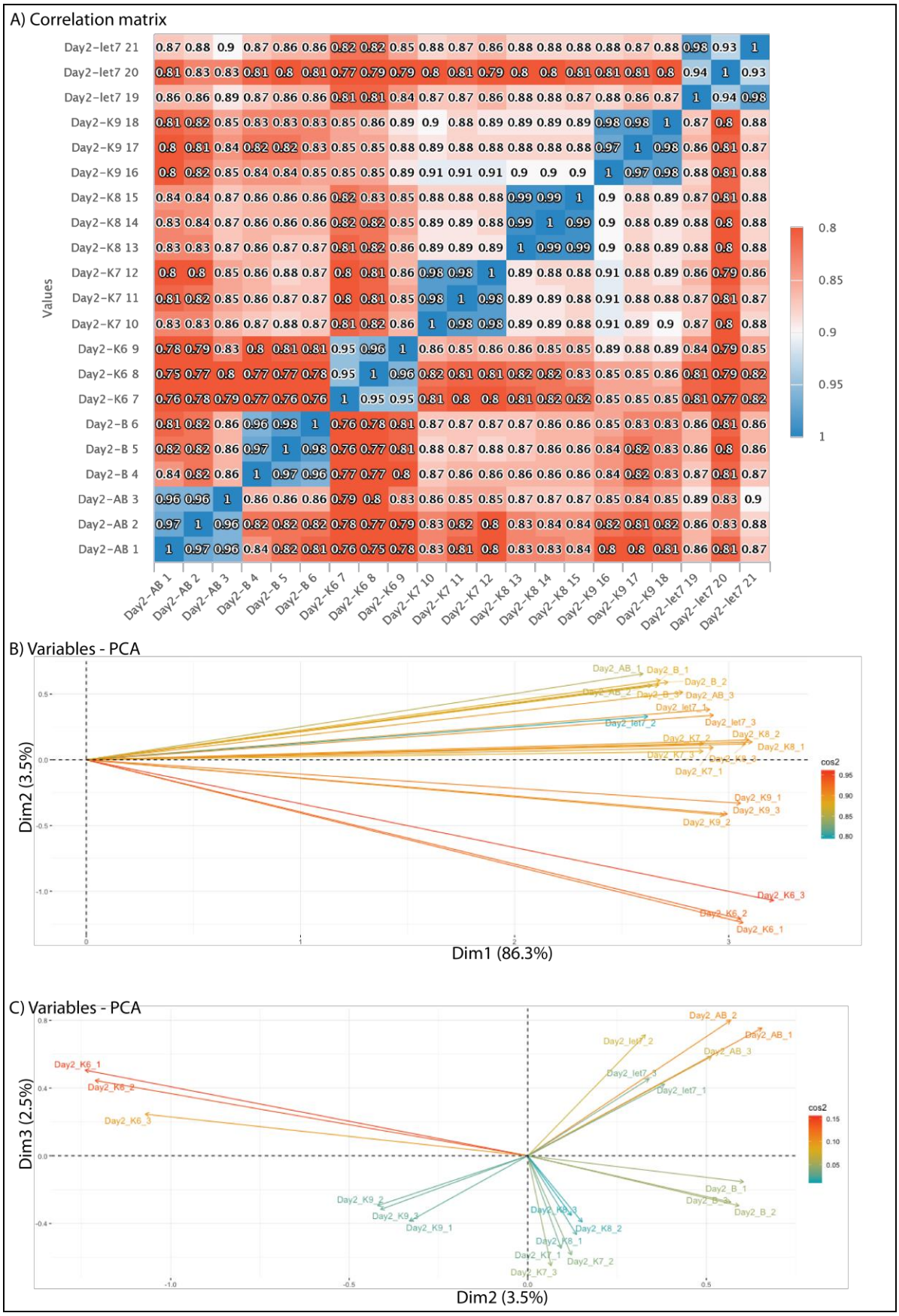


Figure 29. Descriptive analysis of the Day2 chromatography dataset. (A) Correlation matrix of the chromatography baits. (B) Principal component analysis (PCA) with the first and second dimensions. (C) Principal component analysis with the second and third dimensions.

As before, we can take a closer look at the cell specific and significantly enriched proteins by a given SL-miRNA.

-**SL-miRNA-K6** managed to specifically enrich, among others, a potential splicing factor involve in innate immune response CATIN (Zanini *et al.*, 2017), a helicase DHX36 (Treiber *et al.*, 2017), three heterogenous nuclear ribonucleoprotein HNRH1; HNRH2; HNRPF (Das *et al.*, 2019), another splicing factor PPIE (Jurica *et al.*, 2002).

-**SL-miRNA-K7** managed to specifically enrich, among others, a heterogenous nuclear ribonucleoprotein HNRL1 (Das *et al.*, 2019).

-**SL-miRNA-K8** managed to specifically enrich, among others, a splicing factor SUGP2, and a putative component of the RNA exosome RRP44 (Treiber *et al.*, 2017).

-**SL-miRNA-K9** managed to specifically enrich, among others, far upstream binding element FUBP1 FUBP2 FUBP3, two heterogenous nuclear ribonucleoproteins HNRPL HNRD, and proteins involve in various RNA processing ROA2; ROAA; RBMS1; RBMS2; PTBP2; SF01; TIA1 (Treiber *et al.*, 2017).

-Commonly identified by the different SL-miRNA: Among them we can cite DHX15; YBOX1 which have already been described to be involve in miRNA biogenesis (Treiber *et al.*, 2017) and the VIR protein which recruit methyl-transferases on RNA (Yue *et al.*, 2018) (Figure 30.D).

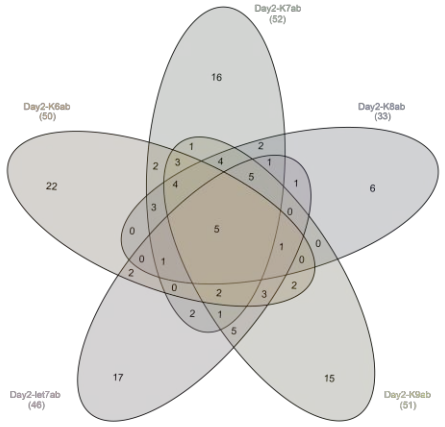
A)

SL-K6	SL-K7	SL-K8	SL-K9	SL-let7	Entry name
NA	Enriched	NA	Enriched	NA	AHNK_HUMAN
Enriched	NA	NA	Enriched	NA	ANLN_HUMAN
NA	Enriched	NA	NA	Enriched	ANXA1_HUMAN
NA	Enriched	Enriched	Enriched	NA	ASAH1_HUMAN
Enriched	NA	NA	NA	NA	ATF7_HUMAN
NA	Enriched	NA	NA	NA	BLM_HUMAN
NA	NA	Enriched	NA	NA	BLMH_HUMAN
NA	Enriched	NA	NA	NA	BRI3B_HUMAN
NA	NA	NA	NA	Enriched	CALL5_HUMAN
NA	NA	Enriched	NA	NA	CARME_HUMAN
NA	NA	Enriched	NA	NA	CASPE_HUMAN
Enriched	NA	NA	NA	NA	CATIN_HUMAN
NA	NA	Enriched	NA	Enriched	CATL2_HUMAN
NA	Enriched	NA	NA	NA	CCNK_HUMAN
Enriched	NA	NA	NA	NA	CE290_HUMAN
NA	Enriched	Enriched	Enriched	NA	CHLE_HUMAN
NA	Enriched	Enriched	Enriched	NA	CMTR1_HUMAN
NA	NA	NA	Enriched	NA	DAZP1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	DDX1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	DHX15_HUMAN
NA	Enriched	Enriched	NA	NA	DHX16_HUMAN
Enriched	NA	NA	NA	NA	DHX36_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	DHX9_HUMAN
Enriched	NA	Enriched	Enriched	Enriched	DIAP2_HUMAN
NA	Enriched	NA	NA	NA	DJC10_HUMAN
NA	NA	NA	Enriched	Enriched	DSRAD_HUMAN
Enriched	NA	NA	Enriched	Enriched	E2AK2_HUMAN
Enriched	Enriched	NA	Enriched	NA	ECHD1_HUMAN
NA	Enriched	NA	NA	NA	EF2_HUMAN
Enriched	Enriched	NA	Enriched	NA	EFHD1_HUMAN
Enriched	NA	NA	NA	NA	EI2BD_HUMAN
NA	NA	Enriched	NA	NA	EOGT_HUMAN
NA	Enriched	NA	NA	NA	F16A1_HUMAN
NA	NA	NA	NA	Enriched	FABP5_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	FBH1_HUMAN
NA	NA	NA	Enriched	NA	FUBP1_HUMAN
NA	NA	NA	Enriched	NA	FUBP2_HUMAN
NA	NA	NA	Enriched	NA	FUBP3_HUMAN
NA	Enriched	NA	NA	NA	GGCT_HUMAN
NA	NA	NA	NA	Enriched	GLNA_HUMAN
NA	NA	NA	NA	Enriched	GSDMA_HUMAN
Enriched	NA	NA	NA	NA	HBS1L_HUMAN
NA	NA	NA	Enriched	NA	HNRDL_HUMAN
Enriched	NA	NA	NA	NA	HNRH1_HUMAN
Enriched	NA	NA	NA	NA	HNRH2_HUMAN
Enriched	NA	NA	NA	Enriched	HNRH3_HUMAN
NA	Enriched	NA	NA	NA	HNRL1_HUMAN
NA	NA	NA	Enriched	Enriched	HNRLL_HUMAN
NA	NA	NA	Enriched	NA	HNRPD_HUMAN
Enriched	NA	NA	NA	NA	HNRPF_HUMAN
NA	NA	NA	Enriched	NA	HNRPL_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	HS71L_HUMAN
NA	Enriched	Enriched	Enriched	NA	HSP76_HUMAN
NA	NA	NA	NA	Enriched	HSPB1_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	HUTH_HUMAN
NA	NA	NA	NA	Enriched	IGHG3_HUMAN
Enriched	NA	NA	Enriched	Enriched	ILF2_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	ILF3_HUMAN
NA	Enriched	NA	NA	NA	KTN1_HUMAN
Enriched	NA	NA	Enriched	Enriched	LA_HUMAN
NA	NA	NA	NA	Enriched	LAP2B_HUMAN
NA	NA	NA	NA	Enriched	LEG7_HUMAN
NA	NA	NA	NA	Enriched	LN28B_HUMAN

B)

SL-K6	SL-K7	SL-K8	SL-K9	SL-let7	Entry name
Enriched	NA	NA	NA	NA	LYSC_HUMAN
NA	NA	NA	NA	Enriched	MA2A1_HUMAN
NA	Enriched	NA	NA	NA	MED1_HUMAN
NA	NA	NA	NA	Enriched	MOES_HUMAN
Enriched	NA	NA	NA	Enriched	MSI2H_HUMAN
Enriched	Enriched	NA	Enriched	Enriched	MYH15_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	NAA25_HUMAN
Enriched	NA	NA	NA	NA	NMD3A_HUMAN
Enriched	NA	NA	NA	NA	NP1L1_HUMAN
Enriched	NA	NA	NA	NA	NP1L4_HUMAN
NA	NA	NA	Enriched	Enriched	NUCL_HUMAN
Enriched	Enriched	NA	NA	NA	NUP54_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	NUP98_HUMAN
Enriched	NA	NA	Enriched	NA	PARVB_HUMAN
Enriched	NA	NA	NA	NA	PLOD1_HUMAN
Enriched	Enriched	Enriched	NA	NA	POP1_HUMAN
NA	Enriched	NA	NA	NA	PPE1_HUMAN
Enriched	NA	NA	NA	NA	PPIE_HUMAN
NA	NA	NA	NA	Enriched	PRDX2_HUMAN
NA	NA	NA	NA	Enriched	PREX1_HUMAN
NA	NA	NA	Enriched	NA	PTBP1_HUMAN
NA	NA	NA	Enriched	NA	PUM2_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	RAVR1_HUMAN
NA	NA	NA	Enriched	NA	RBMS1_HUMAN
NA	NA	NA	Enriched	NA	RBMS2_HUMAN
Enriched	NA	NA	NA	NA	RCC1_HUMAN
NA	NA	NA	NA	Enriched	RCOR1_HUMAN
NA	Enriched	NA	NA	NA	RECQ1_HUMAN
Enriched	NA	NA	NA	NA	RFX5_HUMAN
Enriched	Enriched	NA	NA	NA	RHG05_HUMAN
NA	NA	NA	Enriched	Enriched	ROA1_HUMAN
NA	NA	NA	Enriched	NA	ROA2_HUMAN
NA	NA	NA	Enriched	Enriched	ROA3_HUMAN
NA	NA	NA	Enriched	NA	ROAA_HUMAN
NA	Enriched	NA	NA	NA	RPA1_HUMAN
NA	Enriched	NA	NA	NA	RPA2_HUMAN
NA	Enriched	NA	NA	NA	RPB2_HUMAN
Enriched	Enriched	NA	Enriched	Enriched	RPB3_HUMAN
Enriched	NA	NA	NA	NA	RPR1B_HUMAN
NA	NA	Enriched	NA	NA	RRP44_HUMAN
Enriched	NA	NA	NA	NA	RTCB_HUMAN
NA	NA	NA	NA	Enriched	S10A9_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	SC23B_HUMAN
NA	Enriched	Enriched	NA	NA	SCRB2_HUMAN
Enriched	NA	NA	NA	NA	SEH1_HUMAN
NA	NA	NA	Enriched	NA	SF01_HUMAN
NA	Enriched	NA	Enriched	Enriched	SPB3_HUMAN
NA	NA	NA	NA	Enriched	SR140_HUMAN
Enriched	Enriched	NA	Enriched	NA	SRRM1_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	STRBP_HUMAN
NA	NA	Enriched	NA	NA	SUGP2_HUMAN
Enriched	NA	NA	NA	NA	SUZ12_HUMAN
NA	Enriched	NA	NA	NA	SYFB_HUMAN
Enriched	Enriched	Enriched	NA	NA	SYQ_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	TCPG_HUMAN
NA	Enriched	NA	NA	Enriched	TGM1_HUMAN
NA	NA	NA	Enriched	NA	TIA1_HUMAN
Enriched	NA	NA	NA	NA	TRIM4_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	VIR_HUMAN
Enriched	Enriched	Enriched	NA	NA	WWP2_HUMAN
NA	Enriched	Enriched	NA	Enriched	YBOX1_HUMAN
NA	NA	NA	NA	Enriched	ZC3H4_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	ZN292_HUMAN

C)



D)

Common enriched
DDX1
DHX15 DBP1 DDX15
FBH1 FBX18 FBXO18
NUP98 ADAR2
SEC23B

Figure 30. Differential analysis of the Day2 dataset. (A) and (B) display the bait enriched protein compare to the control -AB selected from volcano plots with a fold change and p-values threshold of respectively 2 and 5 %. C)Venn diagram of the bait enriched proteins. D) Common enriched proteins.

III.A.4.b.1.5 Day3: SL-miRNA-K11, -155, -let7

On Day3, BC-3 and DG75 nuclear extracts were used for the chromatographies. Since I want to compare the two chromatographies, I used homogeneously prepared SL-miRNA baits for both pulldowns. Concerning the Day3 (BC-3) pulldown, 1489 different proteins have been identified containing ~8 % of NA values (Table 6).

Table 6. Mass spectrometry protein identification overview of the Day3 chromatography performed with BC3 nuclear extract. Five baits condition have been used (-B, -AB, -let7, -K11, -155).

Definition	Value
Number of samples	15
Number of conditions	5
Number of lines	1489
Number of missing values	1705
% of missing values	7.63
Number of empty lines	13

This percentage is lower than in the previous experiments indicating that each bait were equivalently and efficiently able to identify proteins during respectively the chromatography or the mass spectrometry analysis. Surprisingly, this chromatography allows the identification of a more heterogeneous protein set composed of more than 1400 proteins, as opposed to the previous chromatography experiments. This information suggests that the protein identification worked better regarding the important protein heterogeneity and the low % of NA values. This increase in numbers of identified proteins can also come from a richer nuclear extract than previously used. Indeed, all the triplicate chromatography conditions were able to identify close to 1200 proteins and the number of NA values per condition is homogeneously distributed (Figure 31.A). Prior to any imputation step, clustered heatmap reveals a good similarity between triplicate that almost perfectly clustered together (Figure 31.B). Interestingly, -AB and -B condition group together under the dendrogram node which makes sense because both baits do not possess any SL-miRNA. The mean density of the log(intensity) is around 25 such as the previous experiments, indicating that a similar protein signal than in the previous pulldowns was retrieved during the mass spectrometry identification (Figure 31.C). As previously, I imputed the POV using the SLSA algorithm and a fixed value of 10 for the MEC protein (Figure 31.E). As usual, the imputation step smoothens out triplicate differences and allows a perfect condition clustering, with small dendrogram branches suggesting again a close proximity of each samples within a triplicate (Figure 31.E).

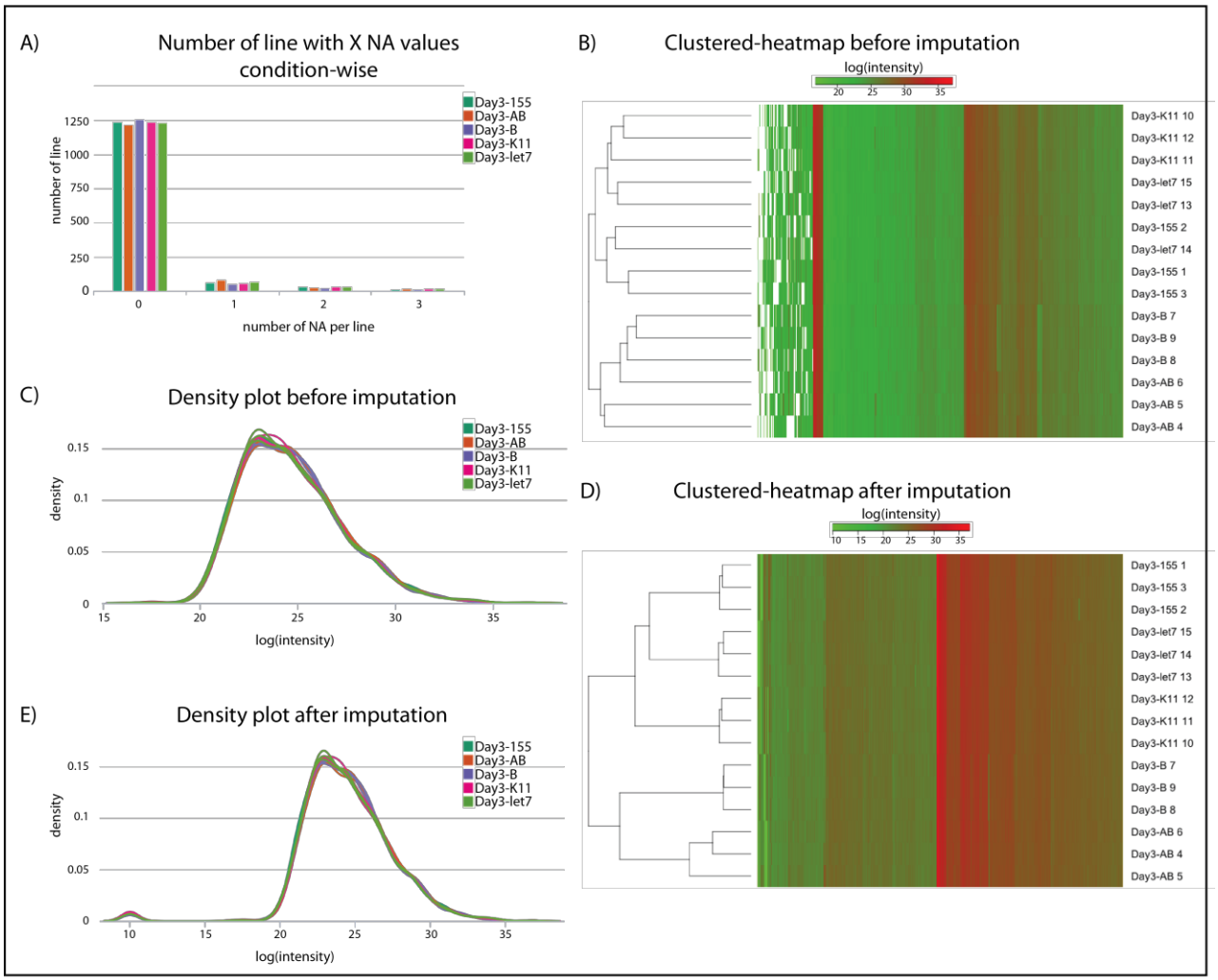


Figure 31. Overview of the Day3 matrix dataset before and after imputation procedure. Prior to any data imputation, (A) displays the non-available values among the matrix, (B) represent the Euclidian clustered heatmap, (C) shows the density plot of the global protein intensity among conditions. After the imputation of the partially observed values (POV) and the missing on entire condition values (MEC), (D) displays an Euclidian clustered heatmap, (E) shows the density plot of the global protein intensity among conditions.

As excepted, matrix correlation as well as PCA reveal a high degree of similarity among triplicate samples (Figure 32.A). In addition, PCA clearly indicates that bait identity is potentially the main factor explaining samples differences (Figure 32.B.C).

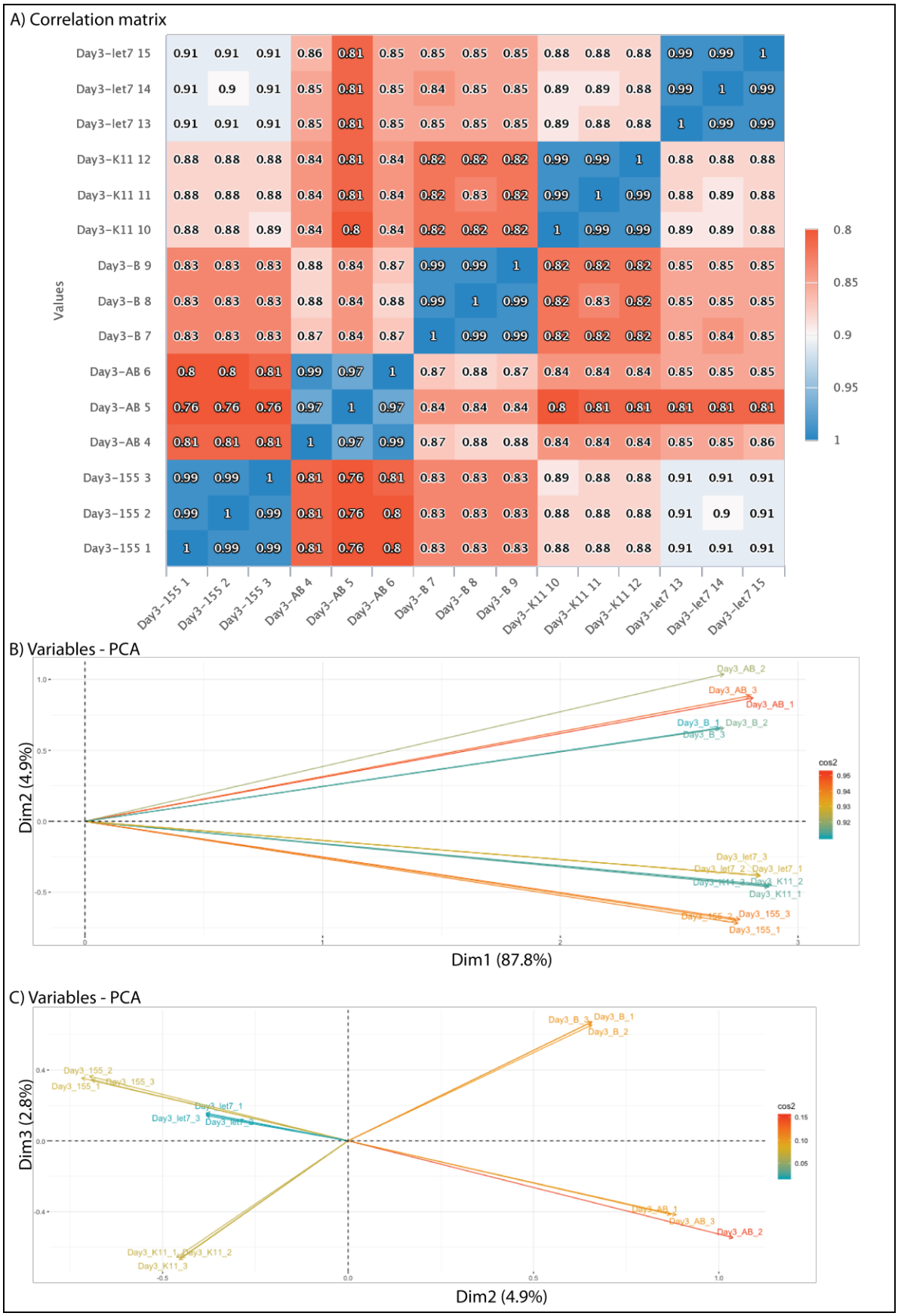


Figure 32. Descriptive analysis of the Day3 chromatography dataset. (A) Correlation matrix of the chromatography baits. (B) Principal component analysis (PCA) with the first and second dimensions. (C) Principal component analysis with the second and third dimensions.

Overall dataset description clearly indicates that the chromatography and the protein identification worked fine and allows me to do the differential analysis as described before. Enriched datasets are produced for all the SL-miRNA bait condition after being compare to the control ssDNA-bead (-AB). This time SL-miRNA-let7 bait allows the enrichment of LN28B and ROA1 but not FUBP2 (Figure 33.A.B). However, all the SL-miRNA managed to significantly enrich both ILF2 and ILF3 proteins. Taken together, all these results nicely validate the chromatography. Finally, each SL-miRNA enriched datasets from the same chromatography are compared using Venn diagram. Global list of enriched protein is also generated (Figure 33.A.B.C.D).

As before, we can take a closer look on the cell specific and significantly enriched proteins by a given SL-miRNA.

-SL-miRNA-K11 managed to specifically enrich, among others, a splicing factor MBNL1, a regulator of mitochondrial post-transcriptional process GRSF1, and a protein involve in snRNA and snoRNA maturation PHAX.

-SL-miRNA-155 managed to specifically enrich, among others, a nucleo-mitochondrial RNA metabolism protein LPPRC with its cofactor SLIRP, a 3' UTR binder CIRBP, a heterogenous ribonucleoprotein HNRPD, a protein implicated in numerous RNA metabolic processes PNPT1 and splicing factors PTBP3; QKI; YBOX1; YBOX3.

-Commonly identified by the different SL-miRNA: Among them we can cite a methyltransferase MET16, and DHX15; ZNF346 which have already been described to be involved in miRNA biogenesis (Chen *et al.*, 2004; Alarcón *et al.*, 2015; Treiber *et al.*, 2017) (Figure 33.D).

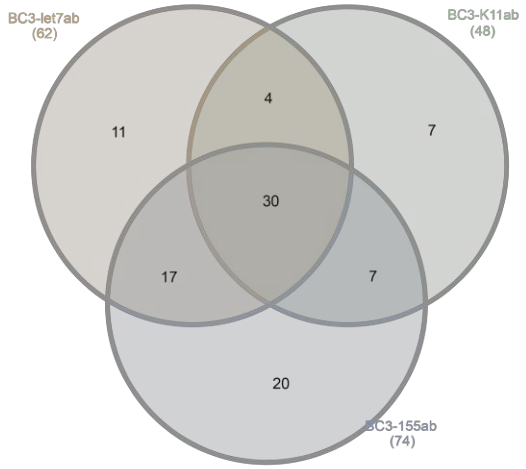
A)

SL-K11	SL-155	SL-let7	Entry name
Enriched	NA	NA	1B08_HUMAN
Enriched	NA	Enriched	3MG_HUMAN
NA	Enriched	NA	ABC3D_HUMAN
Enriched	Enriched	NA	ABCE1_HUMAN
NA	Enriched	NA	ADT2_HUMAN
NA	Enriched	Enriched	ADT3_HUMAN
NA	Enriched	NA	ANX11_HUMAN
NA	Enriched	Enriched	AT5F1_HUMAN
NA	Enriched	Enriched	ATP5H_HUMAN
NA	Enriched	Enriched	ATP5I_HUMAN
NA	Enriched	Enriched	ATP5L_HUMAN
Enriched	Enriched	Enriched	ATP6_HUMAN
NA	Enriched	Enriched	ATPA_HUMAN
NA	Enriched	NA	ATPB_HUMAN
NA	Enriched	Enriched	ATPG_HUMAN
NA	Enriched	Enriched	ATPK_HUMAN
NA	Enriched	Enriched	ATPO_HUMAN
NA	NA	Enriched	CE290_HUMAN
NA	NA	Enriched	CELF1_HUMAN
NA	Enriched	NA	CIRBP_HUMAN
NA	Enriched	Enriched	CMC2_HUMAN
NA	Enriched	NA	CMTR1_HUMAN
Enriched	Enriched	Enriched	CNOT6_HUMAN
NA	Enriched	NA	COX2_HUMAN
NA	Enriched	NA	DAZP1_HUMAN
Enriched	Enriched	Enriched	DHX15_HUMAN
Enriched	Enriched	Enriched	DHX9_HUMAN
Enriched	Enriched	Enriched	DSRAD_HUMAN
Enriched	Enriched	Enriched	E2AK2_HUMAN
Enriched	Enriched	Enriched	EF1A1_HUMAN
NA	NA	Enriched	ELAV1_HUMAN
Enriched	Enriched	NA	ERI1_HUMAN
Enriched	NA	NA	EXOC7_HUMAN
NA	Enriched	NA	GARS_HUMAN
Enriched	NA	Enriched	GRSF1_HUMAN
Enriched	Enriched	Enriched	HBS1L_HUMAN
Enriched	NA	Enriched	HNRH3_HUMAN
Enriched	Enriched	Enriched	HNRL1_HUMAN
NA	Enriched	NA	HNRPD_HUMAN
Enriched	Enriched	Enriched	IF2B_HUMAN
NA	Enriched	Enriched	IF2G_HUMAN
Enriched	Enriched	Enriched	ILF2_HUMAN
Enriched	Enriched	Enriched	ILF3_HUMAN
Enriched	Enriched	Enriched	LA_HUMAN
Enriched	Enriched	Enriched	LFA3_HUMAN
Enriched	NA	Enriched	LN28B_HUMAN
NA	Enriched	NA	LPPRC_HUMAN
NA	Enriched	NA	LSG1_HUMAN

B)

SL-K11	SL-155	SL-let7	Entry name
Enriched	NA	NA	MBNL1_HUMAN
Enriched	Enriched	Enriched	MET16_HUMAN
Enriched	Enriched	NA	MOT1_HUMAN
NA	Enriched	Enriched	MPCP_HUMAN
NA	NA	Enriched	MSI2H_HUMAN
NA	NA	Enriched	NEPR1_HUMAN
NA	Enriched	NA	NUCL_HUMAN
Enriched	NA	NA	PHAX_HUMAN
NA	Enriched	NA	PHB_HUMAN
Enriched	Enriched	NA	PHS2_HUMAN
NA	Enriched	NA	PNPT1_HUMAN
Enriched	Enriched	Enriched	PRKRA_HUMAN
Enriched	Enriched	Enriched	PTBP1_HUMAN
NA	Enriched	Enriched	PTBP3_HUMAN
Enriched	Enriched	NA	PURB_HUMAN
NA	Enriched	Enriched	QKI_HUMAN
Enriched	NA	NA	RAB2A_HUMAN
Enriched	Enriched	NA	RAB6B_HUMAN
NA	Enriched	Enriched	RAB7A_HUMAN
Enriched	Enriched	Enriched	RAB8A_HUMAN
Enriched	Enriched	Enriched	RAN_HUMAN
Enriched	Enriched	Enriched	RAP1B_HUMAN
NA	Enriched	NA	RB11A_HUMAN
Enriched	NA	NA	RBMS1_HUMAN
Enriched	Enriched	NA	RECQ1_HUMAN
Enriched	Enriched	Enriched	RHOA_HUMAN
NA	Enriched	Enriched	ROA0_HUMAN
NA	NA	Enriched	ROA1_HUMAN
NA	NA	Enriched	ROA2_HUMAN
NA	NA	Enriched	ROA3_HUMAN
Enriched	Enriched	Enriched	RPA43_HUMAN
NA	NA	Enriched	RPB3_HUMAN
NA	Enriched	Enriched	SCAM2_HUMAN
NA	Enriched	NA	SLIRP_HUMAN
NA	Enriched	NA	SQOR_HUMAN
NA	NA	Enriched	SRSF1_HUMAN
NA	NA	Enriched	SRSF9_HUMAN
Enriched	Enriched	Enriched	STRBP_HUMAN
Enriched	Enriched	Enriched	TET2_HUMAN
Enriched	Enriched	Enriched	TIM8A_HUMAN
Enriched	Enriched	Enriched	TRBP2_HUMAN
Enriched	Enriched	Enriched	TRIM4_HUMAN
Enriched	NA	NA	TRUA_HUMAN
NA	Enriched	NA	YBOX1_HUMAN
NA	Enriched	NA	YBOX3_HUMAN
Enriched	Enriched	Enriched	ZCCHV_HUMAN
Enriched	Enriched	Enriched	ZFR_HUMAN
Enriched	Enriched	Enriched	ZN346_HUMAN

C)



D)

Common enriched		
ATP6_HUMAN	ILF2_HUMAN	RHOA_HUMAN
CNOT6_HUMAN	ILF3_HUMAN	RPA43_HUMAN
DHX15_HUMAN	LA_HUMAN	STRBP_HUMAN
DHX9_HUMAN	LFA3_HUMAN	TET2_HUMAN
DSRAD_HUMAN	MET16_HUMAN	TIM8A_HUMAN
E2AK2_HUMAN	PRKRA_HUMAN	TRBP2_HUMAN
EF1A1_HUMAN	PTBP1_HUMAN	TRIM4_HUMAN
HBS1L_HUMAN	RAB8A_HUMAN	ZCCHV_HUMAN
HNRL1_HUMAN	RAN_HUMAN	ZFR_HUMAN
IF2B_HUMAN	RAP1B_HUMAN	ZN346_HUMAN

Figure 33. Differential analysis of the Day3 dataset. (A) and (B) display the bait enriched protein compare to the control -AB selected from volcano plot with a fold change and p-values threshold of respectively 2 and 5 %. C)Venn diagram of the bait enriched proteins. D) Common enriched proteins.

III.A.4.b.1.6 DG75: SL-miRNA-K11, -155, -let7

In this assay, 1433 different proteins have been identified, generating a dataset matrix containing ~8% of NA values (Table 7). Such as the previous experiment within BC-3 cell extract, these results strongly emphasized that protein identification occurs well and more importantly that the chromatography appears good. Indeed, descriptive analysis confirm this idea.

Table 7. Mass spectrometry protein identification overview of the Day4 chromatography performed with DG75 nuclear extract. Five baits condition have been used (-B, -AB, -7, -K11, -155)

Definition	Value
Number of samples	15
Number of conditions	5
Number of lines	1433
Number of missing values	1687
% of missing values	7.85
Number of empty lines	11

The number of NA values is homogeneously distributed among bait conditions, and clustered heatmap displays a good similarity between triplicate (Figure 34.A). Again, the mean density of the log(intensity) is around 25 such as the previous experiments, indicating that an overall protein signal is comparable with the previous pulldowns (Figure 34.C). As previously, I imputed the POV using the SLSA algorithm and imputed a fixed value of 10 for the MEC (Figure 34.E). As usual, the imputation step smoothens out triplicate differences and allows a perfect condition clustering, with small dendrogram branch suggesting again a close proximity of each samples within a triplicate (Figure 34.B.D). In addition, triplicate within a bait condition share a dendrogram node which is different for each bait, emphasizing again a good reproducibility of chromatography as well as differences between bait conditions. As excepted, matrix correlation as well as PCA reveal a high degree of similarity among triplicate samples (Figure 35.A). In addition, PCA clearly indicates that bait condition is potentially the main factor explaining samples differences (Figure 35.B.C).

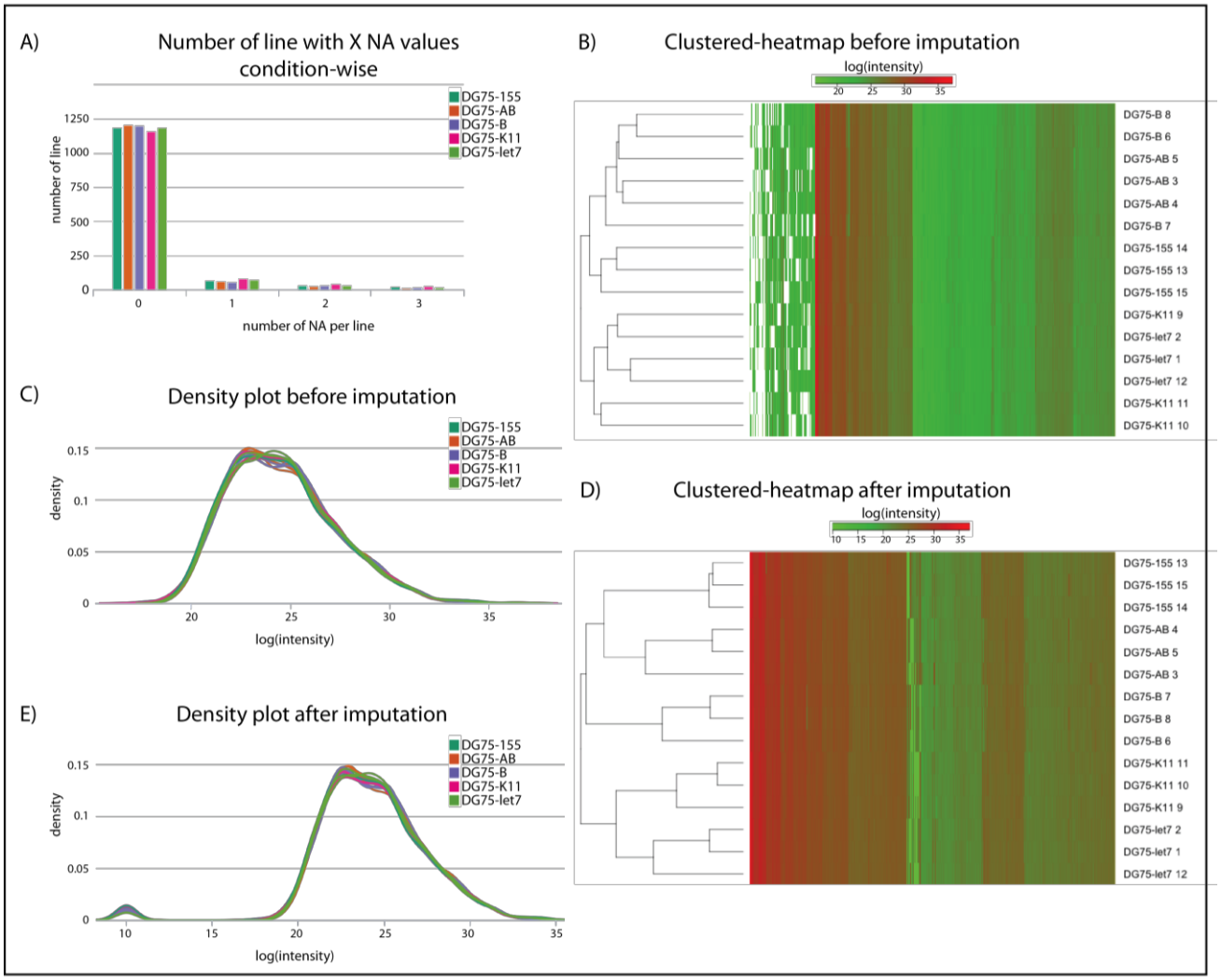


Figure 34. Overview of the DG75 matrix dataset before and after imputation procedure. Prior to any data imputation, (A) displays the non-available values among the matrix, (B) represent the Euclidian clustered heatmap, (C) shows the density plot of the global protein intensity among conditions. After the imputation of the partially observed values (POV) and the missing on entire condition values (MEC), (D) displays an Euclidian clustered heatmap, (E) shows the density plot of the global protein intensity among conditions.

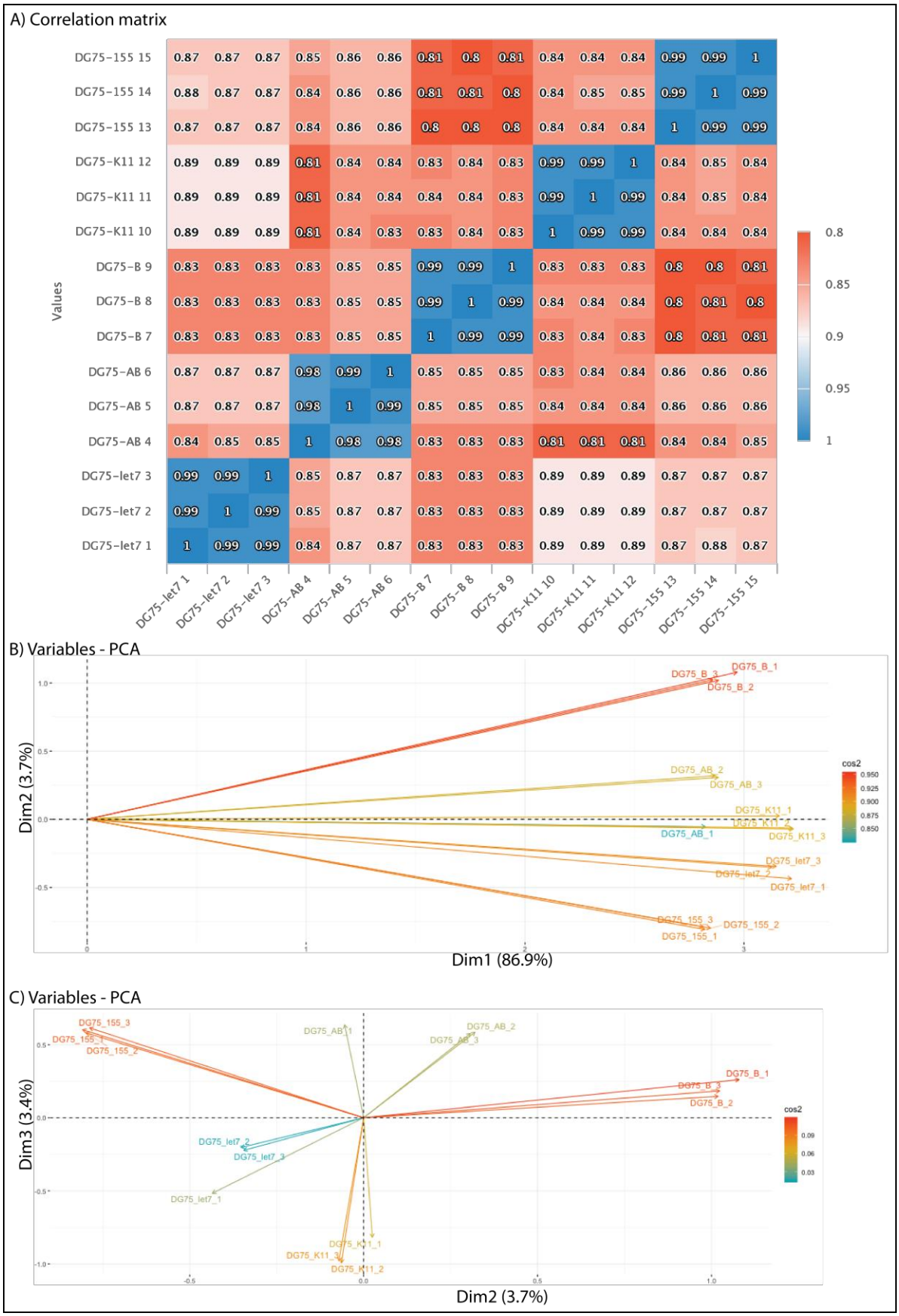


Figure 35. Descriptive analysis of the DG75 chromatography dataset. (A) Correlation matrix of the chromatography baits. (B) Principal component analysis (PCA) with the first and second dimensions. (C) Principal component analysis with the second and third dimensions.

Overall dataset description clearly indicate that the chromatography and the protein identification worked perfectly and allow me to do the differential analysis using. Once the differential statistical test performed, volcano plots are generated with a fold change threshold of 2 and a p-value threshold of 5 %. This time SL-miRNA-let7 bait allows the enrichment of ROA1 but not LN28B and FUBP2 (Figure 36.A.B). The absence FUBP2 and LN28B enrichment could be explain by their relative low abundance in lymphocyte compare to ROA1 (gene expression across species: https://www.ebi.ac.uk/gxa/home). As always, ILF2 and ILF3 are always enriched by the majority of the SL-miRNA used. Finally, each SL-miRNA enriched datasets from the same chromatography are compared using Venn diagram. A global list of enriched proteins is also generated (Figure 36.A.B.C.D).

Here are the cell specific and significantly enriched proteins by a given SL-miRNA:

-**SL-miRNA-K11** managed to specifically enrich, among others, a methyl-transferase MET16 (Alarcón *et al.*, 2015; Tan *et al.*, 2017), a heterogenous ribonucleoprotein HNRH3 (Das *et al.*, 2019).

-**SL-miRNA-155** managed to specifically enrich, among others, splicing factors PTBP1; PTBP3; QKI; DAZAP1 (Treiber *et al.*, 2017), a nucleo-mitochondrial RNA metabolism protein LPPRC (Siira *et al.*, 2017).

-Commonly identified by the different SL-miRNA: splicing factor MBNL1 (Rau *et al.*, 2011), antiviral proteins ZCCHV (Hayakawa *et al.*, 2011); ABCE1; IF2B1 and DHX15 (Treiber *et al.*, 2017); ZNF346 which have already been described to be involved in miRNA biogenesis (Chen *et al.*, 2004; Treiber *et al.*, 2017) (Figure 36.D).

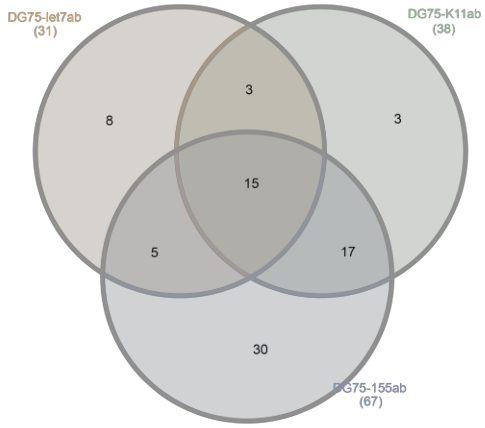
A)

SL-K11	SL-155	SL-let7	Entry name
Enriched	Enriched	NA	ABCE1_HUMAN
Enriched	NA	NA	ACTB_HUMAN
NA	Enriched	Enriched	ACTBL_HUMAN
Enriched	NA	NA	ADA10_HUMAN
NA	Enriched	NA	ARP3_HUMAN
NA	Enriched	NA	ATPA_HUMAN
NA	Enriched	NA	ATPG_HUMAN
NA	Enriched	NA	ATPO_HUMAN
Enriched	Enriched	NA	BMP2K_HUMAN
NA	Enriched	NA	C1TC_HUMAN
NA	Enriched	Enriched	CD2B2_HUMAN
NA	Enriched	Enriched	CELF2_HUMAN
NA	Enriched	NA	CMTR1_HUMAN
NA	Enriched	NA	COEA1_HUMAN
Enriched	Enriched	NA	CSCL1_HUMAN
NA	Enriched	NA	CSTF1_HUMAN
NA	Enriched	NA	CTNA1_HUMAN
NA	Enriched	NA	DAZP1_HUMAN
NA	NA	Enriched	DCD_HUMAN
Enriched	Enriched	NA	DHX15_HUMAN
Enriched	Enriched	Enriched	DHX9_HUMAN
Enriched	Enriched	Enriched	DSRAD_HUMAN
NA	Enriched	NA	DYN2_HUMAN
Enriched	Enriched	NA	DYST_HUMAN
Enriched	Enriched	Enriched	E2AK2_HUMAN
Enriched	Enriched	NA	EF1A2_HUMAN
Enriched	Enriched	NA	EF1A3_HUMAN
NA	Enriched	NA	EHD1_HUMAN
Enriched	Enriched	NA	EPHA4_HUMAN
Enriched	Enriched	Enriched	ERI1_HUMAN
NA	Enriched	NA	FAT3_HUMAN
NA	Enriched	NA	FUBP3_HUMAN
NA	Enriched	NA	FUMH_HUMAN
Enriched	NA	NA	HBB_HUMAN
Enriched	Enriched	NA	HBS1L_HUMAN
Enriched	Enriched	Enriched	HCLS1_HUMAN
Enriched	NA	Enriched	HNRH3_HUMAN
Enriched	Enriched	NA	HNRL1_HUMAN
NA	Enriched	NA	IF2A_HUMAN
NA	Enriched	NA	IF2B_HUMAN
Enriched	Enriched	NA	IF2B1_HUMAN

B)

SL-K11	SL-155	SL-let7	Entry name
Enriched	Enriched	Enriched	IF2G_HUMAN
Enriched	Enriched	Enriched	ILF2_HUMAN
Enriched	Enriched	Enriched	ILF3_HUMAN
Enriched	Enriched	Enriched	LA_HUMAN
NA	Enriched	NA	LPPRC_HUMAN
Enriched	Enriched	NA	LSG1_HUMAN
Enriched	Enriched	NA	MBNL1_HUMAN
NA	Enriched	NA	MCM7_HUMAN
Enriched	NA	Enriched	MET16_HUMAN
Enriched	Enriched	Enriched	MOT1_HUMAN
NA	Enriched	NA	NAC1_HUMAN
NA	Enriched	NA	NEPR1_HUMAN
NA	NA	Enriched	NMD3A_HUMAN
NA	NA	Enriched	PHOCN_HUMAN
Enriched	Enriched	Enriched	PRKRA_HUMAN
NA	NA	Enriched	PSB6_HUMAN
NA	Enriched	NA	PTBP1_HUMAN
NA	Enriched	Enriched	PTBP2_HUMAN
NA	Enriched	NA	PTBP3_HUMAN
NA	Enriched	NA	QKI_HUMAN
Enriched	Enriched	Enriched	RAB7A_HUMAN
Enriched	Enriched	NA	RAN_HUMAN
NA	Enriched	NA	RAP1A_HUMAN
NA	Enriched	NA	RM46_HUMAN
NA	Enriched	Enriched	ROA0_HUMAN
NA	NA	Enriched	ROA1_HUMAN
NA	NA	Enriched	ROA3_HUMAN
Enriched	Enriched	Enriched	RPB2_HUMAN
NA	NA	Enriched	S10A8_HUMAN
Enriched	Enriched	Enriched	STRBP_HUMAN
NA	Enriched	NA	SYFB_HUMAN
Enriched	Enriched	Enriched	TRIM4_HUMAN
NA	Enriched	NA	URP2_HUMAN
Enriched	NA	Enriched	VDAC1_HUMAN
NA	Enriched	NA	XP32_HUMAN
NA	Enriched	NA	YBOX3_HUMAN
Enriched	Enriched	NA	ZCCHV_HUMAN
Enriched	Enriched	NA	ZFR_HUMAN
NA	NA	Enriched	ZG16B_HUMAN
Enriched	Enriched	NA	ZN346_HUMAN

C)



D)

Common enriched
DHX9_HUMAN
DSRAD_HUMAN
E2AK2_HUMAN
ERI1_HUMAN
HCLS1_HUMAN
IF2G_HUMAN
ILF2_HUMAN
ILF3_HUMAN
LA_HUMAN
MOT1_HUMAN
PRKRA_HUMAN
RAB7A_HUMAN
RPB2_HUMAN
STRBP_HUMAN
TRIM4_HUMAN

Figure 36. Differential analysis of the DG75 dataset. (A) displays the bait enriched protein compare to the control -AB coming from volcano plot with a fold change and p-values threshold of respectively 2 and 5 %. B) Venn diagram of the bait enriched proteins. D) Common enriched proteins.

III.A.4.b.1.7 Day3 versus DG75

Since the levels of KSHV mature miRNA expression differ from one lymphocyte cell line to another. For example, we observed previously that kshv-miR-K10-11 is not expressed in the DG75 cell line (Contrant *et al.*, 2014). It is therefore interesting to use our mass spectrometry data to identify lymphocyte-specific enriched proteins. To do so, merged dataset of both chromatography cell line is generated by the Mascot software. Here, I will describe the differential analysis between the two cell lines for a given SL-miRNA bait (-let7, -K11, -155). PCA clearly shows that the cell extract is one of the most factor that explain the differences between the samples, which is expected (Figure 37.A). Another dimension allows us to confirm that the bait condition is the other important factor explaining samples differences (Figure 37.B). Then, I simply produced the corresponding significantly enriched dataset for each SL-miRNA per cell line and generate Venn diagram to detect common or cellular specific SL-miRNA binding proteins. As the PCA highlight major difference between the two cell extracts, if I compare both cell line for each SL-miRNA without checking the enrichment of each SL-miRNA to its own control, I obtain a big list of enriched proteins. For this reason, I only compared enriched protein dataset of each SL-miRNA (e.g. Day3-K11 enriched versus DG75-K11 enriched) (-let7: Figure 38; -K11: Figure 39; -155: Figure 40).

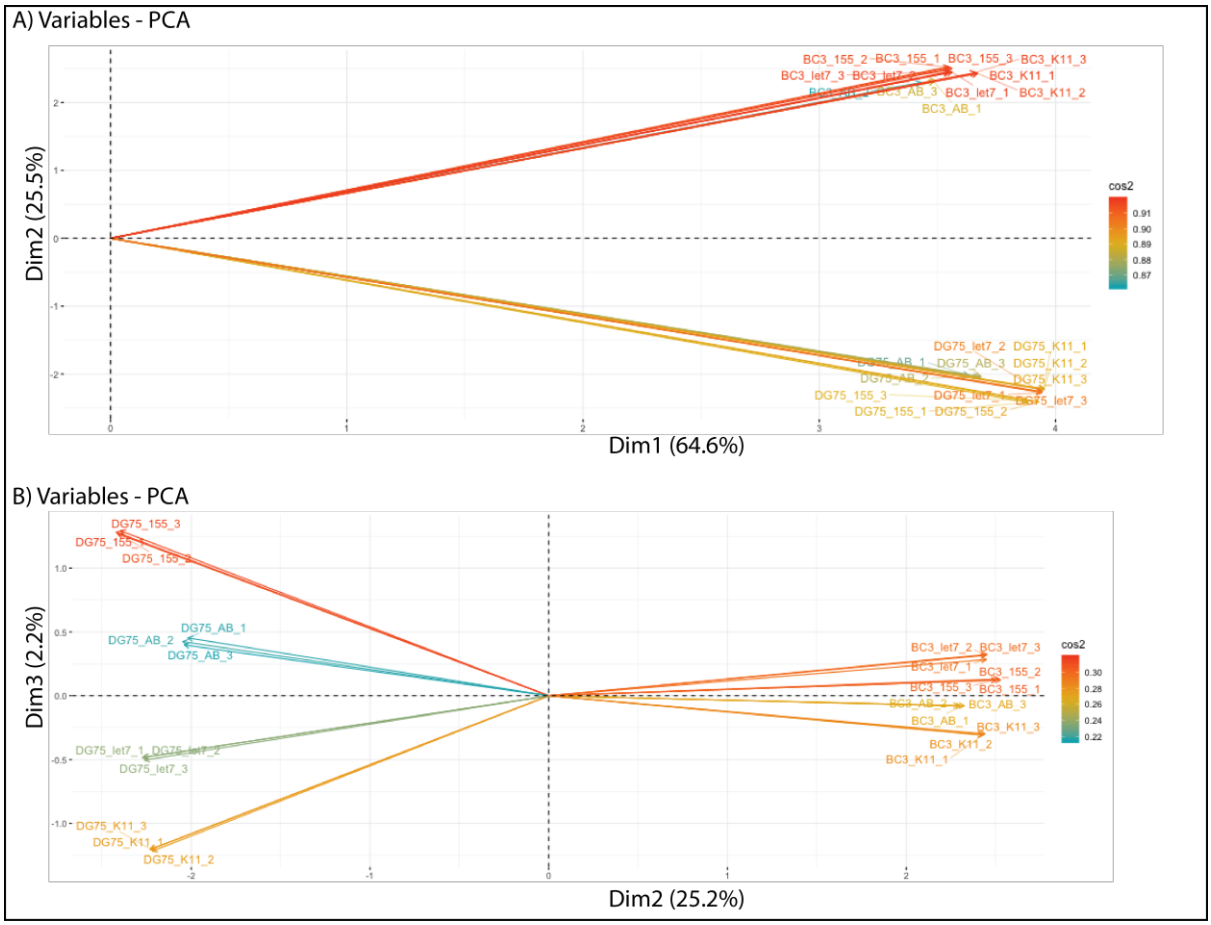


Figure 37. Descriptive analysis of the Day3 and DG75 RNA chromatography. (A) Principal component analysis (PCA) with the first and second dimensions. (B) Principal component analysis with the second and third dimensions.

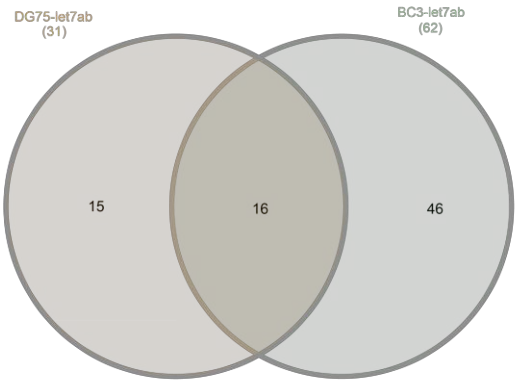
Here are the cell specific and significantly enriched proteins by a given SL-miRNA:

-**SL-let-7** (Figure 38): our analysis reveals that the SL-miRNA-let7 enriched the LN28B protein (Heo *et al.*, 2008) with the BC-3 extract and not with the DG75 extract. If I take a look on small RNA libraries generated in the laboratory, it reveals that the miRNA let-7a-1 is more expressed (5-fold) in the DG75 cell line compared to the BC-3 line. This result can highlight the potential of this approach to discover repressor or enhancer protein for a specific miRNA maturation.

-**SL-miRNA-K11** (Figure 39), the analysis reveals that only the IF2B1 (Weinlich *et al.*, 2009) protein seems DG75 specific. Within the BC-3 chromatography more proteins seem interesting such GRSF1 (Noh *et al.*, 2016), PHAX, LIN28B, PTPB1 and RBMS1 (Treiber *et al.*, 2017).

-**SL-miRNA-155** (Figure 40), we can cite the BC-3 specific proteins MET16, CIRBP, NUCL, YBOX1 and the DG75 specific proteins FUBP3, IF2B1 and MBNL1.

A)



B)

SL-miRNA-let7 BC3 specific candidates	
3MG_HUMAN	MPG AAG ANPG MID1
ADT3_HUMAN	SLC25A6 ANT3 CDABP0051
ATSF1_HUMAN	ATP5PB ATP5F1
ATP5H_HUMAN	ATP5PD ATP5H My032
ATP5I_HUMAN	ATP5ME ATP5I ATP5K
ATP5L_HUMAN	ATP5MG ATP5L
ATP6_HUMAN	MT-ATP6 ATP6 ATPASE6 MTATP6
ATPA_HUMAN	ATP5F1A ATP5A ATP5A1 ATP5AL2 ATPM
ATPG_HUMAN	ATP5F1C ATP5C ATP5C1 ATP5CL1
ATPK_HUMAN	ATP5MF ATP5J2 ATP5JL
ATPO_HUMAN	ATP5PO ATP5O ATPO
CE290_HUMAN	CEP290 BBS14 KIAA0373 NPHP6
CELF1_HUMAN	CELF1 BRUNOL2 CUGBP CUGBP1 NAB50
CMC2_HUMAN	SLC25A13 ARALAR2
CNOT6_HUMAN	CNOT6 CCR4 CCR4a KIAA1194
DHX15_HUMAN	DHX15 DBP1 DDX15
EF1A1_HUMAN	EEF1A1 EEF1A EF1A LENG7
ELAV1_HUMAN	ELAVL1 HUR
GRSF1_HUMAN	GRSF1
HBS1L_HUMAN	HBS1L HBS1 KIAA1038
HNRL1_HUMAN	HNRNPUL1 E1BAP5 HNRPUL1
IF2B_HUMAN	EIF2S2 EIF2B
LFA3_HUMAN	CD58 LFA3
LN28B_HUMAN	LIN28B CSDD2
MPCP_HUMAN	SLC25A3 PHC OK/SW-cl.48
MSI2H_HUMAN	MSI2
NEPR1_HUMAN	CNEP1R1 C16orf69 TMEM188
PTBP1_HUMAN	PTBP1 PTB
PTBP3_HUMAN	PTBP3 ROD1
QKI_HUMAN	QKI HKQ
RAB8A_HUMAN	RAB8A MEL RAB8
RAN_HUMAN	RAN ARA24 OK/SW-cl.81
RAP1B_HUMAN	RAP1B OK/SW-cl.11
RHOA_HUMAN	RHOA ARH12 ARHA RHO12
ROA2_HUMAN	HNRNPA2B1 HNRPA2B1
RPA43_HUMAN	TWISTNB
RPB3_HUMAN	POLR2C A-152E5.7
SCAM2_HUMAN	SCAMP2
SRSF1_HUMAN	SRSF1 ASF SF2 SF2P33 SFRS1 OK/SW-cl.3
SRSF9_HUMAN	SRSF9 SFRS9 SRP30C
TET2_HUMAN	TET2 KIAA1546 Nbla00191
TIM8A_HUMAN	TIMM8A DDP DDP1 TIM8A
TRBP2_HUMAN	TARBP2 TRBP
ZCCHV_HUMAN	ZC3HAV1 ZC3HDC2 PRO1677
ZFR_HUMAN	ZFR
ZN346_HUMAN	ZNF346 JAZ

C)

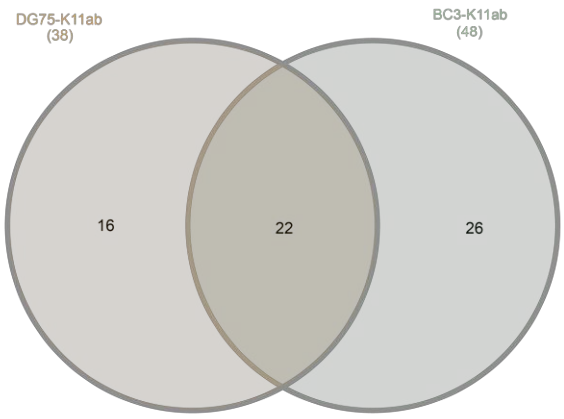
SL-miRNA-let7 DG75 specific candidates	
ACTBL_HUMAN	ACTBL2
CD2B2_HUMAN	CD2BP2 KIAA1178
CELF2_HUMAN	CELF2 BRUNOL3 CUGBP2 ETR3 NAPOR
DCD_HUMAN	DCD AIDD DSEP
ERI1_HUMAN	ERI1 3'EXO THEX1
HCLS1_HUMAN	HCLS1 HS1
MOT1_HUMAN	SLC16A1 MCT1
NMD3A_HUMAN	GRIN3A KIAA1973
PHOCN_HUMAN	MOB4 MOB3 MOBKL3 PHOCN PREI3 CGI-95
PSB6_HUMAN	PSMB6 LMPY Y
PTBP2_HUMAN	PTBP2 NPTB PTB PTBLP
RPB2_HUMAN	POLR2B
S10A8_HUMAN	S100A8 CAGA CFAG MRP8
VDAC1_HUMAN	VDAC1 VDAC
ZG16B_HUMAN	ZG16B UNQ773/PRO1567

D)

SL-miRNA-let7 common candidates	
DHX9_HUMAN	DHX9 DDX9 LKP NDH2
DSRAD_HUMAN	ADAR ADAR1 DSRAD G1P1 IFI4
E2AK2_HUMAN	EIF2AK2 PKR PRKR
HNRH3_HUMAN	HNRNPH3 HNRPH3
IF2G_HUMAN	EIF2S3 EIF2G
ILF2_HUMAN	ILF2 NF45 PRO3063
ILF3_HUMAN	ILF3 DRBF MPHOSPH4 NF90
LA_HUMAN	SSB
MET16_HUMAN	METTL16 METT10D
PRKRA_HUMAN	PRKRA PACT RAX HSD-14 HSD14
RAB7A_HUMAN	RAB7A RAB7
ROA0_HUMAN	HNRNPA0 HNRPA0
ROA1_HUMAN	HNRNPA1 HNRPA1
ROA3_HUMAN	HNRNPA3 HNRPA3
STRBP_HUMAN	STRBP SPNR
TRIM4_HUMAN	TRIM4 RNF87

Figure 38. Differential analysis of the DG75 and Day3 RNA chromatography dataset for the SL-miRNA let-7a-1. A) Venn diagram of the bait enriched proteins. B) Specific BC3 let-7a-1 bait enriched protein. C) Specific DG75 let-7a-1 bait enriched protein. D) Common BC3 and DG75 let-7a-1 bait enriched protein.

A)



B)

SL-miRNA-K11 common candidates	
ABCE1_HUMAN	ABCE1 RLI RNASEL1 RNASELI RNS4I OK/SW-cl.40
DHX15_HUMAN	DHX15 DBP1 DDX15
DHX9_HUMAN	DHX9 DDX9 LKP NDH2
DSRAD_HUMAN	ADAR ADAR1 DSRAD G1P1 IFI4
E2AK2_HUMAN	EIF2AK2 PKR PRKR
ERI1_HUMAN	ERI1 3'EXO THEX1
HBS1L_HUMAN	HBS1L HBS1 KIAA1038
HNRH3_HUMAN	HNRNPH3 HNRPH3
HNRL1_HUMAN	HNRNPUL1 E1BAP5 HNRPUL1
ILF2_HUMAN	ILF2 NF45 PRO3063
ILF3_HUMAN	ILF3 DRBF MPHOSPH4 NF90
LA_HUMAN	SSB
MBNL1_HUMAN	MBNL1 EXP KIAA0428 MBNL
MET16_HUMAN	METTL16 METT10D
MOT1_HUMAN	SLC16A1 MCT1
PRKRA_HUMAN	PRKRA PACT RAX HSD-14 HSD14
RAN_HUMAN	RAN ARA24 OK/SW-cl.81
STRBP_HUMAN	STRBP SPNR
TRIM4_HUMAN	TRIM4 RNF87
ZCCHV_HUMAN	ZC3HAV1 ZC3HDC2 PRO1677
ZFR_HUMAN	ZFR
ZN346_HUMAN	ZNF346 JAZ

C)

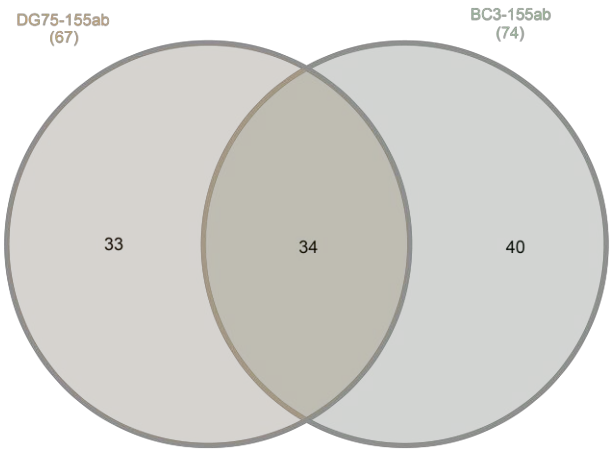
SL-miRNA-K11 BC3 specific candidates	
1B08_HUMAN	HLA-B HLAB
3MG_HUMAN	MPG AAG ANPG MID1
ATP6_HUMAN	MT-ATP6 ATP6 ATPASE6 MTATP6
CNOT6_HUMAN	CNOT6 CCR4 CCR4a KIAA1194
EF1A1_HUMAN	EEF1A1 EEF1A EF1A LENG7
EXOC7_HUMAN	EXOC7 EXO70 KIAA1067
GRSF1_HUMAN	GRSF1
IF2B_HUMAN	EIF2S2 EIF2B
LFA3_HUMAN	CD58 LFA3
LN28B_HUMAN	LIN28B CSDD2
PHAX_HUMAN	PHAX RNUXA
PHS2_HUMAN	PCBD2 DCOH2 DCOHM
PTBP1_HUMAN	PTBP1 PTB
PURB_HUMAN	PURB
RAB2A_HUMAN	RAB2A RAB2
RAB6B_HUMAN	RAB6B
RAB8A_HUMAN	RAB8A MEL RAB8
RAP1B_HUMAN	RAP1B OK/SW-cl.11
RBMS1_HUMAN	RBMS1 C2orf12 MSSP MSSP1 SCR2
RECQ1_HUMAN	RECQL RECQ1 RECQL1
RHOA_HUMAN	RHOA ARH12 ARHA RHO12
RPA43_HUMAN	TWISTNB
TET2_HUMAN	TET2 KIAA1546 Nbla00191
TIM8A_HUMAN	TIMM8A DDP DDP1 TIM8A
TRBP2_HUMAN	TARBP2 TRBP
TRUA_HUMAN	PUS1 PP8985

D)

SL-miRNA-K11 DG75 specific candidates	
ACTB_HUMAN	ACTB
ADA10_HUMAN	ADAM10 KUZ MADM
BMP2K_HUMAN	BMP2K BIKE HRIHFB2017
CSCL1_HUMAN	TMEM63A KIAA0489 KIAA0792
DYST_HUMAN	DST BP230 BP240 BPAG1 DMH DT KIAA0728
EF1A2_HUMAN	EEF1A2 EEF1AL STN
EF1A3_HUMAN	EEF1A1P5 EEF1AL3
EPHA4_HUMAN	EPHA4 HEK8 SEK TYRO1
HBB_HUMAN	HBB
HCLS1_HUMAN	HCLS1 HS1
IF2B1_HUMAN	IGF2BP1 CRDBP VICKZ1 ZBP1
IF2G_HUMAN	EIF2S3 EIF2G
LSG1_HUMAN	LSG1
RAB7A_HUMAN	RAB7A RAB7
RPB2_HUMAN	POLR2B
VDAC1_HUMAN	VDAC1 VDAC

Figure 39. Differential analysis of the DG75 and Day3 RNA chromatography dataset for the SL-miRNA -11. A) Venn diagram of the bait enriched proteins. B) Common BC3 and DG75 SL-miRNA -11 bait enriched protein. C) Specific BC3 SL-miRNA -11 bait enriched protein. D) Specific DG75 SL-miRNA -11 bait enriched protein.

A)



B)

SL-miRNA-155 common candidates	
ABCE1_HUMAN	ABCE1 RLI RNASEL1 RNASELI RNS4I OK/SW-cl.40
ATPA_HUMAN	ATP5F1A ATP5A ATP5A1 ATP5AL2 ATPM
ATPG_HUMAN	ATP5F1C ATP5C ATP5C1 ATP5CL1
ATPO_HUMAN	ATP5PO ATP5O ATPO
CMTR1_HUMAN	CMTR1 FTSJD2 KIAA0082 MTR1
DAZP1_HUMAN	DAZAP1
DHX15_HUMAN	DHX15 DBP1 DDX15
DHX9_HUMAN	DHX9 DDX9 LKP NDH2
DSRAD_HUMAN	ADAR ADAR1 DSRAD G1P1 IFI4
E2AK2_HUMAN	EIF2AK2 PKR PRKR
ERI1_HUMAN	ERI1 3'EXO THEX1
HBS1L_HUMAN	HBS1L HBS1 KIAA1038
HNRL1_HUMAN	HNRNPUL1 E1BAP5 HNRPUL1
IF2B_HUMAN	EIF2S2 EIF2B
IF2G_HUMAN	EIF2S3 EIF2G
ILF2_HUMAN	ILF2 NF45 PRO3063
ILF3_HUMAN	ILF3 DRBF MPHOSPH4 NF90
LA_HUMAN	SSB
LPPRC_HUMAN	LRPPRC LRP130
LSG1_HUMAN	LSG1
MOT1_HUMAN	SLC16A1 MCT1
PRKRA_HUMAN	PRKRA PACT RAX HSD-14 HSD14
PTBP1_HUMAN	PTBP1 PTB
PTBP3_HUMAN	PTBP3 ROD1
QKI_HUMAN	QKI HKQ
RAB7A_HUMAN	RAB7A RAB7
RAN_HUMAN	RAN ARA24 OK/SW-cl.81
ROA0_HUMAN	HNRNPA0 HNRPA0
STRBP_HUMAN	STRBP SPNR
TRIM4_HUMAN	TRIM4 RNF87
YBOX3_HUMAN	YBX3 CSDA DBPA
ZCCHV_HUMAN	ZC3HAV1 ZC3HDC2 PRO1677
ZFR_HUMAN	ZFR
ZN346_HUMAN	ZNF346 JAZ

C)

SL-miRNA-155 BC3 specific candidates	
ABC3D_HUMAN	APOBEC3D
ADT2_HUMAN	SLC25A5 ANT2
ADT3_HUMAN	SLC25A6 ANT3 CDABP0051
ANX11_HUMAN	ANXA11 ANX11
AT5F1_HUMAN	ATP5PB ATP5F1
ATP5H_HUMAN	ATP5PD ATP5H My032
ATP5I_HUMAN	ATP5ME ATP5I ATP5K
ATP5L_HUMAN	ATP5MG ATP5L
ATP6_HUMAN	MT-ATP6 ATP6 ATPASE6 MTATP6
ATPB_HUMAN	ATP5F1B ATP5B ATPMB ATPSB
ATPK_HUMAN	ATP5MF ATP5J2 ATP5JL
CIRBP_HUMAN	CIRBP A18HNRNP CIRP
CMC2_HUMAN	SLC25A13 ARALAR2
CNOT6_HUMAN	CNOT6 CCR4 CCR4a KIAA1194
COX2_HUMAN	MT-CO2 COII COX2 COXII MTCO2
EF1A1_HUMAN	EEF1A1 EEF1A EF1A LENG7
GARS_HUMAN	GARS
HNRPD_HUMAN	HNRNPD AUF1 HNRPD
LFA3_HUMAN	CD58 LFA3
MET16_HUMAN	METTL16 METT10D
MPCP_HUMAN	SLC25A3 PHC OK/SW-cl.48
NUCL_HUMAN	NCL
PHB_HUMAN	PHB
PHS2_HUMAN	PCBD2 DCOH2 DCOHM
PNPT1_HUMAN	PNPT1 PNPASE
PURB_HUMAN	PURB
RAB6B_HUMAN	RAB6B
RAB8A_HUMAN	RAB8A MEL RAB8
RAP1B_HUMAN	RAP1B OK/SW-cl.11
RB11A_HUMAN	RAB11A RAB11
RECQ1_HUMAN	RECQL RECQ1 RECQL1
RHOA_HUMAN	RHOA ARH12 ARHA RHO12
RPA43_HUMAN	TWISTNB
SCAM2_HUMAN	SCAMP2
SLIRP_HUMAN	SLIRP C14orf156 DC23 DC50 PD04872
SQOR_HUMAN	SQOR SQRDL CGI-44
TET2_HUMAN	TET2 KIAA1546 Nbla00191
TIM8A_HUMAN	TIMM8A DDP DDP1 TIM8A
TRBP2_HUMAN	TARBP2 TRBP
YBOX1_HUMAN	YBX1 NSEP1 YB1

D)

SL-miRNA-155 DG75 specific candidates	
ACTBL_HUMAN	ACTBL2
ARP3_HUMAN	ACTR3 ARP3
BMP2K_HUMAN	BMP2K BIKE HRIHFB2017
C1TC_HUMAN	MTHFD1 MTHFC MTHFD
CD2B2_HUMAN	CD2BP2 KIAA1178
CELF2_HUMAN	CELF2 BRUNOL3 CUGBP2 ETR3 NAPOR
COEA1_HUMAN	COL14A1 UND
CSCL1_HUMAN	TMEM63A KIAA0489 KIAA0792
CSTF1_HUMAN	CSTF1
CTNA1_HUMAN	CTNNA1
DYN2_HUMAN	DNM2 DYN2
DYST_HUMAN	DST BP230 BP240 BPAG1 DMH DT KIAA0728
EF1A2_HUMAN	EEF1A2 EEF1AL STN
EF1A3_HUMAN	EEF1A1P5 EEF1AL3
EHD1_HUMAN	EHD1 PAST PAST1 CDABP0131
EPHA4_HUMAN	EPHA4 HEK8 SEK TYRO1
FAT3_HUMAN	FAT3 CDHF15 KIAA1989
FUBP3_HUMAN	FUBP3 FBP3
FUMH_HUMAN	FH
HCLS1_HUMAN	HCLS1 HS1
IF2A_HUMAN	EIF2S1 EIF2A
IF2B1_HUMAN	IGF2BP1 CRDBP VICKZ1 ZBP1
MBNL1_HUMAN	MBNL1 EXP KIAA0428 MBNL
MCM7_HUMAN	MCM7 CDC47 MCM2
NAC1_HUMAN	SLC8A1 CNC NCX1
NEPR1_HUMAN	CNEP1R1 C16orf69 TMEM188
PTBP2_HUMAN	PTBP2 NPTB PTB PTBLP
RAP1A_HUMAN	RAP1A KREV1
RM46_HUMAN	MRPL46 C15orf4 LIECG2
RPB2_HUMAN	POLR2B
SYFB_HUMAN	FARSB FARSLB FRSB HSPC173
URP2_HUMAN	FERMT3 KIND3 MIG2B URP2
XP32_HUMAN	XP32 C1orf68

Figure 40. Differential analysis of the DG75 and Day3 RNA chromatography dataset for the SL-miRNA -155. A) Venn diagram of the bait enriched proteins. B) Common BC3 and DG75 SL-miRNA -155 bait enriched protein. C) Specific BC3 SL-miRNA -155 bait enriched protein. D) Specific DG75 SL-miRNA -155 bait enriched protein.

III.A.4.b.1.1 SL-miRNA-let7a biological triplicate

As we always used the let-7a-1 SL-miRNA in our pulldown experiments, we have a biological triplicate for this miRNA precursor upon RNA chromatography within BC3 lymphocyte nuclear extract. It is thus interesting to analyse this biological triplicate to gain some insight on the reproducibility of the DNase assisted RNA chromatography experiment. The analysis of this biological triplicate is interesting to discuss the importance of the cell extract upon chromatography assay. Indeed, the proteome content of the cell is prone to variations thereby reducing protein identification reproducibility. (This is the reason why for a perfect experiment and upon biological triplicate, cells have to be synchronized to ensure a proteome stability). In addition, the nuclear extract preparation can also be a source of variability, as well as the pulldown procedure itself. Here, each let-7 chromatographies are compared without taking into account their corresponding control require for the enrichment ratio calculation. As we can see on the correlation matrix as well as the PCA (Figure 41.A.B.C), each day group is different from the next. Surprisingly, the differences are not bigger than various bait conditions within a specific chromatography day, which suggests that the cell extracts are not as different as what was expected.

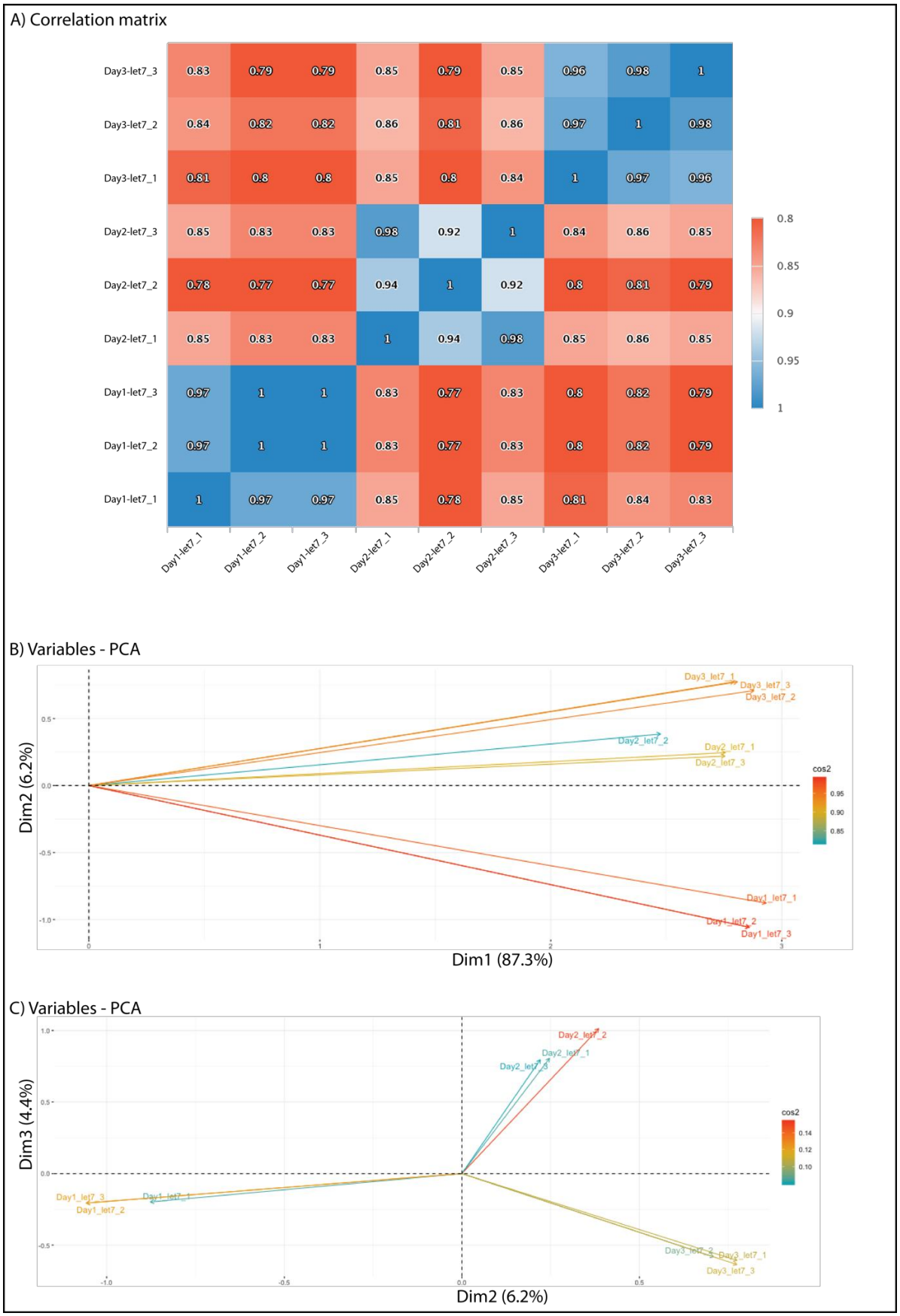


Figure 41. Descriptive analysis of the SL-miRNA-let7 biological triplicate dataset. (A) Correlation matrix of the chromatography baits. (B) Principal component analysis (PCA) with first two dimensions. (C) Principal component analysis with the second and third dimensions.

Here, I compare the enrichment ratio of each let-7 bait relative to their own control. As we can see on the Venn diagram and the protein list enriched within each day still display variation highlighting the variations of the experimental approach (Figure 42). This biological triplicate shows that 14 proteins are always enriched. Among this list, there are only relevant proteins such as microprocessor's cofactor (ILF2, ILF3, DHX9, DHX15), dsRNA binding protein (ADAR, PKR), let-7 specific cofactor (LN28B, ROA1), interesting potential SL-miRNA binding protein (STRBP, MSI2, ROA3, HNRH3), commonly identified contaminant (SSB) and finally a subunit of the polymerase II (POLR2C) which is implicated in ncRNA biogenesis. This result greatly emphasizes the importance of biological triplicate to increase relevant candidate list confidence. It is important to keep in mind that miRNA biogenesis is greatly impacted by proteome changes allowing miRNA to be expressed according to the cell status.

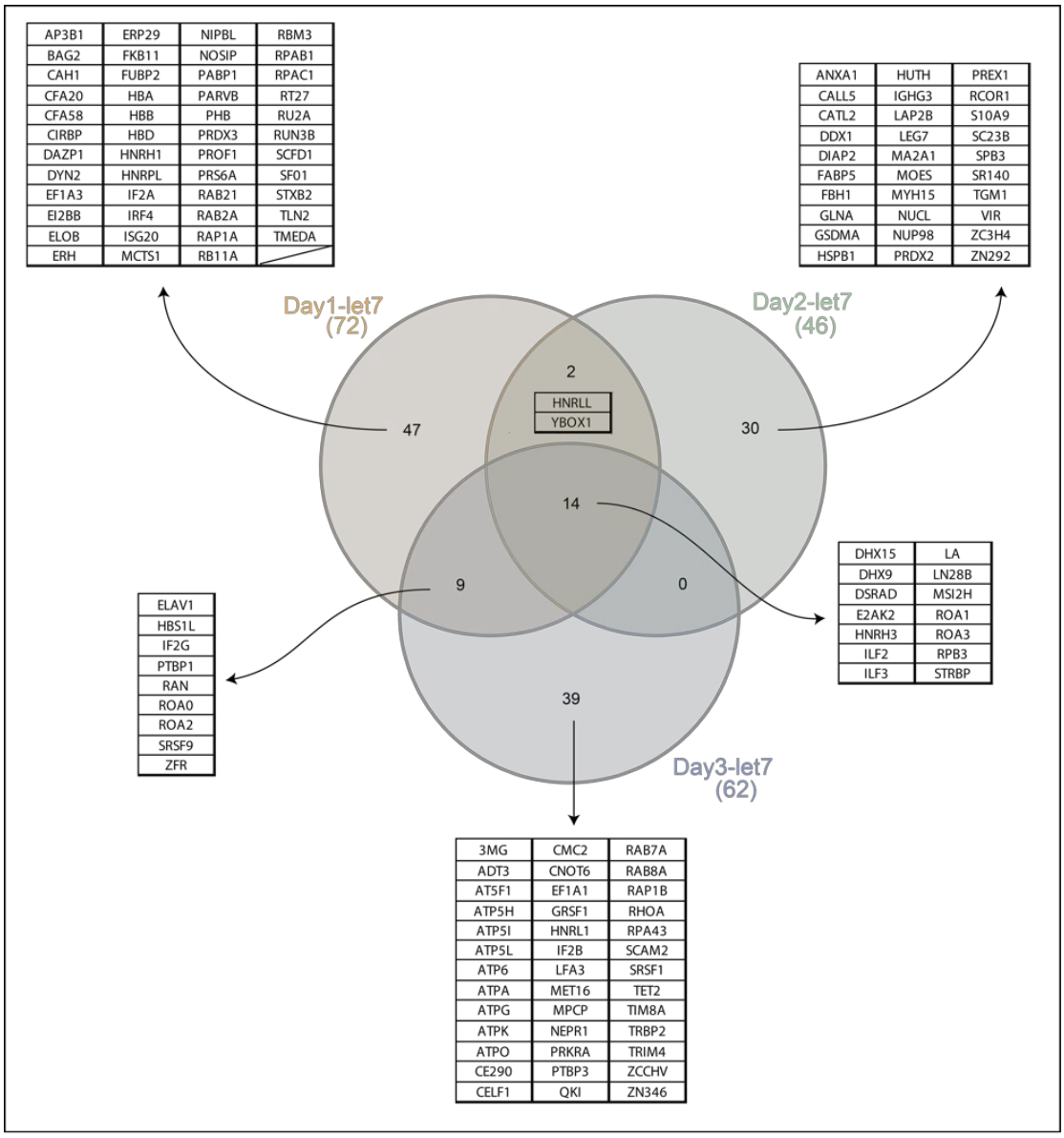


Figure 42. Enriched dataset comparison of the SL-miRNA-let7 biological triplicate generated with the DNase assisted RNA chromatography. Venn diagram shows the common or specific significantly enriched protein of each technical replicate (uniport IDs).

III.B. Candidates validation

Prior to assess the role of a candidate in miRNA biogenesis, the proteins identification by our mass spectrometry approach can be verified by western blot analysis. In the event that no antibody is readily available for the endogenous protein, we can also resort to the expression of a tagged version of the candidate to verify its interaction with a specific pre-miRNA by RNA pulldown. Once the identification and the interaction has been validated, the involvement of the candidate proteins on miRNA maturation can be assessed. This can be done by modulating the expression of the candidate either by silencing or overexpression and measuring the effect on miRNA levels with different methods such as luciferase reporters, qRT-PCR or northern blot analysis.

III.B.1 Tools developed for the validation

Here, I will describe the different tools that we generated to modulate candidate expression as well as the tools require to assess any variation in the studied miRNA biogenesis.

III.B.1.a Modulation of candidate expression

Even though proteins with no known RNA-binding activity have been showed to act as RBP, as described for metabolic enzymes displaying moonlighting activity (Jeffery, 2015; Ribeiro *et al.*, 2019), I cautiously limited my selection to proteins with well-characterized RNA binding activity. I picked twelve candidates, some of which have already been implied in regulating miRNA biogenesis. These are listed here together with publications link to their RNA binding ability or molecular functions (Table 8). Selected candidates are discussed in part IV.C.1. For those twelve candidates, I either cloned their cDNA within a lentiviral vector (pLenti6.3) for overexpression or an shRNA targeting their mRNA within another lentiviral vector (pLKO.1-TRC) to silence them.

Table 8. Selected candidates.

Candidate	SL-miRNA	Functions	Reference
CIRBP	-K1, -155	3'UTR binding, translation modulation, mRNA localization, p53 interaction, inflammation modulation, miRNA biogenesis	(Lujan *et al.*, 2018; Downie Ruiz Velasco *et al.*, 2019)
ELAV1	-K2, -let7	Nuclear export, mRNA localization, miRNA biogenesis, miRISC destabilization, splicing	(Kundu *et al.*, 2012; Choudhury *et al.*, 2013; Noh *et al.*, 2016)
FUS		mRNA localization, splicing, transcription, DNA repair	(Morlando *et al.*, 2012; Ederle and Dormann, 2017; Rizzuti *et al.*, 2018; Zhang *et al.*, 2018)
HNRPD	-K9, -155	RNA chaperone, RNA annealing, AU-rich binding, DNA binding, mRNA translation, mRNA turnover	(Lee *et al.*, 2014; Min *et al.*, 2017; White *et al.*, 2017; Meyer *et al.*, 2019)
HNRPF	-K6	Splicing, mRNA localization, G-rich binding, mRNA turn over	(Huang *et al.*, 2017; Du *et al.*, 2018; Das *et al.*, 2019)
LN28B	-K1, -K2, -K11, -let7	miRNA biogenesis, loop binding, mRNA binding	(Heo *et al.*, 2008; Busch *et al.*, 2016; Peters *et al.*, 2016)
MET16	-K11, -155, -let7	m^6A modification, MAT2A transcript regulation, U6 snRNA modification, specific RNA structure recognition, lncRNA binding	(Alarcón *et al.*, 2015; Doxtader *et al.*, 2018; Mendel *et al.*, 2018)
MSI2H	-K6, -let7	miRNA biogenesis, translation regulation, mRNA binding, cell cycle regulation	(Choudhury *et al.*, 2013; Kharas and Lengner, 2017; Kudinov *et al.*, 2017)
QKI	-K11, -155	mRNA stability, splicing, mRNA localization, translation regulation, miRNA biogenesis	(Wang *et al.*, 2013; Conn *et al.*, 2015; Tili *et al.*, 2015; F. Wang *et al.*, 2017)
RBM45	-K1	DNA damage, poly(C) binding, splicing, translation control	(Li, Collins, *et al.*, 2016; Gong *et al.*, 2017)
ROA1	-K1, -K2, -let7	miRNA biogenesis, splicing, mRNA localization, translation regulation	(Guil and Caceres, 2007; Michlewski and Cáceres, 2010; Kooshapur *et al.*, 2018; Levengood and Tolbert, 2019)
ZNF346	-155	DNA binding, dsRNA binding, RNA processing	(Chen *et al.*, 2004; Burge *et al.*, 2014)

III.B.1.a.1.1 siRNA & shRNA

Loss of function is generally used to assess protein implication for a given mechanism. Knocking-out candidate is the more efficient method to characterize its role, but it is more time-consuming to setup and cannot be done for essential gene as opposed to knock down approaches. I decided to knock-down candidates with either siRNA duplexes or short-hairpin-RNA (shRNA) for respectively transient or stable candidate silencing. SMARTpool siRNAs have been purchased from Dharmacon (ON-TARGETplus siRNA). siRNAs that are not supposed to silence any mRNA are used as negative controls (si-CTRL). Concerning the stable knock-down, I picked the pLKO.1-TRC backbone plasmid composed of a polymerase III promoter used to express shRNA designed by the Genetic Perturbation Platform website (Broad Institute) (Root *et al.*, 2006). shRNAs are designed with an algorithm that assesses the best 21mer shRNA starting 25 bp after the start codon of the candidate coding sequence. Scores are calculated for all the potential shRNA to penalize or reward knock-down *in silico* specificity via miRNA seed matches. I selected the best shRNA proposed by the algorithm.

The pLKO.1-TRC vector allows the production of lentiviral particles. Prior to any pLKO lentiviral production, pLKO-shRNA directed against my candidates are transfected in HEK293-Grip cell to assess the efficiency of the silencing in a transient manner. I checked the mRNA level of the target candidate by qRT-PCR analysis (Figure 43). siRNAs, for their part, all display a silencing activity superior to 70 % on every candidate mRNA tested. The silencing kinetic of an shRNA is generally slower than the siRNA because it has to be transcribed and processed to form the siRNA duplex contrary to the siRNAs. For this reason, the use of shRNA-pLKO.1 is not optimal to silence mRNA in a short period of time. However, pLKO.1 constructs will be incorporated within lentiviral particles that will be used to generate stable cell lines that will be used in long-term experiment. Here, only the sh-LN28B does not show any silencing activity. Other shRNAs display slight (sh-MET16 ~20%) to efficient (sh-HNRPF ~silencing activity, except sh-HNRPF which is efficient (Figure 43).

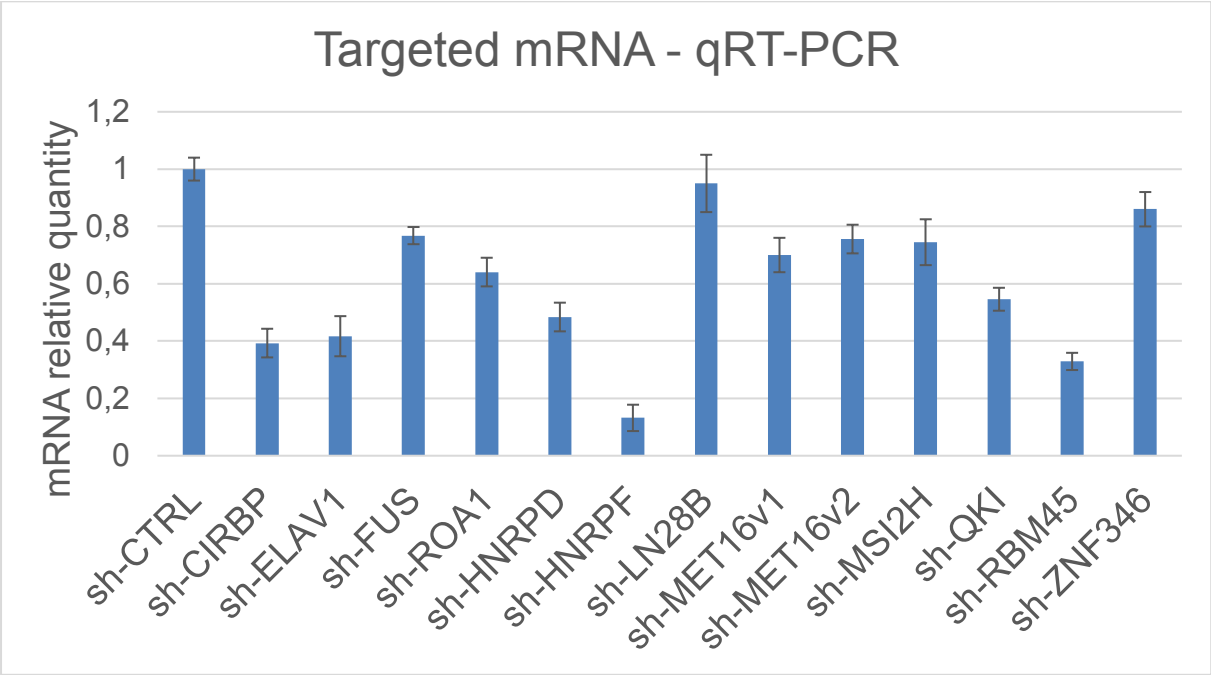


Figure 43. Silencing activity of the shRNA on the candidate mRNA. Reverse transcription steps are done with oligo random nonamers. Each specific pLKO-shRNA are compared to the non-targeting sh-CTRL control.

III.B.1.a.1.2 Overexpression of HA-tag

In order to modulate candidate expression in different system, some candidates have been cloned downstream a CMV promotor and a Flag/Ha tag sequence within pLenti6.3 expression vector, to either over-express the candidate in any transfectable cell line or to produce lentiviral particles. Here, cloning steps have been done with gateway recombination sequences. Thus, additional sequences are present upstream and downstream candidate's ORF. Moreover, Flag/Ha-tag sequence also increases the amino-acid sequence length of the candidate of about 36 amino-acids (~4.1 kDa). Presence of these different "N-ter extensions" can disturbed candidate's activity. A GFP construction is also generated, using the exact same approach, to act as a transfection/transduction control.

For my project, I simply controlled the protein molecular weight (Figure 44.A) as well as the SL-miRNAs interacting properties previously identified by RNA-pulldown (Figure 49). In the Figure 44.A, protein concentration has not been assessed prior to SDS-PAGE loading. Nonetheless, western blot analysis shows that proteins seem expressed at different level. In addition, lower bands appear for some construction (HNRPD, ELAV1, RBM45). These lower bands could correspond to degradation products. On the other hand, FUS construct displayed a greater length than expected but antibody's provider also reveals FUS protein around 70 kDa (Abcam ab23439). When transfected, the pLenti6.3-GFP allows a good detection of the GFP fluorescent signal easily detected less than 24 h post transfection or after 48 h post transduction (Figure 44.B). Overall protein length, together with correct nucleic acid sequence of the candidate within the plasmids suggest that those construction are usable. pLenti6.3 backbone seems efficient to express our proteins of interest and will allow the production of lentiviral particles.

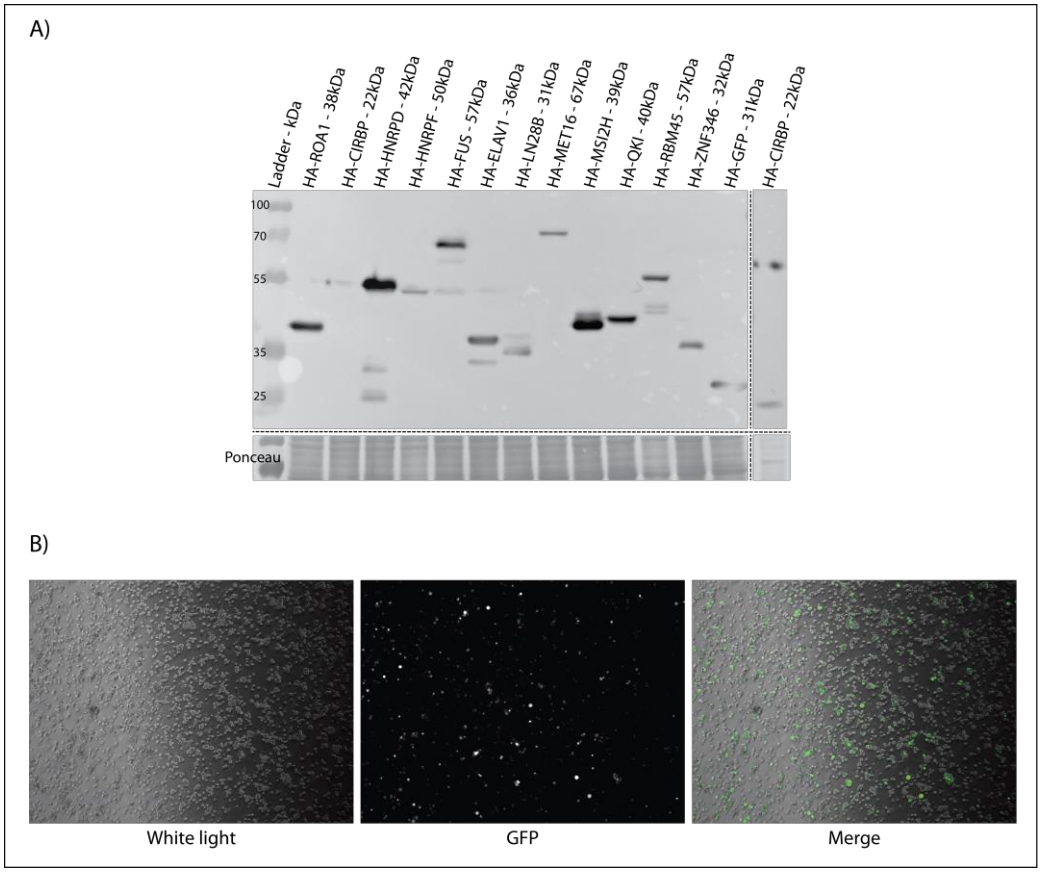


Figure 44. pLenti6.3 expression control. A) Western blot analysis of the tag-candidates expressed within HEK-Grip cell during 48 h. B) Transduced lymphocyte with pLenti6.3-GFP. GFP signal is obtained 48h post-transduction and after 300 ms of exposition.

III.B.1.b Luciferase reporters

Luciferase reporters have been generated to screen the involvement of the identified pulldown proteins in the regulation of miRNA processing. The purpose of luciferase reporters is to easily screen for candidates having a negative or positive impact on miRNA biogenesis in order to reduce the candidates list and select only biologically suspected relevant candidate for

further characterization. Two different kind of reporters are used to either assess the DROSHA cleavage activity of a pre-miRNA inserted within the 3'UTR of firefly luciferase mRNA (pmiRGLO, Drosha cleavage reporter) or assess the activity of the mature miRNA matching a corresponding miRNA binding site within the 3'UTR of firefly luciferase mRNA (psiCHECK-2, miRNA activity reporter). Both pmiRGLO and psiCHECK-2 possess two luciferase reporters. First, *firefly* mRNA is used as the primary reporter gene modulated by DROSHA cleavage (pmiRGLO) or mature miRNA (psiCHECK-2) activity. Second, *Renilla* luciferase allows reporter transfection normalization of the *firefly* luciferase expression. This feature is fundamental to correctly compare our experimental conditions. Reporters are transfected concomitantly with candidate proteins silencing or over-expression. Modulation of the *firefly* activity upon candidate expression variations compared to controls allows the determination of the positive or negative effect of the candidate on the SL-miRNA Drosha cleavage (pmiRGLO) or the SL-miRNA global biogenesis (psiCHECK-2). As both reporters express their luciferase reporters from a strong viral promotor either SV40 or HSV-TK for respectively pmiRGLO and psiCHECK-2, candidate-induced transcriptional modulation can be excluded.

III.B.1.b.1.1 DROSHA cleavage reporter

We cloned the SL-miRNAs used for the different RNA-pulldown assays in the 3'UTR of the pmiRGLO's *firefly* mRNA using Gateway technology. Cleavage of the SL-miRNA by the microprocessor allows to monitor the effect of potential co-factors by measuring luciferase activity (Figure 45). Candidate effect on the microprocessor activity can be positive or negative thus inducing decrease or increase of the luciferase signal respectively. pmiRGLO possessing the 3'UTR SL-miRNA is compared to a control pmirGLO without SL-miRNA upon candidate silencing or over-expression when it is available.

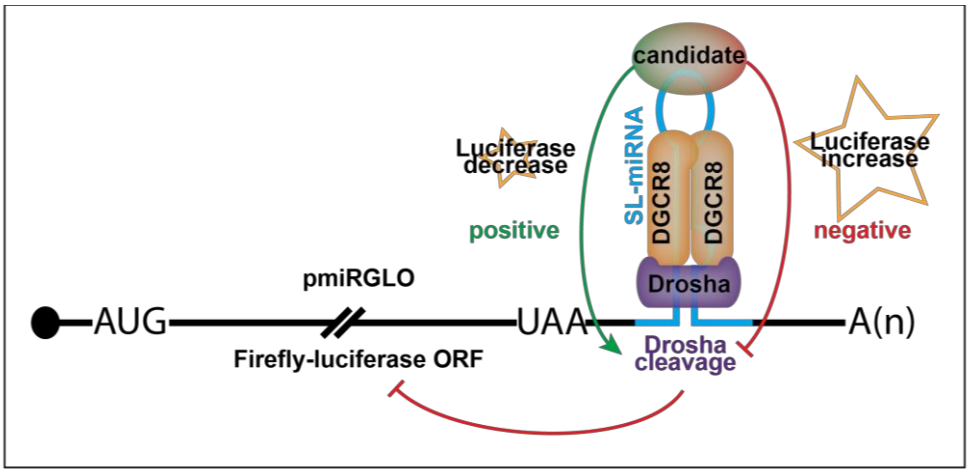


Figure 45. Schematic representation of the pmiRGLO reporter system. Candidate modulation is performed prior to pmiRGLO transfection. Candidate binding to the SL-miRNA can act as a positive or negative factor on the DROSHA processing inducing increase or decrease of the DROSHA cleavage respectively and thus luciferase mRNA destabilization or stabilization.

III.B.1.b.1.2 miRNA activity reporter

I also generated miRNA activity luciferase reporter, the psiCHECK-2 vector (Suffert *et al.*, 2011). Here, instead of inserting the SL-miRNA within the 3'UTR of the *firefly* luciferase, I inserted an imperfect miRNA binding site (bulge position 13 to 15) (Martin *et al.*, 2014) for the KSHV miRNAs, using gateway technology. psiCHECK-2 act as a mature miRNA activity reporter which implies that miRNA precursors have to expressed by the cell line and have to be processed by DROSHA and DICER before acting on the firefly mRNA (Figure 46). Knowing that each maturation steps of miRNA precursors can be modulated, psiCHECK-2 reporter will not discriminate processing alteration from DROSHA cleavage to ARGONAUTE loading and silencing. As the HEK293 cells used are not infected by KSHV, it is required to transfect a plasmid expressing either the full miRNA cluster or each individual pri-miRNA. It is important to mention here that if a plasmid expressing the KSHV cluster is used, modulation of a miRNA co-factor candidate can act either directly on its bound SL-miRNA or on its more or less distant neighbouring SL-miRNAs. Indeed, 3D organization of the cluster can bring a candidate, binding a specific SL-miRNA, in close proximity to a distant one. This close localisation can modulate DROSHA processing in a cis-acting distant manner. To solve this potential issue, it is therefore important to use single SL-miRNA expressing plasmids.

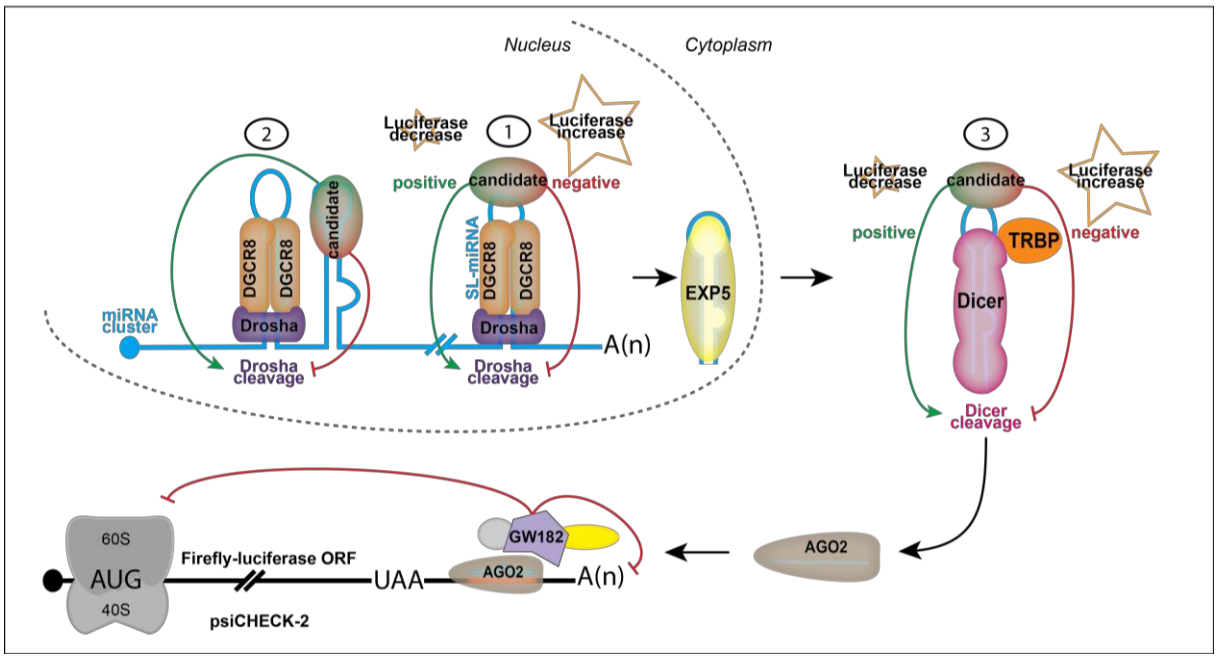


Figure 46. Schematic representation of the psiCHECK-2 reporter system. Candidate modulation is performed prior to psiCHECK-2 and plasmid expressing KSHV's full cluster or single SL-miRNA transfection. Candidate binding to the SL-miRNA can act positively or negatively on all the different miRNA processing steps. Positively acting candidate on the SL-miRNA processing will decrease the luciferase signal contrary to negatively acting candidate. Candidate can either act on its bound SL-miRNA (1), on its more or less distant neighbouring SL-miRNA (2) or on the DICER processing step (3).

III.B.1.c Northern blot & qRT-PCR

Pri-miRNA processing can be assessed with different approaches. Northern blot and qRT-PCR are the easiest way to do so. Probes and primers used are listed on Table 15, Table 16 and Table 17. Depending on the cell used for the experiments the normalization procedure varies.

For the northern blot analysis performed in a non-infected cell line, mature miRNA signal could be normalized by all the KSHV miRNA signals studied within the sample (e.g. $miRNA_K1_{norm} = miRNA_K1 / \Sigma (miRNA_{KSHV})$). Even though a specific cofactor can act or not on more or less distant neighboring SL-miRNA, overall KSHV miRNA signal should display a relative correlation with the cluster-expressing plasmid transfection in a non-infected cell line. With this normalization approach, a candidate that acts as a general cofactor will show no effect. Concerning the infected cell line, normalization procedure will be the same even though we can assume that the number of viral episome do not vary knowing that each experimental condition come from the same initial cell batch.

qRT-PCR analyses are performed with SYBR Green-based technology. In a non-infected cell line, cluster-expressing plasmid transfection could be assessed by looking at a non-processed cluster extremity, if we assume that these extremity parts are not rapidly degraded. Using different primer couples and reverse transcriptase enzymes, pri-miRNA and mature miRNA relative quantity can be determined (Figure 47.A). pre-miRNA molecules were reverse transcribed using the miRScript kit with a forward specific primer and a universal primer. Pri-miRNA are reverse transcribed using the superscript IV with two specific primers. In my hands, I managed to obtain good results only for the 5' and 3' cluster primer couples as well as all the pri-miRNA. For the other couples, I obtained off-target and non-satisfying amplification. Taq-man probes technology can be a good alternative to overcome this issue. For the infected cell line, the normalization procedure is the same as describe previously.

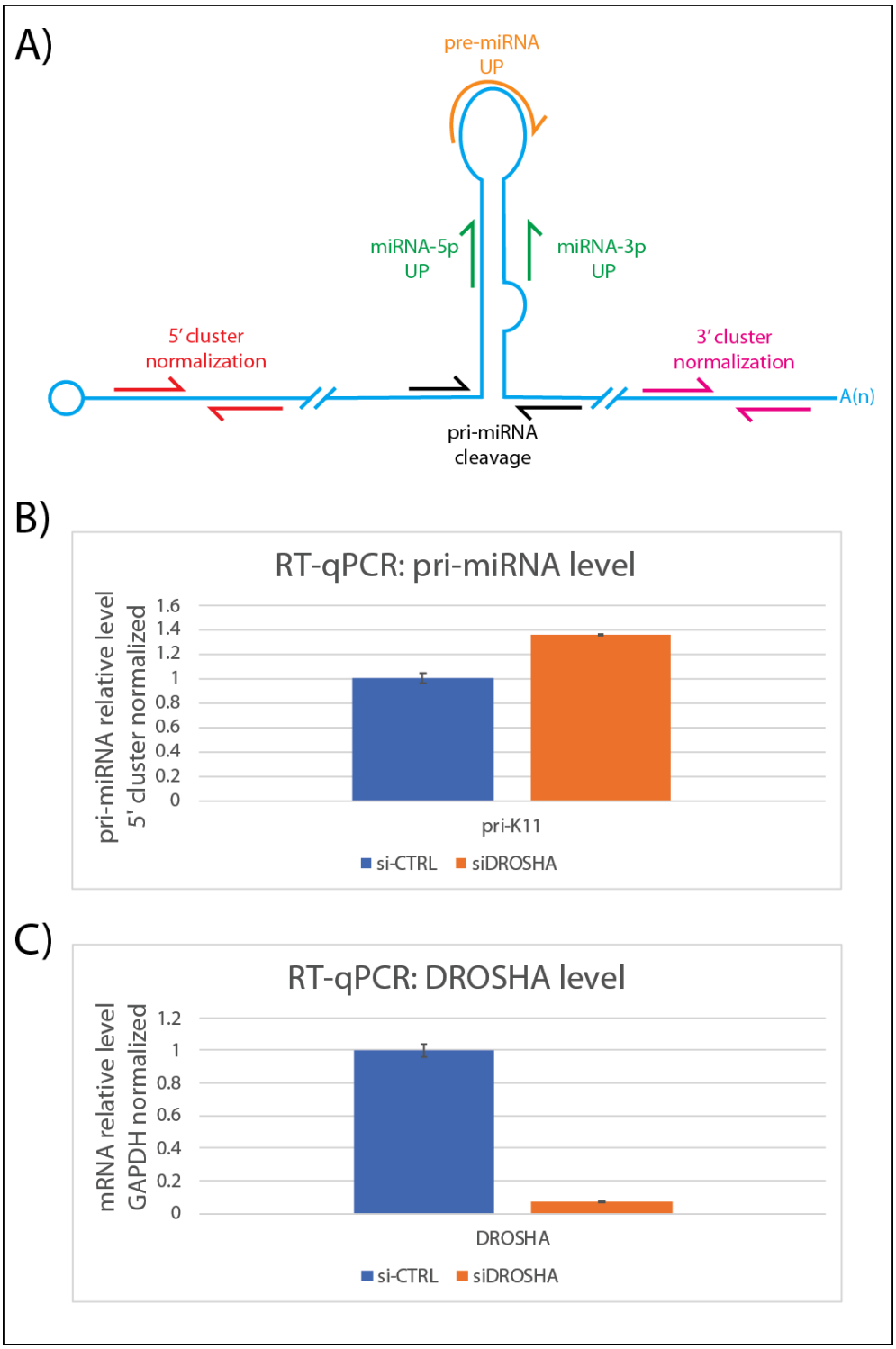


Figure 47. Schematic representation of the primer couples and pri-miRNA normalization. A) miRNA -5p and -3p as well as pre-miRNA can be amplify with a specific reverse transcriptase kit (miScript II) allowing the poly-adenylation and universal tag addition (UP) respectively in greem and orange. B) pri-miRNA level normalized with the 5' cluster signal. C) DROSHA mRNA level normalized with GAPDH sample signal.

With the qRT-PCR assays, it is fundamental to know if the primer couples used to assess the SL-miRNA excision from the cluster transcript are indeed modulated by a variation in DROSHA expression. As mentioned, I normalized the pri-miRNA signal with the 5' cluster amplification one. This normalization is correct only if the transcript is not rapidly degraded after each SL-miRNA DROSHA excision. With the qRT-PCR approach, each time a candidate displays a positive effect on DROSHA SL-miRNA-cleavage, the corresponding pri-miRNA signal should drop contrary to a negative-candidate effect on DROSHA which increases the pri-miRNA signal. In the Figure 47.B, it seems that siRNA treatment directed against DROSHA (Figure 47.C) does increase the pri-K11 signal as expected suggesting that the 5' cluster signal is stable enough even when the pri-miRNAs get cleaved.

III.B.2 Validation of protein involvement

All the tools previously describe are used to decipher candidate implication in miRNA biogenesis. First, I will start by describing the mass spectrometry identification validation, then confirm the RNA binding properties of the cloned candidates. Secondly, the luciferase screens are exposed. Finally, the impact of the candidate expression modulation on the miRNA biogenesis is described.

III.B.2.a Confirmation of interaction

Mass spectrometry protein identification can be control using western blot analysis. The RNA binding feature of the cloned candidate is also control with RNA pulldown to validate their usage for the following experiments.

III.B.2.a.1.1 Endogenous proteins

We performed a new RNA pulldown with agarose beads using BC3 nuclear extract in order to validate the mass spectrometry identification. The purpose here is to validate the identification by another approach. Antibodies are listed in VI.E.2.

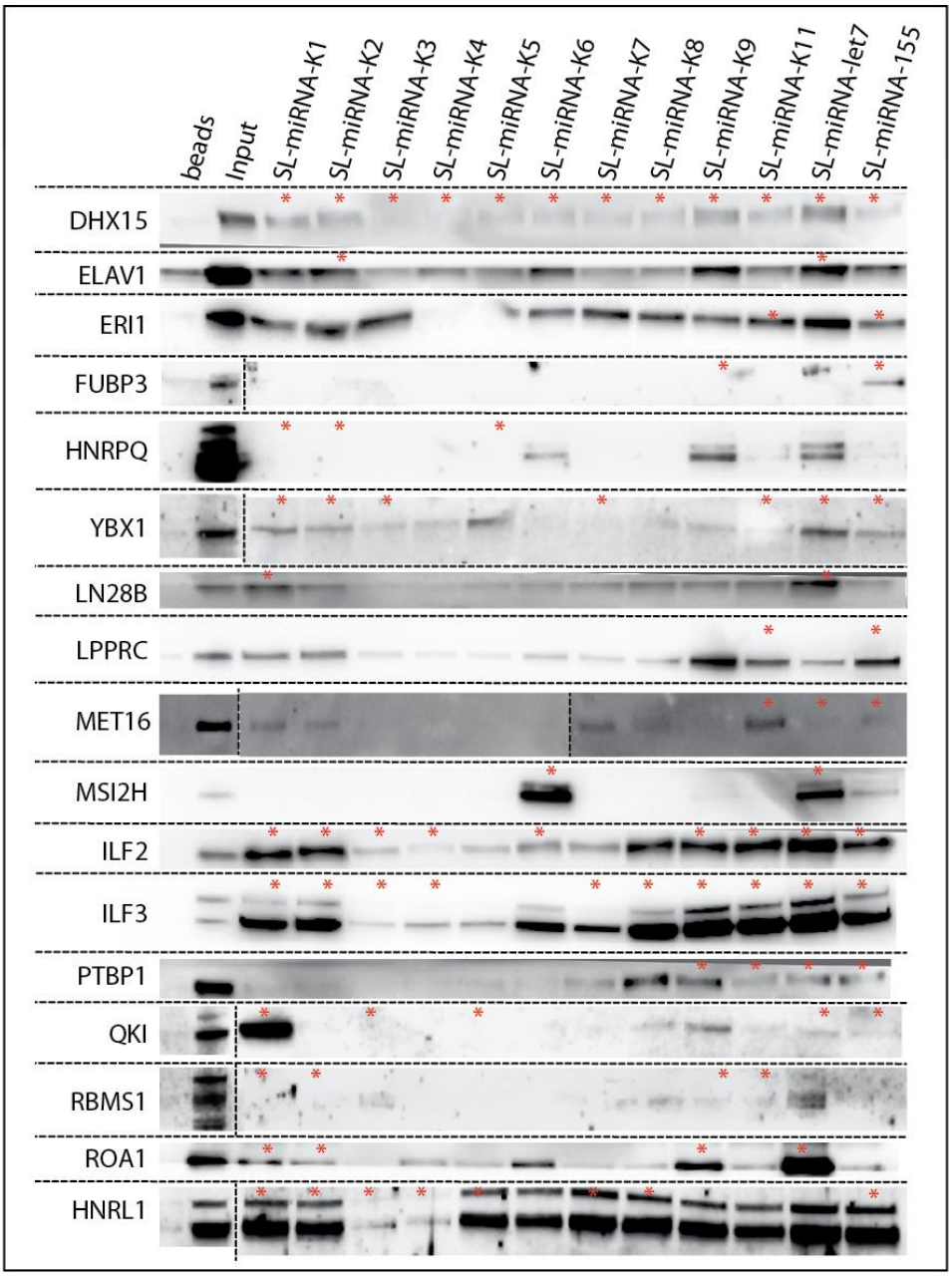


Figure 48. Western blot analysis of the agarose RNA-pulldown. Samples from the agarose RNA-pulldown are analysed by SDS-PAGE. Red stars represent the protein consider as significantly enriched by the mass spectrometry identification. Antibodies are listed in VI.E.2. Black dashed lines represent a cut on the raw data for the sake of the figure understanding.

On the Figure 48, we can see that the negative control (beads) mostly fail to interact with the tested proteins with the exception of ELAV1, which showed some background interaction. The known pri-let-7a-1 cofactors, ROA1 and LN28B clearly show a good enrichment for the SL-miRNA-let7. As expected from the MS analysis, the -K1 and -K2 baits also interact with LN28B and ROA1. Several proteins have been enriched, with varying efficiency, by all the SL-miRNA used. This result is not surprising for the protein ILF2, ILF3 and DHX15 which can be considered as general microprocessor cofactors (Gregory *et al.*, 2004). In the enriched dataset, proteins ERI1 and LPPRC were only enriched by the SL-miRNA-K11 and SL-miRNA-155. Here when assess by western blot it seems that all SL-miRNA managed to fish those proteins. When I explored the mass spectrometry raw data, both proteins were simply not identified by the mass spectrometer within the chromatography Day1 and Day2. Thus, it is possible that LPPRC and ERI1 interact with all the KSHV SL-miRNAs. The HNRL1 protein was almost enriched by all the SL-miRNA except SL-miRNA-K9 and -K11 in the mass spectrometry data. On this western blot, HNRL1 signal displays similar intensity in all the lanes except the SL-miRNA-K3 and -K4. ELAV1 shows a signal for all the SL-miRNA with different intensity. However, ELAV1 was supposed to only display a enriched signal for the SL-miRNA-let7 and -K2. On the mass spectrometry data, ELAV1 is identified within all the eluates but only consider as enriched for SL-miRNA-let7 and -K2. Thus, it could make sense to obtain signal on this western blot even though we can clearly see a better enrichment for SL-miRNA-K6 and -K9. The SL-miRNA-155 but not the -K9 show a signal for FUBP3 protein as it is expected by the mass spectrometry data. The protein HNRPQ shows a signal for the SL-miRNA-K6, -K9 and -let7 whereas it was expected to be enriched by SL-miRNA-K1, -K2 and -K5. In the raw data, HNRPQ has been identified by the different chromatography days with a majority of true MSMS spectra. I cannot explain this different output between the mass spectrometry and the western blot. This result can come from an important protein concentration difference between the eluate. Anyway, it highlights the importance to keep an open mind on the mass spectrometry data. HNRL1 was enriched in the mass spectrometry data with SL-miRNA-K1, -K2, -K3, -K7, -let7, -K11 and -155. On the western blot, baits -K1, -K2, -K5 and -let7 show a good enrichment, thereby partially validating the results. MET16 was correctly identified on the western blot by the SL-miRNA-K11 and -155 as expected. In addition, SL-miRNA-K1, -K2 and -K7 managed to give a signal for the MET16. In the raw data, MET16 was simply not identified in the first two chromatography. In the Day3, MET16 peptides (true MSMS) have been detected only for the SL-miRNA-K11 with few occurrences contrary to the other conditions in which the identification was done by peptide matching. This result demonstrates that the mass spectrometry identification misses peptide identification certainly due to low peptide separation upon exclusion chromatography performed before the ionisation step. MSI2H protein displays one of the better results with a great enrichment for both bait -K6 and -let7 as expected. It is interesting to see that a good MSI2H enrichment is correlated with a good enrichment of the ELAV1 protein which is clearly expected as those two proteins are already known to participate together to the maturation of the pri-miRNA-7-1 (Choudhury *et al.*, 2013). PTBP1 show important signal on the western blot for the bait -K8, -K9, -let7, -155, whereas the MS data showed that it interacts with baits -K1, -K2 and -K3. Here again, it can come from a similar issue as YBX1 protein. QKI was supposed to be enriched by the SL-

miRNA-K1, -K3, -K5, -let7 and -155. On this western blot, it shows a signal for -K1 and -K9 and very weak signal for -let7 and -155.

Altogether, our western blot analysis allowed us to partially validate some of the mass spectrometry data, while it invalidates other ones, which therefore must have been false positives.

III.B.2.a.1.2 HA-tagged proteins

As the cloned candidates possess an HA tag on their N-ter part, it is important to control if the candidates kept their SL-miRNA binding ability. New RNA-pulldown have been performed in a smaller scale (6-wells plate) with the agarose method due to its cheaper cost compared to the magnetic method (Figure 49). Few SL-miRNAs have been selected since not all of them are supposed to be bound by the candidate. Thus, I selected 2 SL-miRNAs per pLenti6.3 construct for which the first one is supposed to be a greater interactant than the second according to the mass spectrometry quantitative results. The beads condition act as a specificity control as well as the GFP candidate that is not supposed to interact with any SL-miRNA. Since candidate list have been generated using the DNase assisted RNA chromatography based on magnetic beads, these heat-eluted agarose pulldowns can display different results. It's important to precise that the candidate list correspond to significant enriched proteins, thus it is normal to detect some interaction between the candidate protein and the SL-miRNA which didn't allow candidate significant enrichment. Knowing that the *in vitro* transcribed SL-miRNAs come from the same batch used and validated for the DNase assisted RNA chromatography, it can suggest that SL-miRNA also correctly fold. Thus, if the pLenti6.3 cloned candidate protein is still able to interact and preferentially bind with the "positive" SL-miRNA, it suggests that candidate molecular function is maintain at least for the RNA binding property.

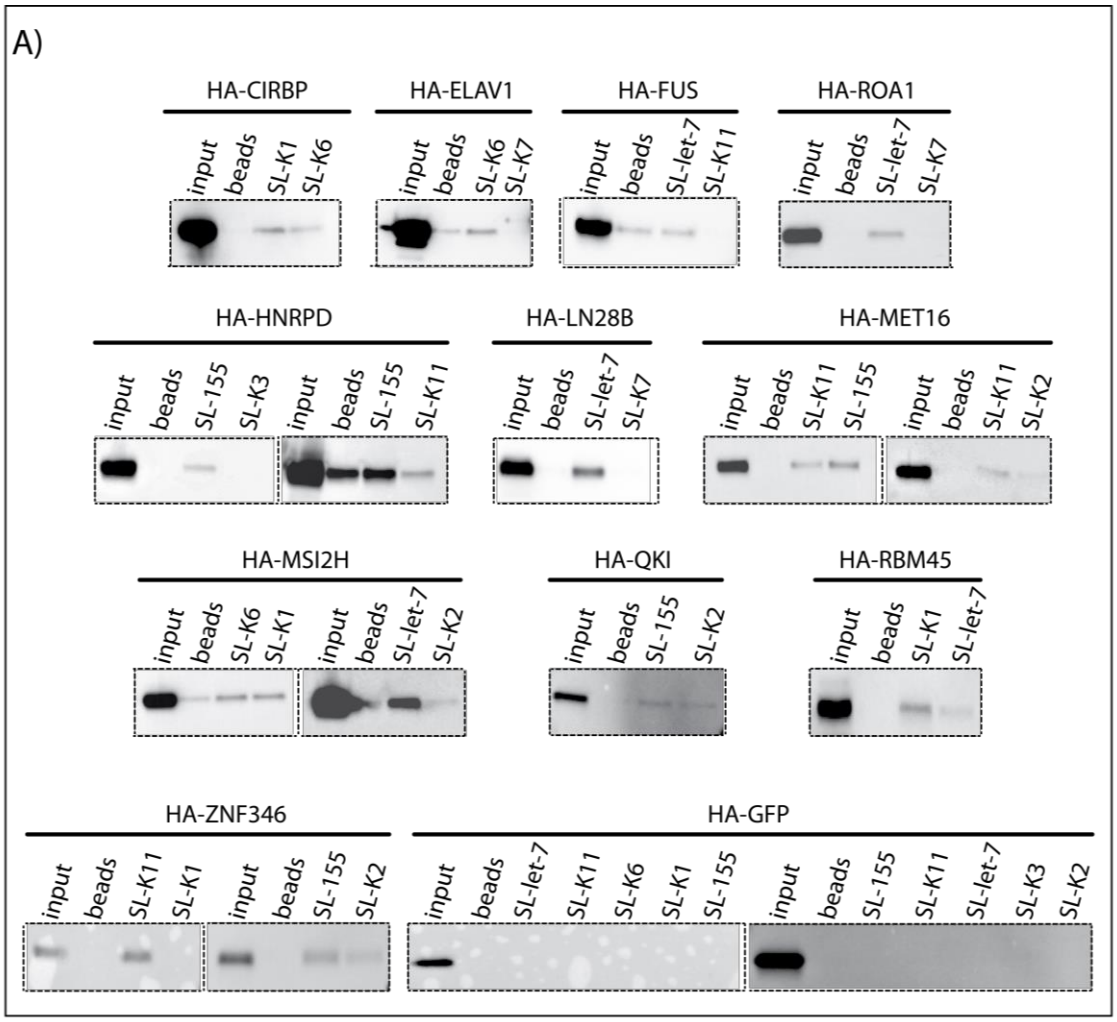


Figure 49. Agarose RNA-pulldown with HA-candidate. A) Agarose RNA pulldown followed by Flag/HA tag detection using WB. Plasmids expressing Flag/Ha-candidate are transfected in HEK293-Grip cells. Agarose RNA pulldown are performed with beads and 2 SL-miRNAs. 10% of the eluate is loaded in an SDS-PAGE. Finally, presence of the Flag/Ha-tag is revealed with specific HA-HRP antibodies.

On the Figure 49, -let7 bait allows me to validate this chromatography as it managed to perfectly fish HA-ROA1 and HA-LN28B proteins. HA-GFP, for its part, act as a good negative control as it is never pull downed by any SL-miRNA (Figure 49). CIRBP is slightly more enriched by SL-miRNA-K1 compare to -K6 which was expected. ELAV1 display similar results as for western blot on the endogenous protein. ELAV1 is able to interact with the beads and is a better binder for SL-miRNA-K6 than -K7. FUS is also retrieve by the beads. However, FUS is not identified by the SL-miRNA-K11. Here, it suggests that the steric hindrance of the SL-miRNA on the beads limit the non-specific binding of the protein on the beads. Thus, HA-FUS construction seems functional. HNRPD is supposed to be enriched by the SL-miRNA-155 and not -K11. Here, two different pulldowns have been performed. The second one display important signals for all the condition including the beads compare to the first one for which signal is only obtain for SL-miRNA-155. A washing issue can be at the origin of this result. MET16 display a similar result as the previous western blot. Since MET16 mass spectrometry

identification is not deep enough it is difficult to conclude on the binding specificity. However, the protein seems to keep it SL-miRNA binding ability. MSI2H displays a similar intensity between SL-miRNA-K6 and -K1 which was not expected. However, during a second pulldown, MSI2H shows a better enrichment for SL-miRNA-let7 compare to -K2. MSI2H seems to still possess its binding specificity. QKI shows a very weak enrichment whatever the SL-miRNA. QKI seems to still possess an RNA binding ability, but it is difficult to conclude on its specificity. RBM45 shows a better enrichment with SL-miRNA-K1 compare to -let7 which was expected. Finally, ZNF346 displays a good enrichment for SL-miRNA-K11 compare to -K1. Taken together, it seems that the cloned candidate still possessed their RNA binding feature. However, it is difficult to conclude on their SL-miRNA binding specificity notably for the construction HA-QKI and HA-MET16. Overall cloned candidates within pLenti6.3 seem functional and will be used for the following experiments.

III.B.2.b Impact on miRNA biogenesis

Any "exploratory" approach (co-IP, RNA chromatography…) produces lots of data and important number of potential candidates in which false positive candidates are present. To reduce the number of relevant candidates, screening assay are usually performed. In this project, we generated two screening approaches to assess either DROSHA cleavage (pmiRGLO) or mature miRNA activity (psi-CHECK-2) based on luciferase reporters. The most relevant reporter is the pmiRGLO reporter that directly assess the SL-miRNA maturation step I am interested in. Both screening assay are presented in the part III.B.1.b. Then, miRNA biogenesis variation assays upon individual candidate modulation in non-infected and infected cell lines are respectively described part III.B.2.b.1.5 and part III.B.2.b.1.6.

III.B.2.b.1.1 Luciferase screen assays

The purpose here is to effectively and rapidly screen all the candidates to restrict the number of candidates.

III.B.2.b.1.2 DROSHA cleavage reporter assay

To validate this pmiRGLO reporter assay, I started to use the pmiRGLO containing the SL-miRNA-let7 (pmiRGLO-let7) within its *firefly* 3' UTR. First, it is important to know if the pmiRGLO reporter is microprocessor-dependent. In the Figure 50.A, we see the effect of silencing DROSHA or its negative regulators ILF2 and ILF3 on pmiRGLO-let7 activity in HEK293-Grip cells (Ha and Kim, 2014). In this experiment, si-DROSHA increases the luciferase signal contrary to si-ILF2/ILF3 which decreases it. It seems therefore that the activity of pmiRGLO-let7 can indeed be modulated in a microprocessor-dependent manner. Since the main objective is to identify cofactors regulating the microprocessor activity, I then did similar experiment by either over-expressing or silencing known pri-let-7a-1 cofactors. In Figure 50.B, pmiRGLO-SL-let7 has been transfected in HEK293-Grip cells upon DROSHA, ROA1 and FUBP2 silencing or over-expression (Michlewski and Cáceres, 2010). siRNA experiments (si-DROSHA, si-ROA1, si-FUBP2) are normalized to the siRNA negative control (si-CTRL). Over-expression experiments (pDROSHA, pROA1, pFUBP2) are normalized to the empty

expression vector (pcDNA3.1). Here, I only show the results obtained in a four-day long experiment within HEK293-Grip cells after 2 successive siRNA treatments at 50 nM or one transfection of the candidate expressing plasmid (150 ng) followed by the pmiRGLO-SL-miRNA-let7 transfection (100ng). As a reminder, ROA1 and FUBP2 are known pri-let-7a-1 maturation cofactors, which act respectively as negative and positive regulators. Thus, over-expression of ROA1 and FUBP2 is supposed to respectively decrease and increase the DROSHA (Michlewski and Cáceres, 2010). Hence, the luciferase signal ratio (F-Luc/R-Luc) is supposed to increase for ROA1 and decrease for FUBP2 over-expression. For the siRNA treatment we expect to see the opposite effect. In Figure 50.B, we can see that DROSHA overexpression reduces the luciferase signal (F-Luc/R-Luc: 0.6) while DROSHA knock-down increases it (F-Luc/R-Luc: 1.2). For both cofactors studied, the luciferase signal is modulated as expected. However, the observed effect is very mild, even when we modulate the expression of DROSHA (20% for the knock-down). Indeed, we can only see a ~10% modulation for the co-factors tested even though all silencing treatments were effective when assessed by qRT-PCR.

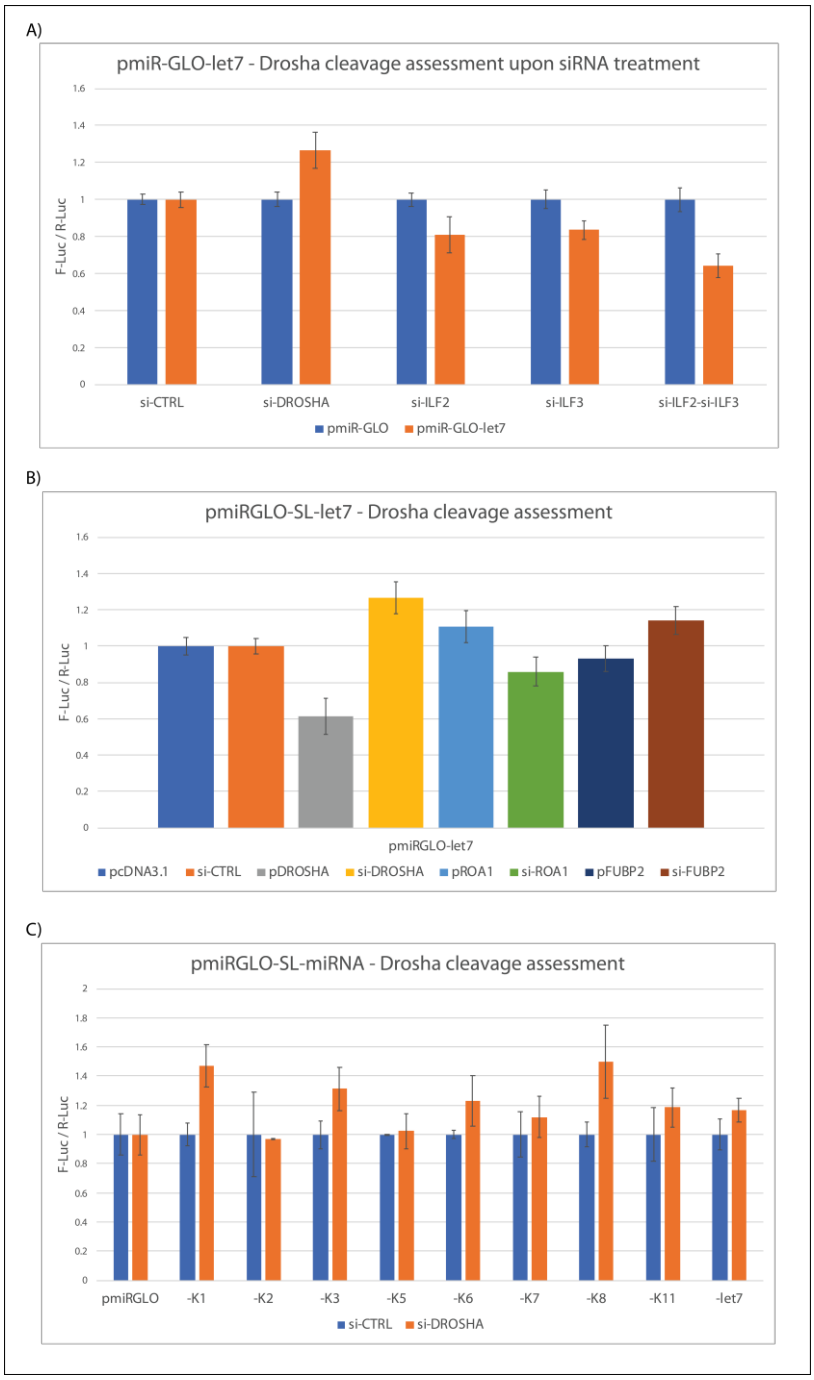


Figure 50. pmiRGLO luciferase assay. Firefly luciferase (F-Luc) signal is mediated by the DROSHA cleavage step. Renilla luciferase (R-Luc) act as a transfection normalizer. Luciferase signal display correspond to the ratio of F-Luc/R-Luc. Data have been subject to a double normalization first on the empty pmiRGLO control, then on the si-RNA control (siCTRL) for the silencing experiment or on the empty vector (pcDNA3.1) for the over-expression experiment. A) pmiRGLO-let7 is transfected within HEK293-GripTite cells upon candidate's microprocessor component silencing. B) pmiRGLO-let7 is transfected within HEK293-GripTite cells upon candidate's microprocessor component or microprocessor cofactors silencing and over-expression. C) pmiRGLO bearing the individual KSHV's SL-miRNA are individually transfected within HEK293-GripTite cells upon DROSHA silencing.

In parallel, we monitored the robustness of pmiRGLO sensors for KSHV SL-miRNAs. It should be noted that pmiRGLO -K2, -K5, -K7 and -K11 do not seem functional, therefore can not be interpreted (data not shown). In Figure 50C, we can see that the pmiRGLO sensors that appeared to be functional do respond to si-DROSHA treatment. However, there were some reproducibility issues with lots of variation between experiments. This figure only represents the better results obtained. Even after multiple setting up namely: siRNA quantity, experimental timing, cell lines and pmiRGLO quantity, it was not possible to obtained results reproducible

enough for statistical analysis. For these reasons, it therefore seems that the use of the pmiRGLO sensor is not well suited for the validation of our results obtained for KSHV pre-miRNAs.

In all experiments, luciferase signal is mildly modulated even when microprocessor components are targeted. I would have expected to see a greater luciferase modulation upon DROSHA silencing knowing that the siRNA treatment was efficient and allowed a decrease of more than 90 % of the targeted mRNA. It is important to keep in mind that the protein half-life could be longer than the duration of the experiment. Technically speaking it means that multiple round of siRNA treatments would have to be done, which implicate splitting cells several times before using the pmiRGLO reporter. Here, it is clear that only looking for mRNA decrease is not sufficient and that my experiment should last longer to eventually obtained better result and pmiRGLO sensitivity. Since, technical improvements were inconclusive and that DROSHA cleavage reporters were not sensible enough to decipher cofactors activity, I decided to test another reporter system.

III.B.2.b.1.3 psiCHECK-2: mature miRNA activity sensor

In Figure 51.A, I transfected the psiCHECK-2 with their corresponding pcDNA vector that expressed a single miRNA precursor. It represents a technical replicate in which luciferase signal is strongly reduced. Biological replicates (data not shown) for their part, display greater error bars that certainly come from the pcDNA transfection efficiency. Indeed, if the pcDNA transfection varies it will directly impact the luciferase signal because the psiCHECK-2 reporter is sensitive to the mature miRNA level which is directly correlated to the amount of pri-miRNA. The Figure 51.B represent also a technical replicate of the psiCHECK-2 reporters transfected with the pcDNA expressing the KSHV cluster which induce correct *firefly* luciferase decrease for most of the sensors. Plasmid allowing the expression of individual KSHV SL-miRNA have been controlled in (Contrant *et al.*, 2014).

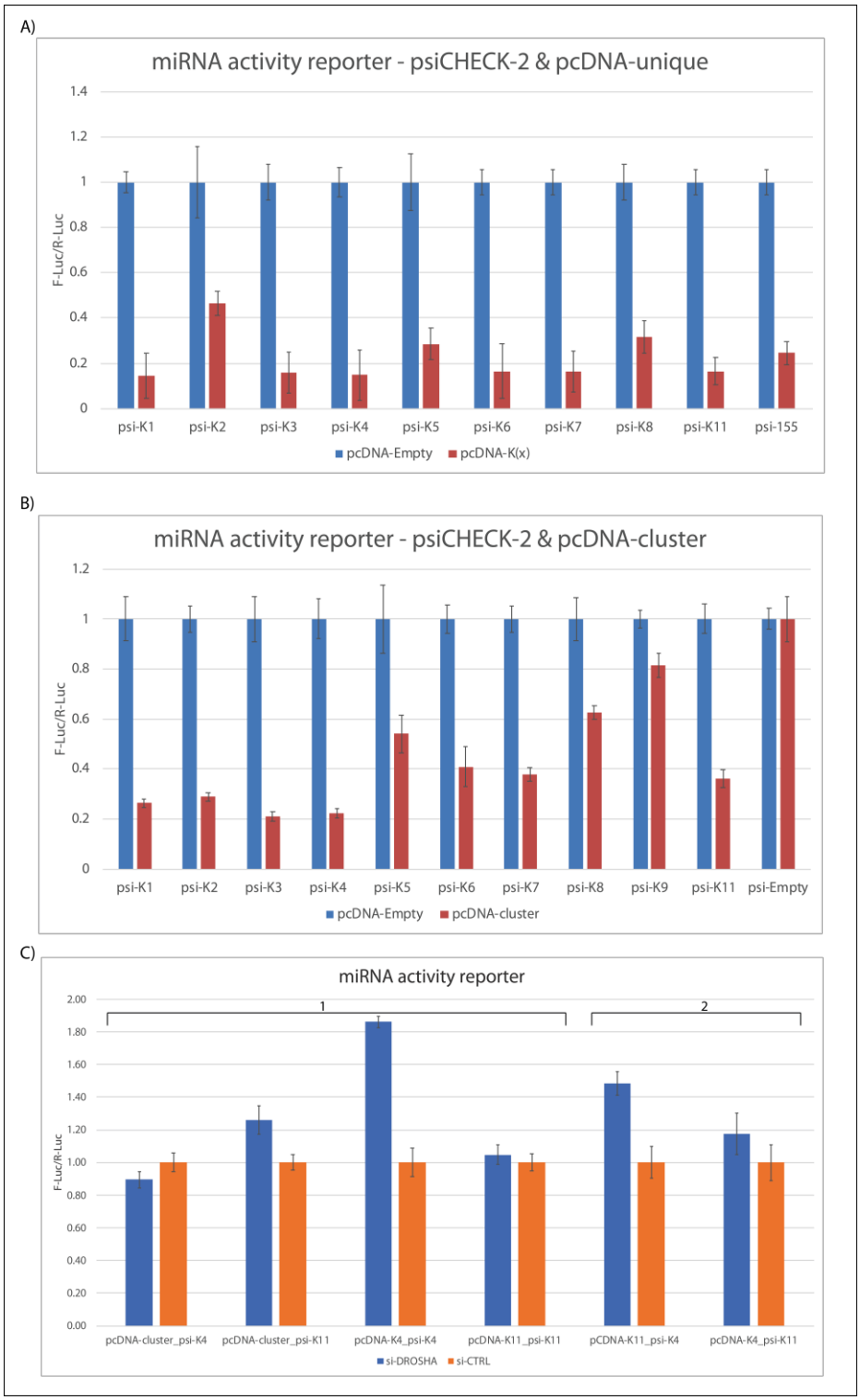


Figure 51. psiCHECK-2 luciferase assays. pcDNA-cluster correspond to the plasmid allowing the expressing of the complete KSHV miRNA cluster. pcDNA-K1 to pcDNA-K11 and pcDNA-155 correspond to the plasmid allowing the expressing of a unique miRNA precursor respectively kshv-miR-K1 to kshv-miR-K11 and hsa-miR-155. Firefly luciferase (F-Luc) signal is mediated by a given mature miRNA corresponding to the seed-match present within the F-Luc 3' UTR (from kshv-miR-K1 to kshv-miR-K11 and hsa-miR-155). Renilla luciferase (R-Luc) act as a transfection normalizer. F-Luc/R-Luc ratio is represented on this bar-chart. Dataset has been normalized by the pcDNA that express no miRNA precursor (pcDNA-Empty or si-CTRL condition. A) psiCHECK-2 signal with the presence of pcDNA expressing unique miRNA precursors. B) psiCHECK-2 signal with the presence of pcDNA expressing the KSHV miRNA cluster. C) psiCHECK-2 signal with the presence of either pcDNA unique or pcDNA cluser upon DROSHA and control silencing. The bars labelled 1 correspond to a matching psiCHECK-2 with miRNA expressing vector. The bars labelled 2 correspond to a non-matching psiCHECK-2 with miRNA expressing vector.

Here, biological replicates also display important variation (data not shown) which correspond to the pcDNA transfection efficiency. Even if the psiCHECK-2 correctly responds to the mature miRNA (Figure 51.A.b), biological replicate suggest that the pcDNA transfection is a considerable bias due to my inability to normalized its transfection. This bias is clearly highlighted in Figure 51.C. Here, I either used the psiCHECK-2 vector possessing the kshv-miR-K4 (psi-K4) or kshv-miR-K11 (psi-K11) binding site upon siRNA treatment. The bars labelled with 1 correspond to a condition in which psiCHECK-2 reporter match with either a single-miRNA expressing vector or the cluster expressing vector. The bars labelled with 2 correspond to a condition in which psiCHECK-2 reporter do not match with a single-miRNA expressing vector. Concerning the matching psiCHECK-2 and pcDNA, DROSHA silencing display important luciferase increase or not depending on the plasmid couples. As mentioned earlier, pcDNA transfection efficiency is the main bias and could be responsible of these important variations. With this reporter approach, this issue is clearly highlight by the pcDNA-cluster and the psi-K4 condition. Here, it does not make sense to detect less luciferase signal upon DROSHA siRNA treatment compare to the siRNA control. It indicates that the transfection and the expression of the pri-miRNA cluster and by extension the expression of the single miRNA can varies a lot. Here, it is likely that more miRNA cluster was express in the si-DROSHA condition compare to the si-CTRL one. Therefore, if I cannot perfectly assess the efficiency of miRNA expression plasmid transfection it will be difficult to correctly interpret the psiCHECK signal. In addition, when the pcDNA and the psiCHECK vector do not match, we can see variation of the luciferase signal suggesting a potential DROSHA and miRNA independent mechanism implicated (Figure 51.C). These variations can come from proteome modulation upon DROSHA alteration, or simply from technical troubles during the luciferase assessment. In conclusion, those vectors seem functional, but they have to be used within infected cell line or cell line stably expressing KSHV miRNA to eliminate this transfection bias and pcDNA transfection normalization.

III.B.2.b.1.4 miRNA biogenesis variation upon individual candidate modulation

Since the luciferase reporter are not functional, we have to modulate candidate expression and assess the different miRNA by northern blot and/or qRT-PCR. This approach is tedious first of all because northern blot and qRT-PCR take more time to perform than the luciferase assay and more importantly selected candidate could be a false positive.

III.B.2.b.1.5 Heterologous system

As mentioned earlier, HEK293-Grip cells are easy to transfect but do not express KSHV miRNAs. Consequently, miRNA cluster-expressing plasmid has to be transfected. The Figure 52 shows results obtained in HEK-Grip cell transfected both by a candidate overexpression pLenti6.3 plasmid and a plasmid expressing KSHV miRNA cluster. In Figure 52.A, we can see on the northern blot that each KSHV miRNA are correctly expressed and at similar levels in the different extracts. One representative northern blot analysis out of three replicates is shown. Each membrane has been blotted several times to reveals several miRNAs on each that explain the presence of two EtBr and tRNA controls. The good overexpression of the candidates is

controlled by western blot analysis (Figure 52.B). Each candidate seems correctly expressed, with the exception of HA-QKI protein, which is due to a loading problem. Since all candidate were expressed, I assumed that was the case for the HA-QKI protein as well, since I managed to express it in several other experiments (Figure 44). Without quantification, northern blot analysis does not reveal any striking increase or miRNA decrease upon overexpression of the tested candidates. Figure 52.C represents the quantification of the biological triplicate. As we can see, big error bars clearly suggest that there is an issue with the reproducibility of the assay except for the cellular Let-7a miRNA. In this case, overexpression of ROA1 significantly decreases its mature miRNA signal, as described in the literature. In addition, overexpression of HA-MSI2H seems to increase the level of mature Let-7a expression (Figure 52.D). Concerning KSHV miRNAs, the variability in expression suggests that the candidate or cluster-expressing vector transfections are not reproductible. Thus, it emphasizes the need to work within naturally KSHV infected cell line. Nonetheless, since total RNAs samples were available, I analyzed them by RT-qPCR as well (Figure 52.E).

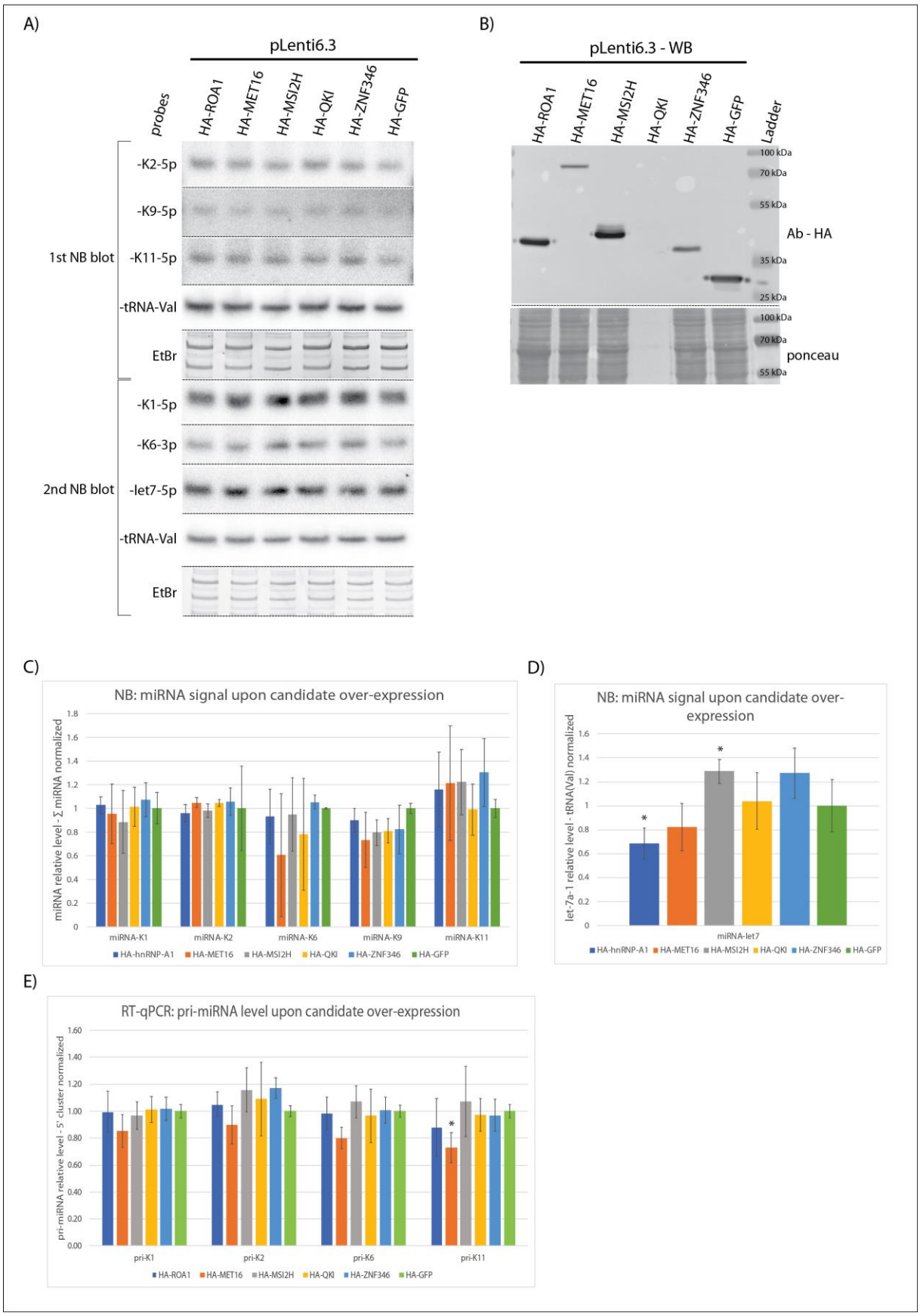


Figure 52. Candidate over-expression in heterologous system. Legend continues on the next page.

A) Northern blot assays from different total candidate-overexpressed RNA. Samples have been blotted two times to reveal several transcripts. Only one northern blot experiment is display for simplicity. B) Western blot assays validate the overexpression of the samples used in previous northern blot. C) Northern blot quantification chart displaying miRNA relative level of a biological triplicate. Each miRNA signal has been normalized by the sum of the different KSHV miRNA revealed. D) Northern blot quantification of let-7a-1 signal normalized by the tRNA(Val) signal. All blot signals have been quantified with the ImageQuant-TL software proposed by GE healthcare. E) pri-miRNA assessment of the biological triplicate by qRT-PCR. pri-miRNA signals have been normalized by the 5' cluster signal.

On the Figure 52.E, we can see no effect of the candidate's overexpression on pri-miRNA-K1 and -K2 cleavage. For the pri-miRNA-K6, HA-MET16 overexpression seems to increase its cleavage from the complete cluster transcript. This increase processing is significant for pri-miRNA-K11. This result needs to be interpreted with great caution as this heterologous system is far from relevant. However, it is interesting to note that both pri-miRNA-K6 and -K11 variated the same way upon HA-MET16 overexpression (miRNA: Figure 52C; pri-miRNA: Figure 52.E). Since those two SL-miRNAs are in close proximity on the cluster it could make sense that affecting cleavage of one pre-miRNA might affect the neighboring stem-loop. With this heterologous system, I did not try to use the shRNA-expressing vectors as their silencing ability is weak in transient transfection as previously shown.

Taken together these two approaches, northern blot and RT-qPCR therefore seem adapted to decipher pri-miRNA cleavage variation upon candidate modulation but not particularly well-suited for this heterologous system. Since I cloned the different shRNAs and HA-candidates within lentiviral vectors, I produced lentiviral particles to transduce and make stable KSHV-infected B lymphocytes, in order to study the implication of these candidates in KSHV miRNA biogenesis in a more natural context.

III.B.2.b.1.6 KSHV naturally infected lymphocyte system

Lentiviral particles have been generated with both second and third generation systems. Viral production was done in HEK293T cells and viral supernatant directly used to transduce $0.5.10^6$ BC-3 cells. pLKO.1-TRC and pLenti6.3 transduced cells are respectively cultured with 1 µg/mL puromycin and 16 µg/mL blasticidin. Since this candidate gave me some promising results in the heterologous system, I started by generating lentiviral particles for MET16 overexpression and silencing as well as both controls HA-GFP and sh-CTRL. In the next figures, stable BC-3 cells obtained with the second-generation lentiviral system are labeled (1), and the ones obtained with the third generation are labeled (2). In addition, I used two different version of the sh-MET16 referred to as (a) and (b).

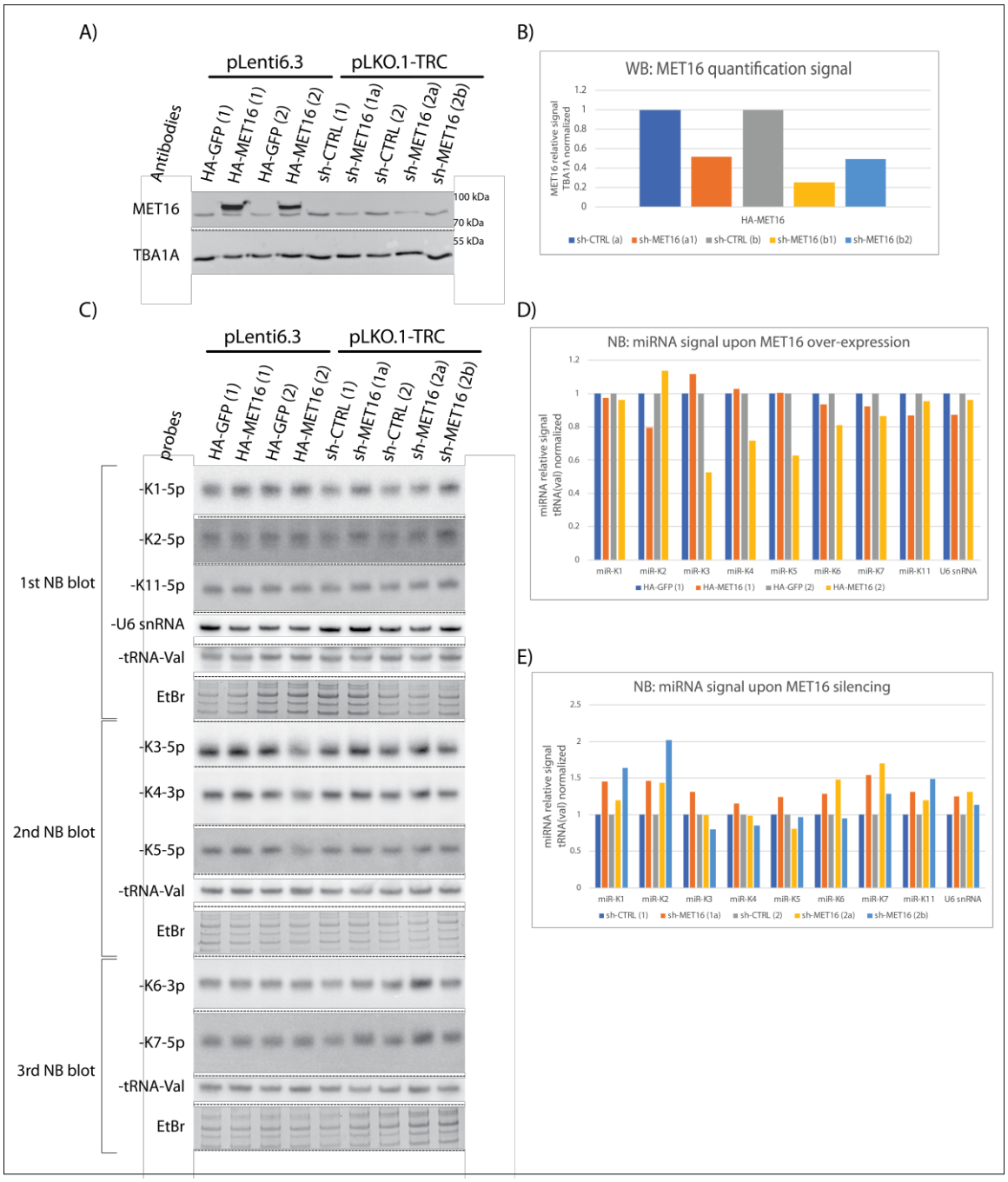


Figure 53. Candidate over-expression and silencing within infected BC-3 cell line. A) Western blot reveals with HRP-antibody directed against endogenous MET16 protein. Protein samples are retrieved from transduced BC-3 infected lymphocytes using RIPA buffer. Tubulin (TBA1A) protein is used to control the sample loading. B) WB quantification using ImageQuant-TL software. TBA1A signal is used as a normalizer. C) Northern blot assays from different transduced BC-3 cells. Samples have been blotted three times to reveal multiple transcripts. D) Northern blot quantification of the overexpressed HA-MET16 using ImageQuant-TL software. miRNA signals have been normalized by the tRNA-Val. Bar chart represent the average signal obtained with the different biological replicates (a) and (b) E) Northern blot quantification of the sh-MET16 transduced BC-3 cell using ImageQuant-TL software. miRNA signals have been normalized by the tRNA-Val. Bar chart represent the average signal obtained with the different biological replicates (a) and (b)

We checked MET16 overexpression and silencing in polyclonal cells by western blot analysis. In Figure 53.A, we can see that the HA-tagged MET16 protein is well overexpressed in cells transduced with the corresponding lentivirus. Cells transduced with the pLKO lentiviruses targeting MET16 show a decrease in MET16 expression ranging from 50 to 75% which reflects the polyclonal characteristic of the cells (Figure 53.B). We then analyzed the impact of MET16 overexpression or silencing on KSHV miRNA by northern blot analysis (Figure 53.C). Figure 53.D and Figure 53.E represent the quantification of the average signal obtained with each condition. Concerning the over-expression, MET16 displays no effect for miR-K1, -K7 and -K11 or contradictory results. Even though the pLenti6.3 was the same to produce lentiviruses with the second or third generation, miR-K2, -K3, -K4 seem less or more abundant depending on technology generation used to produce the viral particles. Concerning the silencing conditions (Figure 53.E), the different versions of the hairpin show an increase of mature miRNA except for miR-K3, -K4, -K5. For those miRNAs either the same hairpin or different hairpin display contradictory even though they target the same mRNA. It seems that overall signal upon MET16 overexpression decreases contrary to the silencing condition in which overall miRNA signals increase, suggesting a potential negative impact of MET16 on the KSHV miRNA biogenesis. However, the contradictory results obtain clearly reflect the polyclonal feature of our B lymphocyte transduced cell lines.

RT-qPCR analysis was also performed on the same samples. Figure 54 displays pri-miRNA amplification within MET16-overexpressing BC3 cells. As seen by northern blot, MET16 overexpression does not modulate the levels of assessed pri-miRNA or of KSHV mRNAs (RTA and vCyclin). MET16 silencing for its part, shows contradictory results. For example, pri-miRNA-K6 and -K11 are less processed with the sh-MET16 (2a) but more cleaved with the sh-MET16 (2b). The only transcript that seems to display a consistent result is the KSHV RTA mRNA which always decrease upon mET16 silencing. As the northern blot suggested, monoclonal cell line have to be generated from our transduced B lymphocytes. Nonetheless, it seems that the qRT-PCR result obtains with condition sh-MET16 (2b) correlate with northern blot. Upon this condition, the pri-miRNA-K11 is more process and the miR-K11 increases.

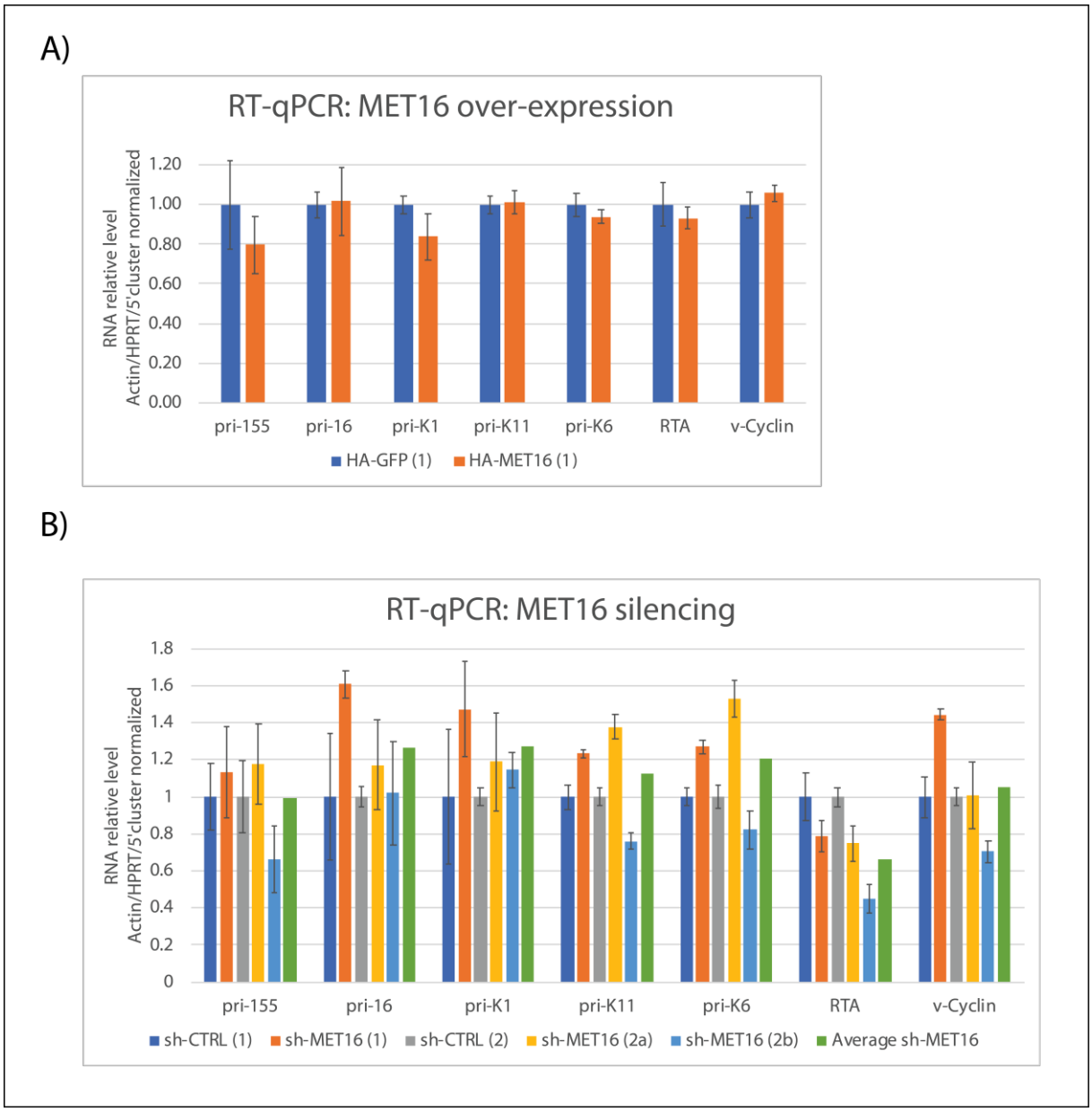


Figure 54. KSHV pri-miRNA assessment by QRT-PCR of transduced BC-3. A) Average pri-miRNA cleavage assessment of HA-MET16 transduced BC-3 cells. B) Average pri-miRNA cleavage assessment of sh-MET16 transduced BC-3 cells.

Taken together these results suggest that it will be possible to study miRNA precursor maturation within transduced BC-3, which was the main purpose. However, the MET16 protein does not display a specific impact on the SL-miRNA-K6 and -K11 as the previous heterologous experiment could have suggested. Nonetheless, MET16 could still have important function for the KSHV life cycle as the RTA mRNA regulation seems MET16 dependent and as RNA post-transcriptional modification is now consider a major regulation system for viral cycle (Tan *et al.*, 2017; Martin *et al.*, 2018; Dang *et al.*, 2019; Ringeard *et al.*, 2019).

IV. DISCUSSION

All fundamental cellular functions require intensive collaboration between RNAs and RBPs. RNAs and RBPs interactions allow the cell to extend or diminish both RNA and RBPs function. Thus, RNA function or processing can be completly understood only when all the possible RNPs are characterized. It is therefore important to decipher RNA-partners globally, but also in specific conditions. Recent methods have allowed the characterization of previously unknown RNA-binding function in a global and non-biased manner that strongly increases the RBPs repertoire. RNA interactome capture (Castello *et al.*, 2016; Perez-Perri *et al.*, 2018) and identification of RNA-dependent protein by density gradient (Caudron-Herger *et al.*, 2019) are the main global approaches to discover unknown RNA-binding function of proteins. Those methods clearly emphasize the fact that lots of proteins lacking known RNA-binding domain are still able to interact with coding or non-coding RNA molecules. The discovery of RNA binding ability of proteins with no previously known relation with RNA biology, such as metabolic enzymes, clearly suggests that there are still underexplored fields in biology (Castello *et al.*, 2016; Perez-Perri *et al.*, 2018; Caudron-Herger *et al.*, 2019). In addition, RNA-nucleotides post-transcriptional modifications greatly enhance the possibility of RNPs formation (Roundtree *et al.*, 2017; Frye *et al.*, 2018; Ontiveros *et al.*, 2019) and also represent a growing RNA biology research field. In my project, I only focused on proteins, but it is also important to precise that RNA/RNA or RNA/DNA interaction as well as metabolites are also able to extend or repress RNA function. This RNA/RNA regulation is particularly studied in prokaryote cells which regulate a multitude of fitness pathways directly via small ncRNA interacting with coding RNAs (Bronesky *et al.*, 2019). This can also occur in eukaryotic cells, where a given RNA can interact with a transcript to modulate its folding resulting in a different transcript fate such as miR-122 regulating miR-21 maturation (D. Wang *et al.*, 2017). Metabolites can also be bound by RNA molecules such as oleic acid which is able to interact with pri-miR-7 and regulate its processing (Kumar *et al.*, 2017). It suggests that metabolites can be sensed by RNAs which can act as aptamers or riboswitches-like molecules.

Concerning miRNAs, it is already known that they are generated and controlled by sequential action of a large variety of proteins allowing their cleavage, export, and RISC loading. As mature miRNAs act as post-transcriptional mRNA regulators and since dysregulation of miRNA biogenesis is linked to cellular pathology, it is clear that regulation of their abundance is mandatory for all the cellular processes from cell division to signalling pathways output. Thus, miRNA regulation can be considered as a fundamental layer of gene expression and regulation network. As mentioned in the introduction, miRNA biogenesis comes in different flavours but display common features (part I.A.2). The most important feature corresponds to the folding of the miRNA precursor as well as flanking sequence motifs (Auyeung *et al.*, 2013). In my project, I focused on the canonical miRNA biogenesis pathway used by KSHV to express its own miRNA from a clustered miRNA pol-II transcript. As mentioned, structural and sequence features is important but not sufficient to decipher miRNA maturation regulation (Michlewski *et al.*, 2008; Contrant *et al.*, 2014). Multiple studies clearly demonstrate the implication of additional implicated protein cofactors which allow the cell to quickly respond to any signal to repress of enhance specific miRNA maturation without transcriptional regulation (Michlewski *et al.*, 2008).

To briefly summarize, miRNA biogenesis is controlled and regulated by multiple processes. The proteome is the main actor of this regulation as it establishes RBP balances depending on the protein quantity and affinity for a given substrate and also depending on all the competitive RNAs that can titter a specific RBP. Obviously, all these parameters will modulate RNPs formation and function. In this context, miRNA biogenesis is first regulated at the transcriptional step. Next, the different sequence and structural features of the precursor stem loop directs its processing together with specific RBPs and nucleotide modifications. Furthermore, the processing order of clustered miRNA manages the fate of the downstream and/or upstream precursors. Finally, all of these aspects are unbelievably dynamic providing its versatility to the miRNA biogenesis.

Discovering new miRNA maturation cofactors of the KSHV was the main objective of my project. Here, I managed to generate lists of candidate proteins potentially implicated in specific miRNA maturation for all KSHV pre-miRNA as well as two cellular pre-miRNA within different cell types. The experimental approaches, tools and results obtained during this project are discussed below.

IV.A. RNA chromatography

RNA chromatography is the standard go-to method to discover protein partners for a given RNA. Every exploratory approach requires strict and challenging procedures (homogeneous bait and cell extract, identical batch of tools used, same operator, timing procedure etc…) in order to generate correct replicates and obtain the most relevant dataset. An important number of biases are intrinsically linked to this kind of approaches even before mentioning the mass spectrometry and data mining bias. It is therefore mandatory to always couple affinity chromatography methods with lots of control steps notably for the bait construction, cell extract preparation and elution (Hage and Matsuda, 2015). During my project, I tried to produce clean dataset carefully explored and experimentally criticized as those lists will be used in follow-up projects. In my project, I decided to performed chromatography in solution. Even though this approach is biased in favour of irrelevant RNPs, it greatly facilitates multiple bait experimental conditions and comparisons.

Our results indicate that agarose-based RNA chromatography possesses more drawbacks than the magnetic-based approach. As mentioned in part III.A.1.d, agarose beads are fragile and difficult to handle due to all the centrifugation steps required. Moreover, mass spectrometry identification following the pulldown was problematic. Descriptive analysis clearly suggests a weak chromatography reproducibility, sensitivity and specificity. To overcome agarose weak sensibility and specificity, I set up an optimized RNA chromatography based on magnetic beads and a specific elution procedure, which is described in part III.A.3. To correctly assess the advantages and drawbacks of both elution methods, it would have been interesting to compare them using biological triplicates. Here, the new ‘Heat vs DNase’ comparison only provides indication due to the sample duplication (DNase-let7_2 = DNase-let7_3). This duplication greatly limits the statistical approach and the corresponding interpretations. However, this comparison of a technical replicate seems to highlight the advantages of the DNase elution approach, thus has been used with the KSHV SL-miRNA baits. This was my main objective since my original purpose was to perform sensitive and specific RNA chromatography for all

the KSHV SL-miRNAs. Magnetic-based RNA chromatography allows me to generate cleaner eluates in term of sensibility and specificity. Enriched dataset obtained with this approach are far more relevant compared to the agarose one and are now considered in my laboratory as a compendium of potential cofactors for all the KSHV miRNA precursors. Unfortunately, all the chromatographies I performed with the KSHV SL-miRNA are far from perfect. First of all, I cannot compare SL-miRNAs from different experiments (Day1: part III.A.4.b.1.3, Day2: part III.A.4.b.1.4, Day3: part III.A.4.b.1.5, DG75: part III.A.4.b.1.6) due to proteome variation, raw mass spectrometry identification and MaxQuant software matching issues. To overcome this problem, I should have used all the baits within a unique cell extract, but this was experimentally challenging. Here, northern blot analysis of the bait formation and elution (Figure 23) clearly displays technical issues notably for the -K5 bait, which appears to have not been eluted. In addition, as opposed to the DNase versus heat assay, my following chromatographies did not manage to retrieve as many significantly enriched proteins. This clearly highlights the importance of the cell extract and bait quality to perform correct affinity chromatography. Surprisingly, specific cofactors involved in pri-miRNA-let-7a-1 maturation (LN28B and ROA1) are not the most enriched proteins. In solution RNPs formation could be responsible of this weak enrichment. It clearly suggests that relevant candidates are not systematically part of the top candidate in term of enrichment ratio and p-values. Nonetheless, the results obtained upon KSHV SL-miRNA chromatography allowed the identification of a multitude of RBPs implicated in various RNA pathways such as translation initiation, helicases, ribosomal component, tRNA maturation, pol-II component, DNA repair, exon-junction complexes, RNA sensors, cap and poly-adenylation complexes, aminoacyl tRNA synthetase, small GTPase proteins and methyl-transferases. I can highlight in particular the identification of mitochondrial RBPs that could be implicated in miRNA biogenesis. Indeed, protein shuttling from one subcellular compartment to another is commonly observed upon viral infection (Cook and Cristea, 2019; Tessier *et al.*, 2019) or specific cell signalling (Bauer *et al.*, 2015; Jeffery, 2015; An *et al.*, 2019). I can also emphasize the absence of viral protein identification during all of the chromatography as well as of DROSHA and DGCR8.

It is important to precise that despite important improvement in mass spectrometry analysis, the downstream analytical pipelines are in constant evolution. I can highlight several analytical steps that can greatly impact the differential analysis. Raw dataset matrix is always composed of “observed” values which are directly measured by the spectrometer, “recovered” values which can come from mass spectrometry run alignment (by-matching) or computed values (fixed or statistically generated). Thus, observed and recovered values should not be trusted equally. This is why filtration and imputation steps are critical before proceeding with any differential analysis. If there are only two conditions to compare, the filtration step is easier. In this condition, all proteins containing important amount of recovered values are generally discarded from the dataset matrix. However, when there are more than two conditions, this kind of filtering can lead to important loss of relevant information when multiple comparison occurs afterward. For the agarose dataset, I only possessed the tools to simply discard contaminant and reversed identified protein, which is insufficient to avoid analysing irrelevant proteins. However, for the magnetic dataset matrix, ProStaR software allowed me to apply different filters. As I wanted to compare multiple eluates, I decided to keep all the proteins that displayed at least two observed values within a triplicate condition. This filtration means that I could

compare conditions displaying recovered values which is dangerous. However, the ProStaR software possesses a function to apply an additional filtration step during the differential analysis to get rid of very low p-values coming from recovered values. This post-filtration step occurring during the pairwise comparison can be applied on the whole matrix, for every condition, or for at least one condition. Thus, each pairwise comparison should be performed and generated using several types of filtration, imputation and post-filtration procedure. Then, the remaining proteins are associated with a fold change enrichment as well as a p-value which correspond to the reliability of the differential analysis. FDR is systematically linked to the p-value and allows the estimation of proteins falsely considered as differentially enriched (false positive). The lower the FDR is, the better the enriched dataset is. However, as with all statistical approaches, FDR can be interpreted only if the correct statistical assumptions are fulfilled. One of the most important assumption is that the p-values of the non-differentially enriched proteins should be distributed uniformly whereas the differentially enriched proteins should not. Unfortunately, I do not possess the mathematical knowledge to estimate all these parameters. Hence, it is imperative to keep in mind that the differentially enriched datasets I generated are as exhaustive as possible but might display underestimated false positive rates. Furthermore, a normal distribution and homoscedastic datasets are mandatory. Those two conditions are very difficult to assess with triplicate experiments. Knowing that those differential tests have been developed for epidemiology dataset that possess greater number of individuals facilitating the prerequisite validation, it could make sense to still being interested in protein strongly enriched with a p-values above 5 % within our triplicate affinity chromatography experiments.

Following the generation of the significantly enriched datasets, protein features can be explored. GO-term analysis is a good way to gain insight into the nature of enriched proteins. But protein features displayed in every database are, by definition, incomplete and in constant evolution. Indeed, if no one in the scientific community do not looks for a specific feature, it will remain hidden. The same is true for subcellular localization, we can assume that a given protein localizes to a specific cellular compartment, but it can naturally re-localize upon stress. Some features may however be defined with an unbiased approach. For example, RNA binding abilities were assessed for all proteins of a given cell line using a proteome wide approach coupled with density gradient followed by mass spectrometry analysis (Caudron-Herger *et al.*, 2019). I cross-referenced this database with my data and the 15 proteins commonly enriched by the SL-miRNA-let7 chromatography upon heat and the DNase elution all possess RNA binding properties.

Altogether, all the RNA chromatographies performed produced candidates list that seem correct and deep enough to work with. As mentioned earlier, all the significantly enriched candidates should be considered as potential SL-miRNA cofactors whatever their enrichment or p-values. To restrain the number of candidates, an input proteome for each of the cell extract used could have been generated. Indeed, a protein strongly present within the cell extract that is strongly enriched upon affinity chromatography could be considered as a less relevant binder compare to a weakly present protein in the input that is strongly enriched after pulldown, reflecting a better binder affinity. In addition, from the moment a protein binds an RNA, it induces extended or repressed RNA function generally depicted by a fold rearrangement

favouring or not new binder interaction. This is why candidate lists have to be parsed without preconceived notions ideally using screening approaches.

IV.B. Luciferase screen assays

Here comes the fundamental importance of a powerful screening experiment to determine the functional implication of candidate proteins and restrict their number for the following experiments. First, candidate expression should be modulated. Secondly, a measurable output of the process studied should be easily assessed, in a reproducible and sensitive manner. Loss of function screen is generally the standard method used (Root *et al.*, 2006; Varble *et al.*, 2013). Indeed, it seems more relevant and efficient to repress one protein at a time instead of increasing its quantity to being able to decipher its role in the studied process. Whichever the candidate modulation approach used the most important part of a screening assay is the output that you can measure. To perform a relevant screening assay, one can either use global or specific approaches. For an RNA processing event, the global approach could correspond to transcriptomic analysis (micro-array or deep sequencing) (Lowe *et al.*, 2017). Global approaches are by nature more relevant, but they are also more expensive and work-intensive. Here, I chose a restricted approach to measure candidates implication on a specific RNA processing based on a luciferase output signal. All screens based on "restricted" approaches have to be sensitive enough to study the processing event. In this project, I wanted to decipher candidate implication on DROSHA SL-miRNA cleavage. Thus, output signal must be reproducible and modulated with enough efficiency by SL-miRNA specific cofactor. During this project, I worked with two luciferase reporters that either measure SL-miRNA DROSHA cleavage or mature miRNA translation repression function namely pmiRGLO (see part III.B.1.b.1.1) and psiCHECK-2 luciferase vectors (see part III.B.1.b.1.2). It is important to precise that the proteomic data have been generated from BC-3 cell extracts. Here, I used reporter vectors in HEK293 cells. Therefore, one can expect differences in term of RBPs presence and quantity between cell lines. It is possible that silencing a candidate within HEK293 cells will not affect a luciferase sensor while it would have in lymphocyte cells. As a consequence, candidate proteins should be both overexpressed and silenced to try to overcome this issue.

pmiRGLO bearing a given SL-miRNA within its 3'UTR firefly luciferase is theoretically one of the most relevant screening tools that I used during my project. The presence of a SL-miRNA within the luciferase mRNA make it sensible to any DROSHA cleavage event and usable within non infected cells or non-expressing KSHV miRNAs cells. *Firefly* output sensibility was first assessed with DROSHA, ILF2 and ILF3 RNA silencing. Upon these conditions, if the *firefly* luciferase was mainly microprocessor dependent, I should have observed strong luciferase modulation, and increase and decrease respectively for DROSHA silencing and ILF2 ILF3 silencing. Here, its output luciferase follows the correct direction depending on the RNAi treatment. However, luciferase is only modestly changing suggesting that either the silencing treatment was ineffective, or the *firefly* luciferase signal is not sensible enough. Silencing treatment has been controlled by qRT-PCR and shows a correct diminution of the targeted mRNA. Relative protein level, though, was not assessed. Since DROSHA is a distributive enzyme, it could make sense that remaining protein level is still sufficient to control

the luciferase output. In addition, ILF2 and ILF3, two negative DROSHA cofactors, display similar luciferase modulation efficiency upon RNAi as DROSHA silencing. ILF2 and ILF3 do not possess a catalytic activity, thus since ILF2, ILF3 and DROSHA RNAi display similar luciferase output modulation, it suggests a lack of sensibility of the pmiRGLO reporter. This lack of sensibility can come from several factor. Since SL-miRNA within the *firefly* mRNA is out of its cluster context, it can misfold and therefore decrease the DROSHA dependent luciferase output. Moreover, misfold SL-miRNA can be considered as additional sequence that increases protein binding onto the firefly 3' UTR which will modulate mRNA localisation, export and expression in a microprocessor independent manner. I tried to overcome this issue and increase DROSHA dependent luciferase output by inserting the SL-miRNA within the firefly 5' UTR (in another sequence context) or by inserting multiple SL-miRNA within the 3' UTR. Unfortunately, those constructions displayed similar luciferase output modulation. At least in my hands, pmiRGLO seems suited to studied core-microprocessor complex but not specific miRNA binder. However, it could eventually respond in different cases. First, the experiment could be performed within cofactor knock-out cell line. Another solution could be to do both knock-down and over-expression to increase the firefly luciferase modulation. For example, a suspected positive cofactor could be silenced within DROSHA knock-down or over-expressed cell line. On the other hand, a suspected negative cofactor could be over-expressed within DROSHA knock-down or over-expressed cell line. In those cases, cofactors modulation effect would be enhanced by DROSHA over-expression or knock down. Nevertheless, it could also create other biases linked with the reproducibility of the DROSHA silencing or overexpression.

At first, I did not want to use the psiCHECK-2 reporter since it means transfecting yet another additional vector having to express KSHV cluster or single SL-miRNAs. With this reporter system, we could measure important luciferase modulation suggesting a good sensibility. Unfortunately, it turned out that normalizing the expression of the cluster or the single SL-miRNA between experimental conditions was complicated. It thus proved difficult to discriminate the impact of candidate protein modulation on SL-miRNA processing from variations in expression of the miRNA precursors. Therefore, the psiCHECK-2 seems appropriate to study naturally expressed miRNA. Here, it will be mandatory to use a cell line that stably expresses KSHV miRNAs in order to make sure that within each experimental condition the studied miRNA is expressed at the same level. Another solution could be to generate new miRNA-expressing vectors that expressed both SL-miRNA and a fluorescent protein under the same promoter in order to be able to quickly estimate the relative expression of the SL-miRNA between experiments.

In conclusion, both luciferase-based approaches are functional but have to be used in more controlled experimental procedures. Since I could not rely on a potent reporter assay to pre-screen my candidates list, I had to cherry-pick a manageable number of candidates based only on available literature and potential biological relevance.

IV.C. RNPs, candidates and expressing vectors

Here, I wanted to discuss the selection of candidates selection as well as the tools used to modulate their expression.

IV.C.1 Selected candidates

All twelve selected candidates display characterized RNA binding activity as well as nuclear localization (Table 8). In addition, all the selected candidates have been characterized to interact with SL-miRNA (Treiber *et al.*, 2017). I tried to select cofactors for the first 5' KSHV miRNA precursors (SL-miRNA-K1, -K2) as well as SL-miRNA positioned in the middle and the end (SL-miRNA-K6, and -K9). Other candidates have been identified by SL-miRNA-let7 and -155. In my selected candidate, I chose two known let-7a-1 cofactors, ROA1 and LN28B, to continue to use let-7a-1 as a positive control. In addition, those two cofactors are also enriched by the SL-miRNA-K1 and -K2, as well as CIRBP and RBM45. As mentioned in the introduction (see I.A.4.c), processing of clustered miRNA adds another layer of DROSHA processing regulation. Since miRNA processing is a co-transcriptional process, the first SL-miRNA exiting the polymerase II could be or not rapidly cleavage inducing conformational changes for the downstream SL-miRNA. This is why I selected CIRBP and RBM45, in addition to ROA1 and LN28B, since they have been described to respectively impact miRNA processing (Downie Ruiz Velasco *et al.*, 2019) or interact with SL-miRNA (Treiber *et al.*, 2017). Two proteins, HNRPF and MSI2H, have been selected for the SL-miRNA-K6. MSI2H was describe as a specific cofactor for the post transcriptional regulation of miR-7 (Choudhury *et al.*, 2013) together with ELAV1 (HUR). HNRPD and HNRPF, are implicated in various RNA process from splicing to RNA turnover and can potentially act on miRNA biogenesis as other members of their hnRNPs family (ROA1, and ROA2). QKI and ZNF346, selected via SL-miRNA-K11 and -155, were described to respectively control pri-miR-7 processing (Wang *et al.*, 2013) and interact with SL-miRNA loop (Treiber *et al.*, 2017). Finally, since nucleotide modification can have important impact on RNA processing, I selected MET16, an RNA N6-methyltransferase. Furthermore, m^6A has been identified as a pri-miRNA mark for processing and DGCR8 recognition (Alarcón *et al.*, 2015).

In retrospect, it might have been easier to restrict my selection to proteins identified only for one SL-miRNA (e.g. SL-miRNA-K1) with the hope that among those at least one would modulate its biogenesis. Anyhow, among selected candidates, I chose to select and start to work with the methyl transferase MET16. It has to be noted here that another component of methyl transferase complex, namely the VIR (virilizer homolog) protein (Yue *et al.*, 2018) part of WMM (WTAP, METTL3 and METTL14) complex (Schwartz *et al.*, 2014), has been identified as significantly enriched by the baits SL-miRNA-K6, -K7, -K8, -let7. RNA nucleotide modifications correspond to another layer of post transcriptional regulation since studies reveals their importance in T cell homeostasis (J. Yang *et al.*, 2019), inflammatory response (Wang *et al.*, 2019), antiviral immunity (Dang *et al.*, 2019) and antitumor response (Sun *et al.*, 2019). Nucleotide modifications can have important impact on RNA processing. In miRNA biogenesis, the importance of adenosine deamination by ADAR is one of the most studied process (Chawla and Sokol, 2014). Nucleotide modification can modulate RNA folding consequently repressing or favouring protein binding in a sequence- or structure-specific manner. Recently, m^6A has emerged as an abundant and dynamically regulated modification throughout coding and non-coding RNA (Balacco and Soller; Alarcón *et al.*, 2015; Meyer and Jaffrey, 2017; Frye *et al.*, 2018). These modifications are controlled by multiple complexes

composed of writer, reader and eraser proteins. METTL3-METTL14 complex was the first m^6A writer machinery described (Liu *et al.*, 2014). Interestingly, m^6A modification seems highly selective suggesting its important function for RNA fate as well as viral infection (Meyer *et al.*, 2012; Schwartz *et al.*, 2014; Martin *et al.*, 2018; Tan and Gao, 2018; Ringeard *et al.*, 2019). Recent studies demonstrated that many viruses harbour m^6A modification on their genomes for RNA viruses or their transcripts for DNA viruses, and these can display either pro- or antiviral effect (Tan and Gao, 2018). m^6A modifications have been detected on KSHV transcripts during both latent and lytic phases. It appears that the cellular epitranscriptome is also changing upon KSHV infection and favour oncogenic processes and epithelial-mesenchymal transition (Tan *et al.*, 2017). However, the impact of the m^6A on specific KSHV transcript has not been evaluated. It is possible that m^6A modification depending on its position, targeted transcript and involved machinery could have different functions.

Within my enriched dataset, both methyl transferases (MET16 and VIR) identification does not seem reproducible, at least for SL-miRNA-let7. In other words, SL-miRNA-let7 managed to identify MET16 during the third pulldown but not the other ones. This can be due to their distributive activity, proteome differences or it could correspond to a false positive. In the list of candidates, we did not identify m^6A erasers, which makes sense since my SL-miRNA baits are not m^6A modified. Recently, the list of m^6A readers has been expanded with RBPs containing common RNA binding domains such as K homology, RNA recognition motifs, and arginine/glycine-rich domain (Cléry, 2011). Among those, we could identify in our mass spectrometry data: ROA2 (HNRNPA2B1), ELAV1 (HUR), SRSF3, SRSF7 and hnRNPs that are also known to be implicated in miRNA biogenesis (Meyer and Jaffrey, 2017; Berlivet *et al.*, 2019; Chen *et al.*, 2019). Taken together, it could make sense for the cell to methylate specific SL-miRNAs to control their biogenesis.

IV.C.2 Candidates silencing and overexpression

To being able to silence my candidates in a stable manner within BC-3 cells, I cloned shRNAs directed against the candidates within the pLKO.1-TRC lentiviral vector. In my project, I cloned only one shRNA per candidate except for MET16 for which I cloned two different shRNAs. I selected the first output provided by the algorithm of the Genetic Perturbation Platform (Broad Institute), which always targets the coding sequence. It would have been interesting to select shRNA targeting 3' UTR to be able to rescue the stable knock-down.

Concerning the plasmids for candidate proteins overexpression, pLenti6.3 vectors can be considered as a versatile tool for the laboratory since they can be used for overexpression in any given cell type. In addition, the presence of the N-ter HA tag can be used to perform cross-linking immunoprecipitation or recombinant protein production. Cross-linking immunoprecipitation approach can be seen as an RNA-pulldown inverted method (Ule *et al.*, 2018). Thus, RNA interactome of each candidate could be explored to gain insight into their global functions. However, the presence of a tag could destabilize candidate function and interactome. Several approaches can confirm the properties conservation. The candidate's interacting partners, molecular functions, and sub-cellular localization should remain identical as the wildtype protein. Here, I only confirmed the RNA partner interaction via new RNA

chromatography. Identical molecular function can be controlled by a different way. For example, the ability of ROA1 to repress pri-let-7a-1 processing was validated in my hand (see part III.B.2.b.1.5). Finally, subcellular localization can be determined with *in situ* immuno-fluorescence assay. All those control requirements clearly emphasize that loss of function experiment could be easier to set up and especially could be more biologically relevant since all assays are done on endogenous molecules.

Lentiviral transduction is a great tool to insert exogenous genetic information within hard to transfect cells. However, genetic insertion cannot be controlled and can have important consequences for both host cell and transgene. For this reason, it is important to generate independent clones in order to exclude an off-target impact due to the site of insertion. In my case, monoclonal cell lines can be obtained with microfluidic, limiting dilution or manual cell picking. On the last part of this manuscript (III.B.2.b.1.6), I presented the first results obtain with a polyclonal MET16 BC-3 cell lines for both pLenti6.3 and pLKO.1-TRC. ROA1 and MSI2H stably transduced cells have also been produced but not yet analysed.

IV.D. Candidate modulation and effect on miRNA

Several methods are available to study pri-miRNA maturation from a kinetic aspect to a snapshot point of view. One of the most powerful approaches to gain insights on the *in vivo* miRNA biogenesis kinetics is the pulse chase RNA labelling followed by BrU-RNA isolation and deep sequencing. This method allows to discriminate kinetic processing of the pri-miRNA into its corresponding intermediate and mature miRNA in a global or specific point of view, when respectively assessed by deep sequencing or qRT-PCR (Herzog *et al.*, 2017; Louloupi and Orom, 2018). However, for a snapshot point of view and for a given miRNA, northern blot and qRT-PCR experiment could be suited.

Working in a heterologous system, within cells that do not express KSHV's miRNA, could be really difficult as seen for the psiCHECK-2 assays. Concerning the northern blot analysis of overexpressed candidate samples, I normalized each individual KSHV miRNA signal by the sum of all other mature KSHV miRNA signal assessed within a given sample. When biological triplicates are analysed, it displays important variation notably for the miRNA-K6 and -K11. It could have been acceptable if the tendency was consistent. On the Figure 52.C, as the error bars suggest, mature miRNA signal is sometimes considered as more or less expressed compared to the control suggesting a similar issue as the psiCHECK-2 assays. In the same assay, let-7a-1 mature miRNA has been normalized using the tRNA-Val signal as a classical house-keeping gene for endogenous RNA loading control. In these triplicate experiments, let-7a-1 expression is repressed and enhanced respectively for ROA1 and MSI2H overexpression. This result was expected for ROA1 overexpression but not for MSI2H. First, ROA1 output indicates that as at least for this candidate, cloning and expression kept ROA1 molecular function intact at least for its let-7a-1 repression function. MSI2H, for its part, was initially known to repress miR-7 biogenesis through inhibitory pri-miRNA structural rearrangement (Choudhury *et al.*, 2013). It is not a surprising to see contradictory effect of a given cofactor for different miRNAs. ROA1 is the perfect example of this dual-properties since it represses let-7a-1 through KHSRP antagonized function (Michlewski and Cáceres, 2010) and enhance miR-18a though favourable pri-miRNA structural rearrangement (Kooshapur *et al.*,

2018). Here, MSI2H could act as a competitor of ROA1 and favours pri-let-7a-1 processing via structural rearrangement or microprocessor recruitment. Anyway, it will be indispensable to validate the molecular function of my cloned MSI2H. Opposite approach, via interfering RNA, should be also performed to observed opposite effect and bring more observations before concluding on its implication for let-7a-1 processing. I also quantified pri-miRNAs by qRT-PCR analysis. With the qRT-PCR, I normalized each pri-miRNA amplification with an amplification signal obtained on the 5' part of the cluster 200 nts above the first SL-miRNA. At this step, I should have assessed pri-let-7a-1 signal to check if I can retrieve similar result. But since my main objective was to look for KSHV miRNA modulation, I focused on them. On the Figure 47, it seems that the 5' cluster signal is more stable than the pri-miRNA, as the pri-miRNA increases upon DROSHA silencing. Here, DROSHA silencing impact on pri-miRNA assessment could be under-estimate if the 5' part of the cluster is quickly degraded after the first SL-miRNA cleavage, highlighting again the need to work with naturally infected cells. Even though 5' cluster normalization does not seem to be perfectly suited, MET16 overexpression seems to favour pri-miR-K11 cleavage. Since northern blot assays did not display similar results, it is difficult to conclude anything except that it is mandatory to work within infected cells. Thus, at this time, I started to produce lentiviral particles to transduce BC-3 infected cells. In all the different experiments, we are supposed to detect an increase of the mature miRNA when the corresponding pri-miRNA is dropping if the modulated candidate only acts on the nuclear processing. Within the non-infected HEK293 cell line, there is no clear correlation between the modulation of the mature miRNA and its corresponding pri-miRNA, except for the pri-miRNA-K11 upon MET16 overexpression

Modulation of the MET16 protein displays non consistent result depending on the cell line used or the shRNA version. For example, in the HEK cell upon over expression the pri-miRNA-K11 decrease and the mature miRNA increase (Figure 52). We obtained exactly the same result but with the silencing (2b) condition within lymphocyte cells (Figure 53). This can suggest that a specific cell line m^6A reader is present and drive positively the processing in the HEK cells or negatively the processing in the lymphocyte. It is also possible that the presence of the m^6A does not preferentially select an RBP but rather modify its function. Both ideas could explain the differential action of ROA1 which is negative for pri-let-7a-1 and positive for pri-miR-18a. Unfortunately, within the infected lymphocyte MET16 overexpression displays no impact on all the transcripts assessed. However, knock-down of MET16 cells, seems to decrease RTA mRNA. RTA is the main KSHV transcription factor controlling virus reactivation from latency to lytic replication (Xu *et al.*, 2006; Purushothaman *et al.*, 2015). RTA mRNA could be, among other mechanism, directly impacted by MET16 as well as indirectly through translational inhibition via miRNAs -K7 and -K9 (Dölken *et al.*, 2010; Umbach and Cullen, 2010; Gay *et al.*, 2018). Knowing that RTA induces lytic cycle, its expression is correlated to other delayed early viral gene (e.g. ORF45, K8) as well as intermediate viral gene (e.g. ORF44). In the opposite way, LANA (latency associated nuclear antigen) the main KSHV's protein implicated in latency maintenance by repressing RTA transcription, should be decrease upon RTA expression. Thus, it could be easier to validate the RTA inhibition upon MET16 overexpression rather than characterizing MET16-mediated RTA inhibition mechanism.

Taken together we can say that miRNA have to be studied in their natural context, and that monoclonal cell lines have to be generated to facilitate the study of their impact on miRNA biogenesis. We can also say that northern blot and qRT-PCR can eventually be optimized or improved to obtain better results even though our approach seem to work with let-7a-1. Finally, it will be important to continue to include let-7 in all the following experiment as it acts as a good positive control. MET16 could be important for the KSHV viral cycle and eventually for the pri-miRNA-K11 maturation as well as its neighbour stem loop precursors.

IV.E. RNPs fascinating world

To finish this discussion, I wanted to briefly talk about the fascinating RNPs biology. As mentioned, every RBP and RNA interactions are dictated by inherent affinity, RNA and RBP concentration and competition. Furthermore, any RNA is generally bound by an important variety of proteins that act simultaneously, after one another, or in a mutually exclusive manner. As a specific protein interacting with a given RNA can modify its function (and vice versa), RBP-RBP interaction can also modify their RNA binding patterns (Hentze *et al.*, 2018). Thus, the number of theoretical RNPs is remarkably large. It is also important to cautiously use terms like specific and non-specific interaction, generally considered respectively as relevant or non-relevant interaction. Biologically speaking it is more complex than this binary point of view (Jankowsky and Harris, 2015). Indeed, lot of important RBPs involved in fundamental processes such as the translation initiation complex or the exon junction complex can be considered as displaying non-specific interaction with RNAs as specific structural and sequence features do not seem required. Concerning so-called specific RBPs displaying conventional RNA binding domains, they can recognize their RNA partners via an RNA particular sequence or shape such as respectively LN28B and the apical loop of pri-let-7a-1 or the 40S ribosomal subunit and internal ribosome entry site. In addition, many known RBPs do not necessarily possess well-characterized or conventional RNA binding domains such as ribosomal proteins. Finally, for a few years now, new RNA binding domain are discovered such as intrinsically disorder protein region, that can influence RNA metabolism, protein aggregation and liquid-like droplet or granule formation (Calabretta and Richard, 2015; Järvelin *et al.*, 2016). These granules generate hub structures which can be considered as new organized structures within crowded nucleoplasm or cytoplasm to enable efficient and regulated molecular processes in a spatial and temporal manner. In addition, their composition can vary depending on the cell cycle or stimuli (Dellaire *et al.*, 2006), and ATPase proteins appear to play important roles in these granule structures (Jain *et al.*, 2016). Thus, it indicates that molecules are not homogeneously distributed among cellular compartment due to the enthalpy driven force of interacting molecules which is stronger than the entropy of individual molecules. Since the microprocessor have been detected within paraspeckles (Jiang *et al.*, 2017), one can easily imagine that miRNA maturation, even though consider as a co-transcriptional process, can also occur in this subcellular granules. It could be interesting for the cell to eventually mature its miRNAs in different subcellular area depending on various stimuli since the protein and RNA composition can differ, therefore modulating the biological process (Figure 55).

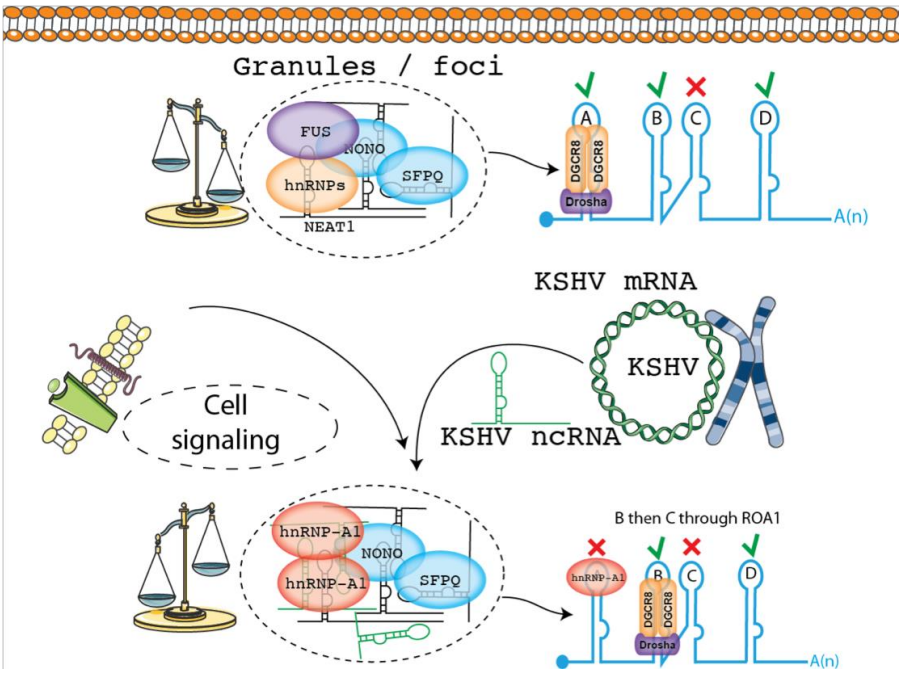


Figure 55. Representation of specific subcellular localization of RNP complexes (granules, foci…). Upon different stresses or presence of ncRNA, those RNP complexes composition is modulated.

To conclude, I want to emphasize that depending on the physical parameters (temperature, ionic forces, concentration…), partner binding, competitor, subcellular localization, RNP complexes could vary as well as their corresponding processes. This is why all RNPs features should be always considered in a given context and cannot be generalized therefore explaining, at least partially, RBP moonlighting properties.

V. CONCLUSION AND PERSPECTIVES

During this project, I performed RNA chromatography experiments for the different KSHV SL-miRNAs as well as 2 cellular SL-miRNAs (-let7 and -155). All the chromatography, done in either endothelial or lymphocyte cells, have been validated thanks to the positive control RNA bait: SL-miRNA-let7. For all the RNA baits used, I obtained lists of potential binders (see part III.A.1.e and III.A.4.b.1.2). In addition, I proposed an improved chromatography elution method based on DNase treatment as well as a data-mining procedure (see part III.A.3). Reporter genes for screening assays have been generated and extensively tested without relevant output in the experimental conditions used (see part III.B.2.b.1.1). However, both reporter systems could potentially be used in other cellular contexts as described in part IV.B.

I selected several candidates for which I designed multiple tools to work with. Lentiviral backbone vectors have been chosen and validated to generate either stable silencing or overexpression respectively with pLKO.1-TRC and pLenti6.3 plasmids (see part III.B.2.b.1.4). Cloning procedure, transfection and transduction protocols have been set up for the upcoming experiments as well as tools to assess miRNA modulation (see part III.B.1). These procedures have been tested with MET16 which, unfortunately, does not seem to be, at least directly, implicated in KSHV miRNA biogenesis modulation. The results I obtained indicate however

that the MSI2H protein could be involved in the maturation of let-7a-1 precursor. Transduced lymphocyte cells have been generated for this candidate at the end of my thesis time as well as for ROA1.

The following suggested experiments could be started for MSI2H and SL-miRNA-let7 for example (Figure 56). First, positive or negative impact of the candidate modulation on miRNA biogenesis should be determined. It can be done within a cell line that express the SL-miRNA with luciferase reporter, northern blot or qRT-PCR as I tried to do during my project. Transcriptional effect has to be excluded. pri-miRNA level should not be modulated by the candidate. If it is the case, it means that the candidate can favour transcription through promoter binding and polymerase II recruitment or epigenetic modification. Transcriptional mechanism can be explored with chromatin immunoprecipitation (CHIP) and co-immunoprecipitation (co-IP). However, if the candidate does not modulate pri-miRNA transcription, it clearly suggests an implication at the post-transcriptional level that could be studied following the pipeline described in Figure 56. Once a candidate has been characterized to impact positively or negatively the miRNA biogenesis at the post-transcriptional level interaction with the microprocessor can be assessed with co-IP. Candidates can either interact directly with the microprocessor (ILF2/ILF3) or indirectly in a RNA dependent (KHSRP, LN28B). Cofactor interacting with the microprocessor do not seem to modulate SL-miRNA fold but rather directly impact DROSHA recruitment and cleavage. Concerning the cofactor that interact with the pri-miRNA, they can act with two mechanisms, SL-miRNA structural rearrangement or binding competition. Folding rearrangement could be determined with 2'-hydroxyl acylation analysed by primer extension sequencing (SHAPE). Binding competition can be assessed with electro mobility shift assay (EMSA) or pulldown experiment. In addition, it would be interesting to determine affinity parameters. Isothermal titration calorimetry is the main approach that study variation of enthalpy of two individual molecules interacting with each other. Enthalpy variation is directly correlated to affinity parameters (K_{on}, K_{off}, K_d) (Sakamoto *et al.*, 2018). Once the affinity parameters are determined it could explain RNPs formation balance. Once integrated to proteomic and transcriptomic data it can help to understand why at limiting concentrations of a given protein, low affinity non-consensus sites in highly expressed RNAs can efficiently compete for protein binding with high affinity consensus sites in an RNA expressed at a lower level (Jankowsky and Harris, 2015). In other words, it can explain how mild or slight modifications of the proteome can modulate RNA processes, highlighting the complexity of RNP dynamics. I strongly believe that RNPs should be studied through a biophysic point of view together with integrated proteomic and transcriptomic data in order to correctly decipher RNA processes and RNP dynamic at the cellular level. In the miRNA biogenesis field, switchSENSE technology could be a relevant approach to deeper characterize binding kinetics and affinity, conformational changes, nuclease activity and competition binding (Clery *et al.*, 2017; Ponzo *et al.*, 2019). Finally, biological implications of the candidate can be assessed to first determine its global role in the miRNA biogenesis pathway as well as its impact on cellular or viral replication and fitness.

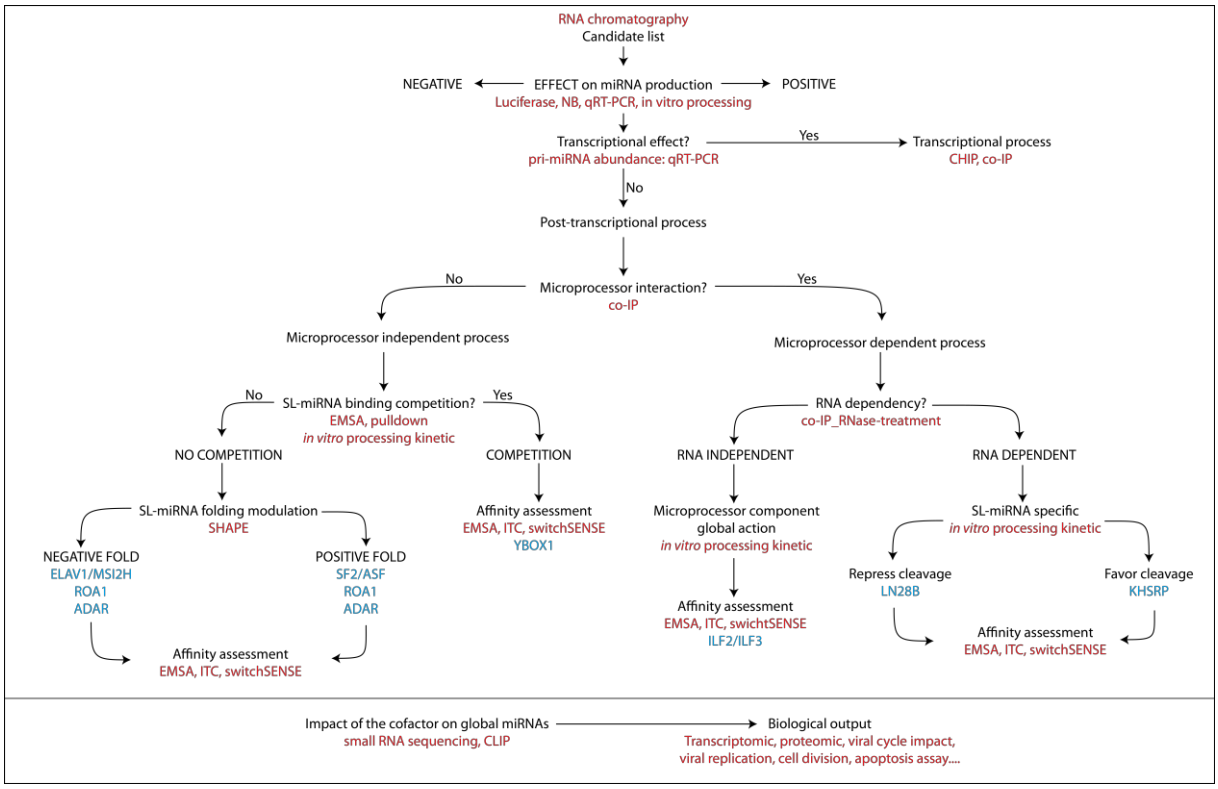


Figure 56. Representation of experiments needed to assess cofactor involvement in specific miRNA biogenesis. Red labelled correspond to technical approaches. Blue labelled correspond to RBP example.

Taken together, I overcame and resolved multiple technical issues and proposed an analytical pipeline to begin the candidate validations. This project is far from being complete and more work will be needed to define co-factors modulating the biogenesis of KSHV miRNAs. As herpesviruses co-evolved with their hosts for millions of years, their genetic expression is perfectly adapted and deeply intertwined with their host-cell proteome and transcriptome. Consequently, the regulation network between viral and cellular genes is incredibly complex. The purpose for a virus to expressed clustered miRNAs seems clear. In the cases of KSHV, with only one transcript unit, it can regulate a number of fundamental cellular processes through its mature miRNAs expression in a dynamic manner while remaining hidden to the immune responses. The complete understanding of the impact of KSHV clustered miRNA on its fitness, requires the study of a multitude of processes. Regardless of all transcription regulation, first, all of KSHV SL-miRNA features (structures, sequences, abundances) have to be defined. Then, RBPs involved in all KSHV SL-miRNAs cleavage have to be determined (binding site, mechanism, affinity parameters). Finally, cleavage order and kinetics need to be assessed. These three requirements should be determined in the different virus host-cells, as well as upon different challenges or viral cycle steps. Taken together, data integration could then allow to understand strategies used by KSHV via its miRNAs for example to hide its genome, induce cell division, control cell cycle, and more generally to respond to changes in the cellular environment and pathology induction. A short-term perspective could be focused on the roles played by co-factors in the maturation on the 5' proximal KSHV SL-miRNAs. Indeed, processing of upstream SL-miRNAs could impact

downstream SL-miRNA maturation. Thus, it could be great to determine if among the KSHV SL-miRNAs one of them could be considered a master-regulator that can impact on other SL-miRNAs processing (Figure 57).

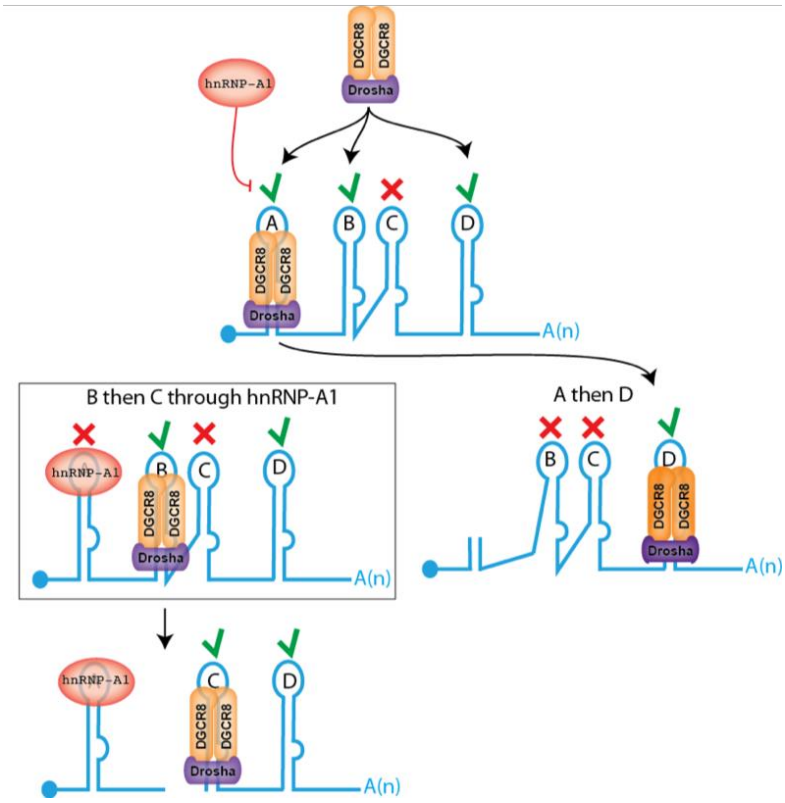


Figure 57. Representation of the pri-miRNA fold dynamism upon cleavage. Each cleavage step induces fold modulation. Cleavage order of SL-miRNA within a SL-miRNA cluster represent another regulation layer. A given RBP can act as a master regulator to force the first cleavage on a given SL-miRNA.

To answer this question, *in vitro* processing of the complete pri-miRNA together with specific SL-miRNA co-factor modulation could be performed. Candidate binding site can be roughly localized using SL-miRNA mutant (e.g. terminal loop or ssRNA flanking sequence mutant) instead of WT SL-miRNA and can be confirmed by electromobility shift assay approach (EMSA). Performing *in vitro* processing with the complete pri-miRNA cluster would allow to determine all SL-miRNA cleavage speed for a given candidate modulation. If the modulation of a candidate changes the kinetic maturation of a given SL-miRNA it could suggest a potential important impact for the downstream SL-miRNA on the cluster. Indeed, as this processing is co-transcriptional, it means that neo-synthesized RNA fold differently at the exit part of the polymerase II depending on the 5' RNA sequence available for interaction. Thus, upstream SL-miRNA cleavage will certainly modulate the fate of the downstream miRNA precursors.

Temporal and spatial miRNAs regulation is fundamental for the cell. Stimuli-mediated miRNA regulation is also of prime importance for the cell and even more for viral miRNAs. Therefore, all the possible induced post transcriptional and translational modifications of molecules related to miRNA biogenesis have to be taken into account to gain insight in miRNA

biogenesis dynamism. Those modifications surely impact RNPs formation as well as their subcellular distribution among a cellular compartment. Together with the required miRNA precursors features (fold and sequence), it creates a complex multilayer regulation process for which important discoveries are still hidden. Furthermore, it is possible that pri-miRNAs possess additional cis-regulatory element, apart from miRNAs *per se* such as pri-miRNA-encoded small peptides, that can play important roles in the regulation of miRNA biogenesis or other processes (Lauressergues *et al.*, 2015). Trans-acting RNA should not be excluded to be implicated in the miRNA biogenesis as well as the specific subcellular localization of the miRNA biogenesis. As other gammaherpesviruses, KSHV also expresses long non-coding RNAs (lncRNAs) (Chavez-Calvillo *et al.*, 2018) and circular RNAs (circRNAs) (Tagawa *et al.*, 2018). lncRNAs can act as decoy, scaffold, guide or enhancer of RNPs formation. Apart from its role during KSHV lytic phase, the PAN viral-lncRNA is retained in the nucleus (Conrad and Steitz, 2005) and act as a regulatory factor of host immune response through cellular and viral proteins interactions (Massimelli *et al.*, 2013). Interestingly, EBV lncRNA, sisRNA-2, interacts with the paraspeckle component NONO (Tompkins *et al.*, 2018). This suggests that viral-lncRNAs can modulate RNP granules composition. All viral-ncRNAs could indirectly impact on viral-miRNA biogenesis and/or function by sequestrating specific RBPs or miRNAs, modifying host or viral protein subcellular localization or directly interacting with viral transcripts. Finally, crosstalk between implicated component of miRNAs, piRNAs, snRNAs, snoRNAs, tRNAs or dsRNA immune sensing should not be overlooked and will surely bring novel insights and discoveries in this already fascinating RNPs world.

VI. MATERIALS AND METHODS

VI.A. Plasmids

VI.A.1 Plasmids expressing KSHV's miRNAs

pCA4 vector expressing KSHV's miRNA cluster under the control of an CMV promotor as well as single KSHV's stem-loop miRNA expressing vector have been previously generated in the lab (Suffert *et al.*, 2011) using pcDNA5-FRT/TO backbone (Invitrogen).

VI.A.2 Plasmids over-expressing candidate

Candidates mature mRNA are cloned, using GATEWAY technology (Invitrogen) following provider's instructions, in a pLenti6.3/FLAG-HA/V5-DEST, modified from a pLenti6.3/V5-DEST (Invitrogen #V53306) expression vector. Briefly, FLAG/HA sequence (ATGGACTACAAGGACGACGATGACAAGTACCCTTATGACGTGCCCGATTACGCT) has been added by PCR mutagenesis PCR above the 5' region of the recombination cassette and below the CMV promotor. Candidates cDNA are either obtained from Addgene plasmid repository or by performing a specific reverse transcription of candidate mRNA on purified BC3 cells RNA. Candidate cDNA is then amplified by PCR with specific primers possessing 5' and 3' extension corresponding to the *attB* recombination sequences (Table 9). pDONR221 (Invitrogen, #12536017) is used as a donor vector to shuttle candidate's cDNA into the expression vector pLenti6.3/FLAG-HA/V5-DEST following Gateway kit (Invitrogen) instructions. pDONR221 and pLenti6.3/FLAG-HA/V5-DEST transformed cells are grown on 50 µg/mL kanamycin and 100 µg/mL ampicillin plate respectively at 37°C. Purified plasmids are sequence (GATC, Eurofins) to confirm the good insertion and sequence of the cDNA insert in frame with the 5' FLAG/HA sequence using CMV promoter complementary primer.

Table 9. pLenti6.3 cloning primers.

Gene	Primer	Sequence	Ref_Seq
CIRBP	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGCATCAGATGAAGGCAAAC	NM_001280.2
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTCGTTGTGTGTAGCGTAAC	
FUS	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGCCTCAAACGATTATACCC	NM_004960.3
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAATACGGCCTCTCCCTGC	
GFP	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGTGAGCAAGGGCGAGGAG	
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTTGTACAGCTCGTCCATGC	
HNRPD	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGTCGGAGGAGCAGTTCG	NM_031370.2
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTAGTATGGTTTGTAGCTATTTTGA	
HNRPF	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGATGCTGGGCCCTGAG	NM_001098208.1
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGTCATAGCCACCCATG	
LN28B	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGCCGAAGGCGGGGCT	NM_001004317.3
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTATGTCTTTTTCCTTTTTTGAACTGAAGG	
MET16	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCATGGCTTTCTGAGTAAATCAATGCA	NM_024086.3
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTAGTTAACTGCAACAAGCCTG	
MSI2H	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGAGGCAAATGGGAGCC	NM_138962.3
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATCAATGGTATCCATTTGTAAAGGC	
QKI	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGTCGGGGAAATGGAAAC	NM_001301085.1
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTAGTTGCCGGTGGCGGC	
RBM45	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTGCATGGACGAAGCTGGCA	NM_152945.3
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAGTAAGTTCTTTGCCGTTTGTTAG	
ROA1	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGTCTAAGTCAGAGTCTCC	NM_002136
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTAAAATCTTCTGCCACTGC	
ZNF346	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGAGTATCCCGCGCCG	NM_012279.3
	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGTCTTCCAAGGTGGTTGA	

VI.A.3 Plasmids expressing candidate's shRNA

Candidates mRNA are silenced with the help of plasmid expressing shRNA with the pLKO.1 – TRC backbone (Addgene #10878). shRNA sequence have been designed with the help of The RNAi Consortium (http://www.broad.mit.edu/genome_bio/trc/). Briefly, the 1.9 kbp stuffer present below the polymerase III promoter of the pLKO.1 – TRC vector can be released with AgeI and EcoRI enzymatic digestion and replaced with your shRNA sequence selected with The RNAi Consortium bearing a 5' AgeI digestion site and a 3' EcoRI digestion site (Table 10). DNA oligonucleotides are mix within 100 mM NaCl; 10 mM Tris-HCl pH 7.5 at a final concentration of 4.5 µM. Oligo mix are then heated at 95°C during 5 min and slowly cooled down at room temperature to ensure a good annealing. Annealed oligos are diluted a 400-fold in 50 mM NaCl; 5 mM Tris-HCl pH 7.5 buffer. 2 µg of pLKO.1 – TRC is double digested with AgeI and EcoRI (Fermentas) within 1 X of Orange buffer (Fermentas) during 1 h at 37°C. Double digested pLKO.1 – TRC is then gel purify on an 1 % TAE 1 X agarose gel to extract the 7 kbp fragment with the Monarch DNA gel extraction kit (New England Biology). Ligation is performed with 40 ng of double digested pLKO.1 – TRC, 2 µL of the diluted (1/400) annealed oligos and 10 U of T4 DNA ligase (Thermo Fisher) within 4 mM Tris-HCl pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 500 µM ATP buffer during 2 h at 22°C. Prior to heat shock bacterial transformation (*E. coli*, Stbl3, Invitrogen), ligation product is denatured at 70°C during 5 min. pLKO.1 – TRC transformed bacteria is grown at 37°C on LB agar plate (Roth) with 100 µg/mL final concentration of carbenicillin antibiotic. Final plasmids are sequence to confirm the good insertion and the correct sequence of the coding shRNA using primer oligo (5'-TCTACTATTCTTTCCCCTGCAC).

Table 10. pLKO.1-TRC cloning primers.

Gene	Primer	Sequence	Pol-III insert	Ref_Seq
CIRBP	Forward	CCGGCGGGTCCTACAGAGACAGTTACTCGAGTAACTGTCTCTGTAGGACCCGTTTTTG	(TRCN0000221637)	NM_001280.1
	Reverse	AATTCAAAAACGGGTCCTACAGAGACAGTTACTCGAGTAACTGTCTCTGTAGGACCCG		
ELAV1	Forward	CCGGTTGTTAGTGTACAACTCATTTCTCGAGAAATGAGTTGTACACTAACAATTTTTG	(TRCN0000276186)	NM_001419.2
	Reverse	AATTCAAAAATTGTTAGTGTACAACTCATTTCTCGAGAAATGAGTTGTACACTAACAA		
HNRPD	Forward	CCGGACTATGGATATGGTGATTATACTCGAGTATAATCACCATATCCATAGTTTTTTG	(TRCN0000293284)	NM_031370.2
	Reverse	AATTCAAAAAACTATGGATATGGTGATTATACTCGAGTATAATCACCATATCCATAGT		
HNRPF	Forward	CCGGAGCGACCGAGAACGACATTTACTCGAGTAAATGTCGTTCTCGGTCGCTTTTTTG	(TRCN0000230710)	NM_004966.3
	Reverse	AATTCAAAAAAGCGACCGAGAACGACATTTACTCGAGTAAATGTCGTTCTCGGTCGCT		
LN28B	Forward	CCGGGCCATAGTATTGGTCCTGTTACTCGAGTAACAGGACCAATACTATGGCTTTTTG	(TRCN0000143619)	NM_001004317.1
	Reverse	AATTCAAAAAGCCATAGTATTGGTCCTGTTACTCGAGTAACAGGACCAATACTATGGC		
MET16-v1	Forward	CCGGCCAAAGTAACGTACACTGAATCTCGAGATTCAGTGTACGTTACTTTGGTTTTTG	(TRCN0000134846)	NM_024086.2
	Reverse	AATTCAAAAACCAAAGTAACGTACACTGAATCTCGAGATTCAGTGTACGTTACTTTGG		
MET16-v2	Forward	CCGGCCTGTACTTACCTACTCACATCTCGAGATGTGAGTAGGTAAGTACAGGTTTTTG	(TRCN0000136957)	NM_024086.2
	Reverse	AATTCAAAAACCTGTACTTACCTACTCACATCTCGAGATGTGAGTAGGTAAGTACAGG		
MSI2H	Forward	CCGGCCCAACTTCGTGGCGACCTATCTCGAGATAGGTCGCCACGAAGTTGGGTTTTTG	(TRCN0000062811)	NM_138962.2
	Reverse	AATTCAAAAACCCAACTTCGTGGCGACCTATCTCGAGATAGGTCGCCACGAAGTTGGG		
QKI	Forward	CCGGTAGGTGCGGTGGCTACTAAAGCTCGAGCTTTAGTAGCCACCGCACCTATTTTTG	(TRCN0000233374)	NM_006775.1
	Reverse	AATTCAAAAATAGGTGCGGTGGCTACTAAAGCTCGAGCTTTAGTAGCCACCGCACCTA		
RBM45	Forward	CCGGTGGGCTACGTACGATACTTAACTCGAGTTAAGTATCGTACGTAGCCCATTTTTG	(TRCN0000416975)	NM_152945.2
	Reverse	AATTCAAAAATGGGCTACGTACGATACTTAACTCGAGTTAAGTATCGTACGTAGCCCA		
ROA1	Forward	CCGGTGTAGTTGAACTGATAGTTACCTCGAGGTAACTATCAGTTCAACTACATTTTTG	(TRCN0000368906)	NM_002136.2
	Reverse	AATTCAAAAATGTAGTTGAACTGATAGTTACCTCGAGGTAACTATCAGTTCAACTACA		
ZNF346	Forward	CCGGCTCATTGGTCCGGGCTAATTCCTCGAGGAATTAGCCCGGACCAATGAGTTTTTG	(TRCN0000219820)	NM_012279.2
	Reverse	AATTCAAAAACTCATTGGTCCGGGCTAATTCCTCGAGGAATTAGCCCGGACCAATGAG		

VI.A.4 Plasmids pmiRGLO: SL-cleavage reporter

SL-miRNA sequence, substrate of the microprocessor complex, has been added within the 3' UTR of the firefly luciferase of the pmiRGLO Dual Luciferase reporter (#E1330, Promega) to assess the cleavage efficiency of the SL-miRNA from the firefly mRNA. SL-miRNA correspond to the pre-miRNA sequence flanked with the 50 nts wildtype 5' and 3' sequence to mimic a Drosha/DGCR8 substrate of a primary miRNA transcript. pmiRGLO Dual Luciferase reporter (Promega) plasmid possessing the Gateway cloning cassette at the 3' end of the firefly luciferase gene has been previously generated in the lab using PCR amplified *AttR1-ccdB-AttR2* recombination site ligated into the *Multi Site Cloning* with restriction sites. SL-miRNA insert is generated by annealing oligonucleotides corresponding to the pre-miRNA of interest with a 5' anchor sequence followed by a nested PCR-based addition of the *attB* recombination site using the anchor sequence (Table 11).

Table 11. pmiRGLO cloning primers.

pmiRGLO SL-miRNA cloning		
SL-miR-K1	Forward	AAAAAGCAGGCTCGGTTGTTTACGCAGGGTGCGGTGC
	Reverse	AGAAAGCTGGGTCGCGCCGGGAACACGCAGGTGGCAG
SL-miR-K2	Forward	AAAAAGCAGGCTAAGTCCGTCAACCAGATTTAAGATTAAG
	Reverse	AGAAAGCTGGGTCGTTGGTTGGGAGTTTGGGGACC
SL-miR-K3	Forward	AAAAAGCAGGCTGTCCCCAAACTCCCAACCAACGCAAC
	Reverse	AGAAAGCTGGGTACCAGGCGGCGTTGTGGCCACGGGC
SL-miR-K4	Forward	AAAAAGCAGGCTCGAACCGCCCGTGGCCACAAC
	Reverse	AGAAAGCTGGGTGTTTGAGAGGCGTAGACATCCG
SL-miR-K5	Forward	AAAAAGCAGGCTCCTCTCAAACCTCGTGGGCACGGCG
	Reverse	AGAAAGCTGGGTACCCGCATAGGTTTTTGTGGCACC
SL-miR-K6	Forward	AAAAAGCAGGCTCGTAAGCGTCTGGATCGACACAAAC
	Reverse	AGAAAGCTGGGTATGATGAAGAACACACAGAACAATAAC
SL-miR-K7	Forward	AAAAAGCAGGCTGTACGCGTGCCCACCGATGAGATAC
	Reverse	AGAAAGCTGGGTTTCAGCTTTGCGTGCTTTCGACGTC
SL-miR-K8	Forward	AAAAAGCAGGCTGGGGTGACAAAGCATGCACTGGAAATC
	Reverse	AGAAAGCTGGGTCACACTTTATGTTGGCCGCGTGCC
SL-miR-K9	Forward	AAAAAGCAGGCTCAGCTGAGTCATCGCAGCCCCTATTC
	Reverse	AGAAAGCTGGGTTTTATACCTTGTTTAATTGTAG
SL-miR-K11	Forward	AAAAAGCAGGCTTTGGGAATCCTACCTCCACGCTC
	Reverse	AGAAAGCTGGGTGCCGTGAAGGTCGGGAACCCGC
SL-miR-let7	Forward	AAAAAGCAGGCTACTGTGATTCCTTTTCACCATTC
	Reverse	AGAAAGCTGGGTAACTGACTTTCTATCAGACCGC

VI.A.5 Plasmids psiCHECK-2: miRNA activity reporter

psiCHECK-2 luciferase miRNA activity reporter (#C8021, Promega) plasmid possessing the Gateway cloning cassette at the 3' end of the firefly luciferase gene has been previously generated in the lab (Suffert *et al.*, 2011). Perfect or imperfect binding site for the mature KSHV's miRNA have been insert within the 3' UTR of the firefly luciferase gene of the psiCHECK-2 plasmid to assess miRNA post transcriptional target repression. Perfect or imperfect binding site for KSHV's mature miRNAs are inserted in the 3' UTR of the firefly gene using the Gateway cassette. Perfect and imperfect miRNA binding site are generated by annealing oligonucleotides corresponding to the binding site (perfect or imperfect) with a 5' anchor sequence followed by a nested PCR-based addition of the *attB* recombination site using the anchor sequence (Table 12). pDONR221 (Invitrogen, #12536017) is used as a donor vector to shuttle binding site PCR products into the Luciferase reporter vector psiCHECK-2 following Gateway kit (Invitrogen) instructions.

Table 12. psiCHECK-2 cloning primers.

psiCHECK-2 cloning imperfect match		
kshv-miR-K12-1-5p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*GCTTACACCCAG**AAA**CCTGTAAT*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*ATTACAGG**TTT**CTGGGTGTAAGC*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-2-5p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*CAGATCGACCC**CCT**CTACAGTT*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*AACTGTAG**AGG**GGGTCGATCTG*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-3-5p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*TCGCTGCCGTCC**AGT**GAATGTGA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TCACATTC**ACT**GGACGGCAGCGA*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-4-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*TCAGCTAGGCC**AGT**GTATTCTA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TAGAATAC**ACT**GGCCTAGCTGA*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-5-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*ACCGGCAAGTTC**GTC**GCATCCTA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TAGGATGC**GAC**GAACTTGCCGGT*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-6-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*CTCAACAGCCC**CTT**AACCATCA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TGATGGTT**AAG**GGGCTGTTGAG*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-7-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*GCGCCAGCA**TGT**TGGGATCA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TGATCCCA**ACA**TGCTGGCGC*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-8-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*TGCTCTCTCA**CAG**GCGCCTAG*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*CTAGGCGC**CTG**TGAGAGAGCA*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-9-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*TTACGCAGCTG**GCA**ATACCCAG*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*CTGGGTAT**TGC**CAGCTGCGTAA*GGCTTAAACAAGGTGGCGAT
kshv-miR-K12-11-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*TCGGACACAGG**GAT**AGCATTAA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TTAATGCT**ATC**CCTGTGTCCGA*GGCTTAAACAAGGTGGCGAT
hsa-let-7a-1-3p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*AACTATACAAC**GAT**CTACCTCA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TGAGGTAG**ATC**GTTGTATAGTT*GGCTTAAACAAGGTGGCGAT
hsa-miR-155-5p	Forward	AAAAAGCAGGCTATCGCCACCTTGTTTAAGCC*ACCCCTATCACG**TAA**AGCATTAA*ATTAGACCTACGCACTCCAG
	Reverse	AGAAAGCTGGGTCTGGAGTGCGTAGGTCTAAT*TTAATGCT**TTA**CGTGATAGGGGT*GGCTTAAACAAGGTGGCGAT

VI.A.6 Bacterial transformation and plasmids purification

E. coli bacteria (Stbl3, Invitrogen) are transformed by heat shock procedure consisting of 30 min incubation in ice followed by 30 s at 42°C (water-bath) and 5 min in ice. Cell are then seeded on LB-agar (Roth) plates containing 100 µg/mL ampicillin/carbenicillin or 30 µg/mL kanamycin and incubated at 37°C. Bacteria clones are then cultivated in LB liquid media (Roth) at 37°C under agitation. Purified plasmids are obtained with the NucleoSpin Plasmid kit (Machery-Nagel). After the plasmid purification, insert presence and sequence are then verified by sequencing (GATC, Eurofins).

VI.B. Cell culture

VI.B.1 Cells growth

HEK293T, HEK293-GripTite and Hela cells are cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10 % foetal calf serum at 37°C in a humidified 5 % CO_2 atmosphere. BCBL-1 and BC3 lymphocytes are grown in a Roswell Park Memorial Institute medium (RPMI 1640, Gibco) supplemented with 10 % of foetal calf serum at 37°C in a humidified 5 % CO_2 atmosphere. Lymphocyte cells transduced with pLenti6.3 constructions are cultured with 16 µg/mL of blasticidin. Lymphocyte cells transduced with pLKO.1 construction are cultured with 1 µg/mL of puromycin.

VI.B.2 Transfection

Adherent cells transfection is performed with either Lipofectamine2000 (Thermo Fisher) or TurboFect (Thermo Fisher) following provider's instructions. Suspension cells transfection is performed with Oligofectamine (Thermo Fisher) following constructer's protocol. The

quantity of plasmid varies depending on the culture format, cells used, plasmids and experimental protocols from 2 µg for a 6-wells plate (candidate plasmid) to 25 ng for a 48-wells plate (Luciferase reporter).

VI.B.3 Lentivirus production

Lentivirus containing candidate cDNA or shRNA are produced using either the 2nd or the 3rd lentivirus generation production technology. For the 2nd generation, psPAX2 plasmid expressing HIV-1 gag as well as HIV-1 pol, pVSV-G plasmid expressing the VSV envelope protein G and the pLenti6.3-candidate or the pLKO.1-shRNA-candidate are co-transfected. For the 3rd lentivirus generation, the psPAX2 plasmid is split in two different plasmids: pMDLg/pRRE expressing HIV-1 gag as well as HIV-1 pol proteins and pRSV-Rev expressing HIV-1 rev protein. 3rd generation lentivirus increases the global security of the system without affecting the lentivirus production. Lentivirus are produced in 2 mL of culture medium in a 6-wells plate using HEK-GRIP cells. 4.10^5 cells are seeded one day before being co-transfected with 1.3 µg of psPAX2 (2nd generation) or 1.2 µg of pMDLg/pRRE; 1.2 µg pRSV-Rev (3rd generation) with 0.33 µg of pVSV-G and 1.7 µg of the pLenti6.3-candidate or pLKO.1-shRNA-candidate using 8.3 µL of lipofectamine2000 following provider's instruction. 48 h post transfection, the supernatant containing the lentiviral virions is filtered at 0.45 µm. Fresh supernatant is then directly used without any freezing step to transduce BC3 cells.

VI.B.4 Lentivirus transduction

$0.5.10^6$ BC3 cells within 200 µL of RPMI 10 % FBS without antibiotic medium are mixed with 500 µL of fresh supernatant containing lentivirus virions with 10 µg/mL final concentration of polybrene. Transduction is performed with a centrifugation step at 600 g for 2 h at 32°C. The supernatant is then discarded, and the cells resuspended with 500 µL of medium containing 50 % of fresh RPMI 10 % FBS without antibiotic and 50 % of filtered growing cells RPMI 10 % FBS without antibiotic medium. 48 h post transduction the antibiotic pressure is applied with 1 µg/mL of puromycin for the pLKO.1 transduced cells or 16 µg/mL of blasticidin for the pLenti6.3 transduced cells.

VI.C. RNA pulldown

SL-miRNA baits are produced by an *in vitro* transcription using a T7-PCR product template.

VI.C.1 PCR template

PCR amplicons harbouring a 5' T7 promotor sequence have been produced by Phusion High-Fidelity DNA polymerase (Thermo Fisher) following manufacturer's instruction to generate transcription templates for each individual SL-miRNA. Forward specific primers possessing a 5' T7 promotor sequence and reverse specific primers were used on either pCA4 plasmid for viral SL-miRNAs or BC3 DNA extract for cellular SL-miRNAs (hsa-let-7a-1, hsa-miR-155).

VI.C.2 In vitro transcription

PCR products served as a template (Table 13) for a T7 transcription reaction within 100 µL reaction volume composed of 0.12 mM MgCl2; 0.16 M Tris HCl pH 7.5; 44 mM DTT; 8 mM spermidine; 0.04 % triton X-100; 25 mM rNTP (Promega); 40 U RNase inhibitor Ribolock (ThermoFisher) and 400 U T7 polymerase (Ambion). Transcription last for at least 3 h at 37°C, then subjected to a DNase I (ThermoFisher) treatment with 2 U during 15 min at 37°C. After 5 min of denaturation in a loading buffer composed of 8 M urea; 50 mM EDTA; 0.05 % bromophenol blue, SL-miRNA molecules were gel purify with a denaturing 8 % acrylamide urea 7 M TBE 1 X gel. UV shadow visualisation (Fluor-coated TLC plate, Life Technologies) allowed SL-miRNA excision from the gel. Gel purified transcripts were then passively eluted from the gel under an overnight agitation within 10 mM EDTA pH 8; 0.5 % SDS; 100 mM sodium acetate pH 5 buffer. Eluted SL-miRNAs are then purified using Phenol/Chloroform/Isoamyl-acid (25/24/1, v/v) and concentrated with classical EtOH precipitation.

Table 13. T7 SL-miRNA primers.

Primers for T7 PCR template		
SL-miRNA-K1	Forward	GAAATTAATACGACTCACTATAGGAGGACGGCCGGATGCGGGCG
	Reverse	GTGTTTTGCCCAAACACGAG
SL-miRNA-K2	Forward	GAAATTAATACGACTCACTATAGGTATAATGGGTCTACTTCGCTAAC
	Reverse	TATAATTGGTGTCACCGCGG
SL-miRNA-K3	Forward	GAAATTAATACGACTCACTATAGGACAATGCCTGTAATGGGCTATC
	Reverse	GTCGCTTGGACCTGGAGGGG
SL-miRNA-K4	Forward	GAAATTAATACGACTCACTATAGGGAACCGGGCAGTATAACTAGC
	Reverse	GCTGGACGGAGGTAGTATAATCAGC
SL-miRNA-K5	Forward	GAAATTAATACGACTCACTATAGGACTAAGGGGGAGTTTGACCTAG
	Reverse	TGGTCTTATGAGCGGGCTTG
SL-miRNA-K6	Forward	GAAATTAATACGACTCACTATAGGGCGGGTTTAGAAAGACTTGTC
	Reverse	ACTGGGCACATGGATGTTTT
SL-miRNA-K7	Forward	GAAATTAATACGACTCACTATAGGAGCCGCGCATATTGGCGTTG
	Reverse	GGCTGGCACACGGGCCGTGAG
SL-miRNA-K8	Forward	GAAATTAATACGACTCACTATAGGCAAGTCAGCGCAGCGCGCGC
	Reverse	CGGCGTGCGCGCCGCGCGCG
SL-miRNA-K9	Forward	GAAATTAATACGACTCACTATAGGAGGTATACCCAGCTGGGTCTAC
	Reverse	CTGGATTTACCCAGCCGGGTTTAC
SL-miRNA-K11	Forward	TAATACGACTCACTATAGGGCCTCATGCGAGGTTCGCTTTGGTCACAGC
	Reverse	CCGCAAGGCACGGGGCCGCATCGG
SL-miRNA-let7	Forward	GAAATTAATACGACTCACTATAGGATGTTCTCTTCACTGTGGGATGAGGTAG
	Reverse	AGACTTTTCTATCACGTTAGGAAAG
SL-miRNA-155	Forward	GAAATTAATACGACTCACTATAGGTTGCTGTAGGCTGTATGCTG
	Reverse	TAGTAACAGGCATCATACAC

VI.C.3 SL-miRNA binding to agarose beads

SL-miRNA 3' end sugar has to be oxidized to be bind to the adipic acid dihydrazide agarose beads (Sigma). 500 pmol of purified SL-miRNAs are oxidized within 400 µL of 5 mM m-periodate (Sigma Aldrich); 0.1 M sodium acetate pH 5 solution in the dark during 1 h at 4°C. After the oxidation step, SL-miRNAs were precipitated with EtOH and resuspended with 500 µL of 0.1 M sodium acetate pH 5 buffer. Simultaneously, 400 µL of 50 % slurry adipic acid dihydrazide agarose beads (Sigma) are washed 4 times with 0.1 M sodium acetate pH 5 solution to finally be resuspended with 300 µL of 0.1 M sodium acetate pH 5 solution. Oxidized SL-miRNAs and washed agarose beads are then incubated together overnight at 4°C. After the

binding step, SL-miRNA-beads are washed 3 times with 1 mL of a 2 M NaCl solution then 3 times more with 1 mL solution composed of 20 mM HEPES-KOH pH 7.6; 6 % glycerol; 0.1 M KCl; 0.2 mM EDTA; 0.5 mM DTT.

VI.C.4 SL-miRNA binding efficiency assessment

Prior to a phosphate^{32} labelling with 10 U of T4 polynucleotide kinase (Fermentas); 25 µCi of [γ-32P]dATP; 1X final concentration of A buffer, SL-miRNA are dephosphorylated with 1 U of Fast Alkaline phosphatase (Thermo Fisher) following manufacturer's recommendation. Gel purified radiolabelled SL-miRNA were then passively eluted from the gel under an overnight agitation within 10 mM EDTA pH 8; 0.5 % SDS; 100 mM sodium acetate pH 5 buffer. Eluted SL-miRNAs are then purified using Phenol/Chloroform/Isoamyl-acid (25/24/1, v/v) and concentrated with classical EtOH precipitation. Radiolabelled SL-miRNA are then oxidized with 5 mM m-periodate (Sigma Aldrich); 0.1 M sodium acetate pH 5 solution in the dark during 1 h at 4°C. After the oxidation step, SL-miRNAs were precipitated with EtOH and resuspended with 0.1 M sodium acetate pH 5 buffer. Radioactive signal was then measured with LS6000SC Cerenkov counting system (Beckman) and followed during all the binding step. Flowthroughs were collected during all the binding assay to calculate binding and elution efficiency.

VI.C.5 Chimeric SL-miRNA generation and beads binding

SL-miRNAs are fixed to a ssDNA-biotinylated sequence to generate a chimeric molecule. ssDNA sequence will be used at the end of the pulldown to specifically elute proteins binding SL-miRNA. For each bait, 500 pmol of a 5' mono-phosphate 3' TEG-biotin ssDNA (/5Phos/TG GAA TTC TCG GGT GCC AAG G/3BioTEG/) are subject to a pre-adenylation step using the 2 times less Mth RNA ligase (New England Biology) than recommended by manufacturer's instructions. Next, pre-adenylated ssDNA is precipitated with EtOH and 0.5 µL glycogen (20 µg/µL stock) before being resuspend with 3 times SL-miRNA excess within 15 % DMSO; 10 mM $MgCl_2$; 10 mM 2-mercaptoethanol; 50 mM Tris HCl pH 7.6; 0,1 mg/mL acetylated bovine serum albumin; 2000 U T4 Rnl2 (1–249) solution to form the chimeric molecule during at least 1 h at 25°C. Chimeric molecule are then purify using Phenol/Chloroform/Isoamyl-acid and precipitated with classical EtOH protocol. Chimera pellet is then resuspended with 200 µL of 0.1 M sodium acetate solution and incubated at 95°C during 1 min followed by a slow cool down step at room temperature to facilitate the SL-miRNA folding. Simultaneously for each bait condition, 75 µL of MagSi-STA 600 (MagnaMedics) magnetic beads are washed with 0.1 M sodium acetate solution. Chimera molecules and washed beads are then incubated together to form the final bait during at least 45 min at room temperature.

VI.C.6 Nuclear extract

For each pulldown, 50.10^6 of either B lymphocyte or endothelial cells are used to obtain a nuclear extract. Cells are centrifuge at 500 g and washed with cold PBS (Gibco). Cell pellet is resuspended with hypotonic buffer composed of 10 mM HEPES pH 7.9; 1.5 mM $MgCl_2$; 10 mM KCl and incubated in ice until the cells have swollen. Cytoplasmic membranes of B

lymphocyte swollen cells are disrupted with a Dounce homogenizer to obtain a nucleus suspension. Cytoplasmic membranes of endothelial swollen cells are disrupted by centrifugation at 1500 g during 10 min at 4°C. Next, nucleus suspension is centrifuge at 1500 g and washed 3 times with hypotonic buffer. Finally, nucleus are lysed within 20 mM Tris HCl pH 7.5; 100 mM KCl; 0.2 mM EDTA; 0.5 mM DTT; 5 % glycerol; protease inhibitor (Roche) solution upon ice water bath sonication procedure (Bioruptor UCD 200®) during 5 min, with cycle of 30 s maximum amplitude and 30 s break. Nuclear sample is then centrifuge at 1000 g during 10 min to get rid of the cellular debris. All steps are controlled by microscopy to ensure a good hypotonic shock as well as the nucleus integrity.

VI.C.7 Nuclear extract quality

At each nuclear extraction steps, cells are observed with an inverted microscope (AXIO-Scope, Zeiss) with DAPI staining using Vectashield (Vectalabs) mounting solution. Nuclear-cytoplasmic ratio is estimated. Decrease of the nuclear-cytoplasmic ratio after the hypotonic shock is required before disrupting the cytoplasmic membrane. If the cells are not swollen, they are put back in the ice with hypotonic buffer for few more minutes. A small fraction of the nuclear fraction is conserved to validate the fractionation procedure using WB analysis.

VI.C.8 Pulldown & elution

Nuclear extract and bait are incubated during 30 min at room temperature on a wheel. 15 µg/mL of yeast total tRNA competitor is then added to decrease non-specific binding. For the agarose beads protocol, beads are centrifuge during 5 min at 100 g and washed 4 times within 20 mM HEPES-KOH pH 7.6; 5 % glycerol; 0.1 M KCL; 0.2 mM EDTA; 0.5 mM DTT solution. Agarose beads are then incubated with 63 mM Tris-HCL pH 6.8; 0.1 % 2-mercaptoethanol; 10 % glycerol; 2 %SDS solution at 95°C during 5min to heat-elute protein bound to the bait. RNA elution from the agarose beads is also performed with 2.5 U and 100 U of RNase A and RNAse T1 treatment respectively within 200 µL of a solution composed of 10 mM Tris-HCl pH 7.2; 1 mM $MgCl_2$; 40 mM NaCl during 1 h at 37°C.

For the magnetic beads protocol, beads are washed 4 times within 20 mM HEPES-KOH pH 7.6; 5 % glycerol; 0.2 M KCL; 0.2 mM EDTA; 0.5 mM DTT solution with the help of a magnetic sample rack. Elution is either performed with 10 U of DNase RQ1 (Promega) following manufacturer's instructions or incubated for the heat elution with 63 mM Tris-HCL pH 6.8; 0.1 % 2-mercaptoethanol; 10 % glycerol; 2 %SDS solution at 95°C during 5min.

VI.D. MSMS identification

VI.D.1 Spectral count

Protein sample are denatured during 5 min at 95° then trypsin digested with 200 ng of trypsin overnight at 37°C. Next, peptides are precipitated with 60 % acetonitrile; 0.1 % formic acid. Peptides are finally resuspended with 0.1 % formic acid and loaded in a hydrophobic high-performance liquid chromatography connected to the mass spectrometer. Institute's MSMS platform identify protein's peptides with the AB SCIEX 5600+ TripleTOF mass spectrometer with the help of ProteinScape (Bruker) and Mascot (Matrix Science) software. Each mass over

electric charge (m/z) of the protein's peptides are assessed and compare to an *in-silico* m/z human peptide database (UniProtKB-SwissProt Taxonomy ID: 9606, 2014). Sequences of common contaminants like keratins as well as KSHV's peptides are also added to the database. The total number of MS/MS fragmentation spectra was used to quantify each protein from three independent biological replicates (spectral count label-free relative quantification). To identify significantly enriched proteins, a statistical analysis by the msmsTests R package using spectral counts was performed. The entire MS dataset was first normalized by the total number of MS/MS spectra (column-wise normalization). The negative binominal model, which is based on the solution provided by the edgeR package, was used. P values were then adjusted using the Benjamini–Hochberg method.

VI.D.2 MaxLFQ

Five µg of protein sample are denatured at 95°C for 5 min before being loaded into a 1-D SDS-PAGE to stack the protein into a single sharp band. Proteins are stained with Colloidal blue after a fixation incubation composed of 50 % EtOH; 3 %phosphoric acid. Stacked band are excised then washed with ammonium hydrogen carbonate; acetonitrile and reduced and alkylated before trypsin digestion. The generated peptides were extracted with 60% acetonitrile in 0.1% formic acid followed by a second extraction with 100% acetonitrile. Acetonitrile was evaporated under vacuum and the peptides were resuspended in 10 µL of H20 and 0.1% formic acid before nanoLC-MS/MS analysis. NanoLC-MS/MS analyses were performed on a nanoACQUITY Ultra-Performance_LC-system (Waters, Milford, MA) coupled to a Q-Exactive Plus Orbitrap mass spectrometer (ThermoFisher Scientific) equipped with a nanoelectrospray ion source. The mass spectrometer was operated in data-dependent acquisition mode by automatically switching between full MS and consecutive MS/MS acquisitions. Survey full scan MS spectra (mass range 300-1800) were acquired in the Orbitrap at a resolution of 70K at 200 m/z with an automatic gain control (AGC) fixed at 3.10^6 ions and a maximal injection time set to 50 ms. All samples were injected using a randomized and blocked injection sequence (one biological replicate of each group plus pool in each block). To minimize carry-over, a solvent blank injection was performed after each sample. Raw MS data processing was performed using MaxQuant software (v 1.5.8.3). Peak lists were searched against a composite database including *Homo sapiens* protein sequences extracted from UniprotKB (Taxonomy ID: 9606, 2017-02-15, 20 195 sequences) using the MSDA software suite (Carapito *et al.*, 2014). Sequences of common contaminants like keratins as well as KSHV's peptides are also added to the database. False discovery rates (FDR) were estimated based on the number of hits after searching a reverse database and was set to 1% for both peptide spectrum matches (minimum length of seven amino acids) and proteins. Data normalization and protein quantification was performed using the LFQ (label free quantification) option implemented in MaxQuant using a "minimal ratio count" of one. The "Match between runs" option was enabled using a 2 minutes time window after retention time alignment. All other MaxQuant parameters were set as default. Among the qualified proteins, descriptive and differential data analysis was performed using the open-source DAPAR and ProStaR softwares (Wieczorek *et al.*, 2017, 2019). Concerning the MSMS_toolkit ("home-made analysis"), each protein LFQ intensity was tested individually for normal distribution and homoscedasticity by

Shapiro-Wilk and Bartlett tests, respectively (p value > 0.01).The proteins with at least three out of four valid values per group as well as the ones “absent” (i.e. 0 valid values) in samples from a given group were sorted out. A Welch’s t-test was applied to compare LFQ intensities of each protein that passed the previous tests between two timepoints. Proteins were considered to be significantly enriched when the Welch’s t-test p-values was below 0.05.

VI.E. Protein analysis

VI.E.1 Protein extraction and quantification

Total proteins are extracted using RIPA buffer composed of 150 mM NaCl; 0.1 % TritonX-100; 0.5 % sodium deoxycholate; 0.1 % sodium dodecyl sulphate; 50 mM Tris HCL pH 8; protease inhibitors (Roche). HEK adherent cells are washed with PBS 1X (Gibco) prior to directly add the lysis buffer on the cell layer. Lymphocyte non adherent cells are centrifuged at 400 g and washed with PBS 1 X before adding the lysis buffer on the cell pellet. Cells resuspended with RIPA buffer are then incubated 10 min in ice with agitation. Next, protein extract is centrifuge at 10,000 g to get rid of the cell debris. Protein extracts are also obtain using Tri-reagent solution (Trizol, Sigma) following manufacturer’s protocol. Protein concentration is assessed by Lowry analysis with the RCDC kit (Bio-Rad) following provider’s instruction using BSA within the same sample’s buffer as standard.

VI.E.2 Western blot

Protein sample are mixed volume to volume with 126 mM Tris-HCL pH 6.8; 0.2 % 2-mercaptoethanol; 20 % glycerol; 4 %SDS; 0.004 % bromophenol blue (2 X Laemmli) buffer at 95°C during 5 min to completely denature proteins prior to SDS-PAGE loading. After migration, proteins are transferred on a nitrocellulose membrane (Hybond nitrocellulose, GE Healthcare). Transfer quality and protein loading visualisation are assessed with a red Ponceau staining composed of 0.1 % red Ponceau; 5 % acetic acid staining solution. Prior to antibody incubation, membranes are blocked with PBS 1 X; 0.2 % tween-20; 5 % milk during at least 1 h at room temperature or overnight at 4°C. Primary and secondary antibodies were incubated respectively overnight at 4°C and 1 h at room temperature (Table 14). After each antibody incubation, blots are washed 3 times with 1 X PBS; 0.2 % tween-20. Antibodies are detected with Fusion FX (Vilber Smart Imaging) high sensitivity camera by adding Luminol (ECL Western blot detection reagents, GE Healthcare) on the blots possessing HRP linked to antibodies.

Table 14. Antibodies.

Provider	Reference	Antibody name	Gene name	Host	Dilution
Thermo Fisher	PA5-29955	ABCF1	ABCF1, ABC50	rabbit	1/500
Santa Cruz	sc-271854	ADAR1 (D-8)	ADAR, ADAR1, DSRAD, G1P1, IFI4	mouse	1/500
Cell Signaling	#13713	CIRBP	cirbp-a, cirbp, cirp, cirp-1	rabbit	1/1000
Santa Cruz	sc-20003	CUG-BP1 (3B1)	CELF1, BRUNOL2, CUGBP, CUGBP1, NAB50	mouse	1/200
Santa Cruz	sc-47731	CUG-BP2 (1H2)	CELF2, BRUNOL3, CUGBP2, ETR3, NAPOR	mouse	1/500
Santa Cruz	sc-393260	CstF-50 (A-5)	CSTF1	mouse	1/200
Santa Cruz	sc-271686	DDX15 (E-6)	DHX15, DBP1, DDX15	mouse	1/200
Cell Signaling	#3364	Drosha (D28B1)	DROSHA, RN3, RNASE3L, RNASEN	rabbit	1/1000
Santa Cruz	sc-5261	ELAV1	ELAVL1, HUR	mouse	1/200
Santa Cruz	sc-137089	ERI-1 (B-10)	ERI1, 3'EXO, THEX1	mouse	1/200
Santa Cruz	sc-55580	Fatty Acid Synthase (A-5)		mouse	1/200
Santa Cruz	sc-398466	FBP3 (E-8)	FUBP3, FBP3	mouse	1/200
Santa Cruz	sc-376411	hnRNP A/B (G-10)	HNRNPAB, ABBP1, HNRPAB	mouse	1/1000
Santa Cruz	sc-374053	hnRNP A2/B1 (B-7)	HNRNPA2B1, HNRPA2B1	mouse	1/500
Santa Cruz	sc-101136	hnRNP E2 (23-G)	PCBP2	mouse	1/200
Santa Cruz	sc-376416	hnRNP H3 (D-4)	HNRNPH3, HNRPH3	mouse	1/200
Santa Cruz	sc-32309	hnRNP F (3H4)	HNRNPF, HNRPF	mouse	1/200
Santa Cruz	sc-28380	hnRNP K (D-6)	HNRNPK, HNRPK	mouse	1/200
Santa Cruz	sc-390699	HNRPLL (A-4)	HNRNPLL, HNRPLL, SRRF, BLOCK24	mouse	1/200
Santa Cruz	sc-56703	hnRNP Q (I8E4)	SYNCRIP, HNRPQ, NSAP1	mouse	1/200
Santa Cruz	sc 393975	hnRNP UL1 (C-2)	HNRNPUL1 E1BAP5, HNRPUL1	mouse	1/200
Cell Signaling	#4196	LIN28B	LIN28B, CSDD2	rabbit	1/1000
Santa Cruz	sc-373827	LGP2 (C-9)	DHX58, D11LGP2E, LGP2	mouse	1/200
Santa Cruz	sc-166178	LRP130 (F-7)	LRPPRC, LRP130	mouse	1/200
Santa Cruz	sc47740	MBNL1 (3A4)	MBNL1, EXP, KIAA0428, MBNL	mouse	1/500
Abcam	ab186012	METTL16	METTL16 METT10D	rabbit	1/1000
Santa Cruz	sc-517212	Msi2 (2C11G4)	MSI2	mouse	1/500
Santa Cruz	sc-365283	NF45 (H-4)	ILF2, NF45, PRO3063	mouse	1/100
Santa Cruz	sc-377406	NF90 (A-3)	ILF3, DRBF, MPHOSPH4, NF90	mouse	1/100
Santa Cruz	sc-365164	p68 RNA Helicase (D-7)	DDX5, G17P1, HELR, HLR1	mouse	1/500
Santa Cruz	sc-8007	PARP-1 (F-2)	PARP1, ADPRT, PPOL	mouse	1/200
Santa Cruz	sc-376316	PTBP-2 (A-10)	PTBP2, NPTB, PTB, PTBLP	mouse	1/200
Santa Cruz	sc-271479	PNPase (D-1)	PNPT1, PNPASE	mouse	1/200
Santa Cruz	sc-390043	PUS1 (A-4)	PUS1, PP8985	mouse	1/200
Santa Cruz	sc 517305	QKI (N147/6)	QKI, HKQ	mouse	1/200
Santa Cruz	sc-393594	Rent1 (C-6)	UPF1, KIAA0221, RENT1	mouse	1/200
Santa Cruz	sc-515495	RBM45 (A-2)	RBM45, DRB1, DRBP1	mouse	1/200
Santa Cruz	sc-101190	RBMS1 (73-K2)	RBMS1, C2orf12, MSSP, MSSP1, SCR2	mouse	1/200
Santa Cruz	sc-517221	RING1 (8C12F4)	RING1, RNF1	mouse	1/200
Santa Cruz	sc-376326	RNase H1 (H-4)	RNASEH1, RNH1	mouse	1/200
Abcam	ab10685	ROA1	HNRNPA1, HNRPA1	mouse	1/200
Santa Cruz	sc-398105	ROD1 (C-1)	PTBP3, ROD1	mouse	1/200
Santa Cruz	sc-393804	U2 SnRNP A (B-3)	SNRPA1	mouse	1/200
Santa Cruz	sc-514508	SLIRP (B-12)	SLIRP, C14orf156, DC23, DC50, PD04872	mouse	1/200
Santa Cruz	sc-398881	splicing factor 1 (H-9)	SRSF1, ASF, SF2, SF2P33, SFRS1, OK/SW-cl.3	mouse	1/200
Santa Cruz	sc-393016	SF4 (B-3)	SUGP1, SF4	mouse	1/200
Santa Cruz	sc-377012	SRP14 (B-3)	SRP14	mouse	1/200
Santa Cruz	sc-52957	TBK1 (108A429)	TBK1, NAK	mouse	1/200
Santa Cruz	sc-166247	TIA-1 (G-3)	TIA1	mouse	1/200
Sigma-Aldrich	T6557	γ-Tubulin	TUBA1A, TUBA3	mouse	1/10000
Santa Cruz	sc-398340	YB-1 (C-3)	YBX1, NSEP1, YB1	mouse	1/200
Santa Cruz	sc-365743	ZFR (B-8)	ZFR	mouse	1/100

VI.E.3 Silver nitrate staining

Protein sample are mixed volume to volume with 126 mM Tris-HCL pH 6.8; 0.2 % 2-mercaptoethanol; 20 % glycerol; 4 %SDS; 0.004 % bromophenol blue (2 X Laemmli) buffer at 95°C during 5 min to completely denature proteins prior to SDS-PAGE loading. After migration, gel is fixed with 40 % EtOH; 10 % acetic acid solution during at least 1 h prior to silver nitrate staining. Silver nitrate staining is performed with the SilverQuest Silver staining (Thermo Fisher) kit following provider's instruction. Due to the sensibility of the silver staining method, it's important to load comparable samples within the same gel staining. For example, input samples will reach staining saturation long before elution's samples.

VI.F. RNA analysis

VI.F.1 RNA extraction and quantification

Total RNA is extracted using Tri-Reagent (Trizol, Sigma) according manufacturer's recommendations. RNA quantification is performed after a freezing/thawing cycle using optic density at 260 nm with the Nanodrop2000 spectrophotometer.

VI.F.2 Northern blot

Northern blotting is performed on either 5 µg or 15 µg of total RNA for respectively infected lymphocytes and non-infected adherent cells. miRNAs are resolved on a 12 % acrylamide; TBE 1 X; 8 M urea gel. Chimera baits are resolved on a 17.5 % acrylamide; TBE 1 X; 8 M urea gel of 20 cm in length. Ethidium bromide staining is systematically done before and after RNA transfer onto Hybond-NX membrane (GE Healthcare) within TBE 0.5 X buffer to assess the transfer efficacy. Blotted miRNAs are then chemically cross-linked to the membrane with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, Thermo Fisher) during 90 min at 65°C. Chimera baits are then cross-linked to the membrane with UV autocrosslink settings of Stratalinker UV Crosslinker (Stratagene) apparatus. Membranes are prehybridized for at least 1 h in PerfectHyb (Sigma) at 50°C in a rotatory oven. DNA probes are phosphate^{32} labelled with T4 polynucleotide kinase (Fermentas), 25 µCi of [γ-32P]dATP followed by a purification step consisting of a size exclusion chromatography using illustra MicroSpin G-25 columns (GE healthcare) (Table 15). Each probe is hybridized overnight at 50°C to the blot with PerfectHyb (Sigma). Blots are washed twice at 50°C for 20 min with 5 X SSC; 0.1 % SDS buffer. Finally, Northen blots are exposed to phosphor-imager plates and scanned using Typhoon FLA 7000 (GE Healthcare).

Table 15. Northern blot probes.

NB probe	
kshv-miR-K12-1-5p	GCTTACACCCAGTTTCCTGTAAT
kshv-miR-K12-2-5p	CAGATCGACCCGGACTACAGTT
kshv-miR-K12-3-5p	TCGCTGCCGTCCTCAGAATGTGA
kshv-miR-K12-4-3p	TCAGCTAGGCCTCAGTATTCTA
kshv-miR-K12-5-5p	CCGGCAAGTTCCAGGCATCCTA
kshv-miR-K12-6-3p	CTCAACAGCCCGAAAACCATCA
kshv-miR-K12-7-5p	AGCGCCAGCAACATGGGATCA
kshv-miR-K12-8-5p	CGTGCTCTCTCAGTCGCGCCTA
kshv-miR-K12-9-5p	TTACGCAGCTGCGTATACCCAG
kshv-miR-K12-11-5p	TCGGACACAGGCTAAGCATTAA
hsa-let-7a-1-5p	AACTATACAACCTACTACCTCA
hsa-miR-155-5p	ACCCCTATCACGATTAGCATTAA
U6 snRNA	GCAGGGGCCATGCTAATCTTCTCTGTATCG
tRNA-Val(CAC)	TCGAACCGGGGACCTTTCGCGTGT

VI.F.3 Reverse transcriptase and quantitative PCR

Prior to reverse transcriptase reaction, 2 µg of total RNA is treated 2 times with DNase I (Fermentas) during 15 min at 37°C within 10 mM Tris-HCl pH 7.5; 2.5 mM $MgCl_2$; 0.1 mM $CaCl_2$ buffer. Reverse transcription is performed on 400 ng of DNase treated total RNA with random nonamer primer (list primer) using SuperScript IV reverse transcriptase kit (Invitrogen) following provider's protocol. Resulting cDNA (1 µL of a 1/10 dilution) was PCR amplified with Maxima SYBR green qPCR Master Mix (2 X) kit (Fermentas) in 10 µL reaction volume with 0.3 µM final concentration of specific primers (Table 16, Table 17). The quantitative real-time PCR analysis is performed with the CFX96 Touch Real-Time PCR detection system (Bio-Rad). PCR program consist of an initial denaturation and hot-start enzyme activation step at 95°C for 10 min followed by 45 cycles at 95°C denaturation for 15 s, 60°C annealing for 30 s, 72°C elongation for 30 s including subsequent melting curve analysis. The data is analysed using the ΔΔCt method using either GAPDH, HPRT, Actin or 5'K10/12-cluster as a reference normalizer, and the overexpressed GFP or shRNA-scramble sample as a calibrator.

Table 16. qPCR KSHV primers.

qPCR KSHV primer		
kshv-primiR-K12-1	Forward	GGTTGTTTACGCAGGGTGC
	Reverse	GCGGATGTGTTTTGCCCAAA
kshv-primiR-K12-2	Forward	AGTCCGTCAACCAGATTTAAGA
	Reverse	TCTATAATTGGTGTCACCGCG
kshv-primiR-K12-4	Forward	GTGGAACCGGGCAGTATAAC
	Reverse	AGAGGCGTAGACATCCGTGT
kshv-primiR-K12-6	Forward	TGCGCGGGTTTAGAAAGACT
	Reverse	ACAGAACAATAACGGGCGACT
kshv-primiR-K12-9	Forward	CGCAGCCCCTATTCCAGTAG
	Reverse	CTGGATTTACCCAGCCGG
kshv-primiR-K12-11	Forward	CTACCTCCACGCTCGCGTAT
	Reverse	AAAATTGCCGCCGTGAAG
kshv-5'-K10/12-Cluster	Forward	GTCGTTTCGGTAGATGGGGG
	Reverse	TGACTCACTAGCTCCCCTCC
hsa-primiR-16	Forward	GCTCTTATGATAGCAATGTCAGCA
	Reverse	ACAACTGTAGAGTATGGTCAACCT
hsa-primiR-155	Forward	ATGCCTCATCCTCTGAGTGC
	Reverse	GCAGCAATTTGTTCCATGTG
hsa-primiR-let-7a-1	Forward	TTTCACCATTCACCCTGGAT
	Reverse	TTTCTATCAGACCGCCTGGA

Table 17. qPCR mRNA primers.

qPCR mRNA primer		
18S rRNA	Forward	CTCTTAGCTGAGTGTCCCGC
	Reverse	CTTAATCATGGCCTCAGTTCCGA
Actin	Forward	AACCCAGCCACACCACAAAG
	Reverse	CACTGACTTGAGACCAGTTGAATAAAA
CIRBP	Forward	CGGGTCCTACAGAGACAGTTAT
	Reverse	GCGAAGCTCCCACAATCTTTA
ELAV1	Forward	CCCTGCTGGATCAGAGAAATAG
	Reverse	CAATGCCTGGTGGACAAATG
FUS	Forward	CAGTTATGTCAGTGGTGGGTAG
	Reverse	CTGGAAGGAAGACATGCCTAAA
GAPDH	Forward	CTTTGGTATCGTGGAAGGACT
	Reverse	CCAGTGAGCTTCCCGTTCAG
HNRPD	Forward	ACAGGTGGTGAAGCAGTATTT
	Reverse	ACTTCCATTCAGGGACTTGATAC
HNRPF	Forward	GAAGCGTTCGTGCAGTTTG
	Reverse	CTCAATGTACCTGTGCCCTATC
HPRT	Forward	TGACCTTGATTTATTTTGCATACC
	Reverse	CGAGCAAGACGTTCAGTCCT
K1 mRNA	Forward	CCAAACGGACGAAATGAAAC
	Reverse	TGTGTGGTTGCATCGCTATT
LN28B	Forward	GGATTTGGATTCATCTCCATGATAA
	Reverse	GAATTCCACTGGTTCTCCTTCTTTT
MET16	Forward	ACAGAAGACACTCCTGATGG
	Reverse	TTAACAGAACTAGGCGGAGG
MSI2H	Forward	GGTCATGAGAGATCCCACTACG
	Reverse	TCTACACTTGCTGGGTCTGC
QKI	Forward	CGATATGCGTGTCCATCCTTAC
	Reverse	CAGGTCATCATTGGGTGAAGAG
RBM45	Forward	GGTTTGGGCTACGTACGATACT
	Reverse	GGTTCAGCCAAGATTGCTCT
ROA1	Forward	CGAAGGTAGGCTGGCAGATA
	Reverse	TTAGACATGACGGCAGGGTG
RTA	Forward	CCCAAACGAAAGCAGAGAAG
	Reverse	GGTGCAGCTGGTACAGTGTG
v-Cyclin	Forward	TAATAGAGGCGGGCAATGAG
	Reverse	ACTGGGCACATGGATGTTTT
v-IL6	Forward	CGGTTCACTGCTGGTATCTG
	Reverse	CAGTATCGTTGATGGCTGGT
ZNF346	Forward	CTAAACTCTCTGCCTCCACAAA
	Reverse	CGTCCTCAATTCTGGGTACATC

VI.G. Western blot and northern blot signal quantification

Antibody or probe signals of Western blot and Northern blot analysis are quantified by ImageQuant TL 1D v.8.1 analysis software (GE Healthcare). Target signals are normalized using the tubulin signal or tRNA-Val for the Western blot or the northern blot respectively.

VI.H. Luciferase assays

Adherent cells are seeded in 48-well plates at ~4.10^4 cells/well. The day after, transfection of siRNA, plasmid expressing candidate, or plasmid expressing Drosha/DGCR8 are performed with lipofectamine2000 (Thermo Fisher). Following day, 150 ng of reporter plasmid pmiRGLO, or 25 ng of reporter plasmid psiCHECK-2 are transfected with lipofectamine2000 (Thermo Fisher). 24 h later, cells are washed with PBS (Gibco) and lysed with 100 µL of a 1 X passive lysis buffer (Promega) compatible with the downstream luciferase assays following provider's instructions. 5 µL of the cellular extract are assayed for firefly and Renilla luciferase activity using the Dual-Glo Luciferase assay system (Promega) and the microplate reader

SAFAS Xenius. Firefly luciferase signal is produced after adding 50 µL of the first substrat (Dual-Glo). Firefly signal reaches a signal plateau 4 s after substrate injection. The average signal of the firefly plateau between 4 and 5 s is calculated. After obtaining the firefly signal, 50 µL of the second substrate (Dual-Glo Stop & Glo) is added to the well to assess the Renilla luciferase signal. Renilla signal reaches a signal plateau 7 s after substrate injection. The average signal of the Renilla plateau between 7 and 8 s is calculated. Transfection efficacy can be normalized by dividing the firefly luciferase signal (reporter possessing the SL-miRNA or the seed match in its 3' UTR) by the renilla luciferase signal (firefly signal / renilla signal). Then, the experimental condition is compared to the control condition.

VII. BIBLIOGRAPHY

Abdi, H. and Williams, L.J. (2010) Principal component analysis: Principal component analysis. *Wiley Interdiscip Rev Comput Stat* **2**: 433–459.

Abend, J.R., Ramalingam, D., Kieffer-Kwon, P., Uldrick, T.S., Yarchoan, R., and Ziegelbauer, J.M. (2012) Kaposi's sarcoma-associated herpesvirus microRNAs target IRAK1 and MYD88, two components of the toll-like receptor/interleukin-1R signaling cascade, to reduce inflammatory-cytokine expression. *J Virol* **86**: 11663–11674.

Abend, J.R., Uldrick, T., and Ziegelbauer, J.M. (2010) Regulation of tumor necrosis factor-like weak inducer of apoptosis receptor protein (TWEAKR) expression by Kaposi's sarcoma-associated herpesvirus microRNA prevents TWEAK-induced apoptosis and inflammatory cytokine expression. *J Virol* **84**: 12139–12151.

Alarcón, C.R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S.F. (2015) N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**: 482–485.

Alberti, C., Manzenreither, R.A., Sowemimo, I., Burkard, T.R., Wang, J., Mahofsky, K., et al. (2018) Cell-type specific sequencing of microRNAs from complex animal tissues. *Nat Methods* **15**: 283–289.

Alberti, S. (2017) Phase separation in biology. *Curr Biol CB* **27**: R1097–R1102.

Altuvia, Y. (2005) Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res* **33**: 2697–2706.

Ambros, V. (2003) A uniform system for microRNA annotation. *RNA* **9**: 277–279.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010) Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**: 1534–9.

An, Z., Zhai, L., Ying, W., Qian, X., Gong, F., Tan, M., and Fu, Y. (2019) PTMiner: Localization and Quality Control of Protein Modifications Detected in an Open Search and Its Application to Comprehensive Post-translational Modification Characterization in Human Proteome. *Mol Cell Proteomics MCP* **18**: 391–405.

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013) Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell* **152**: 844–858.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* **22**: 2773–2785.

Balacco, D.L. and Soller, M. The m6A writer: Rise of a machine for growing tasks. 58.

Baldrich, P., Hsing, Y.-I.C., and San Segundo, B. (2016) Genome-Wide Analysis of Polycistronic MicroRNAs in Cultivated and Wild Rice. *Genome Biol Evol* **8**: 1104–1114.

Ballon, G., Chen, K., Perez, R., Tam, W., and Cesarman, E. (2011) Kaposi sarcoma herpesvirus (KSHV) vFLIP oncoprotein induces B cell transdifferentiation and tumorigenesis in mice. *J Clin Invest* **121**: 1141–1153.

Bartel, D.P. (2018) Metazoan MicroRNAs. *Cell* **173**: 20–51.

Bartel, D.P. (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**: 215–233.

Barton, E.S., White, D.W., Cathelyn, J.S., Brett-McClellan, K.A., Engle, M., Diamond, M.S., et al. (2007) Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature* **447**: 326–329.

Bauer, N.C., Doetsch, P.W., and Corbett, A.H. (2015) Mechanisms Regulating Protein Localization: Mechanisms Regulating Protein Localization. *Traffic* **16**: 1039–1061.

Becker, W.R., Ober-Reynolds, B., Jouravleva, K., Jolly, S.M., Zamore, P.D., and Greenleaf, W.J. (2019) High-Throughput Analysis Reveals Rules for Target RNA Binding and Cleavage by AGO2. *Mol Cell*.

Bellare, P. and Ganem, D. (2009) Regulation of KSHV lytic switch protein expression by a virus-encoded microRNA: an evolutionary adaptation that fine-tunes lytic reactivation. *Cell Host Microbe* **6**: 570–575.

Berlivet, S., Scutenaire, J., Deragon, J.-M., and Bousquet-Antonelli, C. (2019) Readers of the m(6)A epitranscriptomic code. *Biochim Biophys Acta Gene Regul Mech* **1862**: 329–342.

Bhardwaj, A., Myers, M.P., Buratti, E., and Baralle, F.E. (2013) Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Res* **41**: 5062–5074.

Bo, T.H. (2004) LSimpute: accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Res* **32**: 34e–334.

Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlers, M., and Cullen, B.R. (2010) A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. *Mol Cell* **37**: 135–142.

Boss, I.W., Nadeau, P.E., Abbott, J.R., Yang, Y., Mergia, A., and Renne, R. (2011) A Kaposi's sarcoma-associated herpesvirus-encoded ortholog of microRNA miR-155 induces human splenic B-cell expansion in NOD/LtSz-scid IL2Rgammanull mice. *J Virol* **85**: 9877–9886.

Bou-Nader, C., Gordon, J.M., Henderson, F.E., and Zhang, J. (2019) The search for a PKR code-differential regulation of protein kinase R activity by diverse RNA and protein regulators. *RNA N Y N* **25**: 539–556.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005) Principles of MicroRNA–Target Recognition. *PLoS Biol* **3**: e85.

Bronesky, D., Desgranges, E., Corvaglia, A., Francois, P., Caballero, C.J., Prado, L., et al. (2019) A multifaceted small RNA modulates gene expression upon glucose limitation in Staphylococcus aureus. *EMBO J* **38**:.

Buchan, J.R. (2014) mRNP granules. Assembly, function, and connections with disease. *RNA Biol* **11**: 1019–1030.

Burge, R.G., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2014) Structural characterization of interactions between the double-stranded. *Biochemistry* **53**: 1495–1510.

Burger, T. (2018) Gentle Introduction to the Statistical Foundations of False Discovery Rate in Quantitative Proteomics. *J Proteome Res* **17**: 12–22.

Burke, J.M., Bass, C.R., Kincaid, R.P., and Sullivan, C.S. (2014) Identification of tri-phosphatase activity in the biogenesis of retroviral microRNAs and RNAP III-generated shRNAs. *Nucleic Acids Res* **42**: 13949–13962.

Busch, B., Bley, N., Müller, S., Glaß, M., Misiak, D., Lederer, M., et al. (2016) The oncogenic triangle of HMGA2, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. *Nucleic Acids Res* **44**: 3845–3864.

Cai, X. and Cullen, B.R. (2006) Transcriptional origin of Kaposi's sarcoma-associated herpesvirus microRNAs. *J Virol* **80**: 2234–2242.

Calabretta, S. and Richard, S. (2015) Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends Biochem Sci* **40**: 662–672.

Carapito, C., Burel, A., Guterl, P., Walter, A., Varrier, F., Bertile, F., and Van Dorsselaer, A. (2014) MSDA, a proteomics software suite for in-depth Mass Spectrometry Data Analysis using grid computing. *Proteomics* **14**: 1014–1019.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., et al. (2016) Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**: 696–710.

Catala, R., Carrasco-Lopez, C., Perea-Resa, C., Hernandez-Verdeja, T., and Salinas, J. (2019) Emerging Roles of LSM Complexes in Posttranscriptional Regulation of Plant Response to Abiotic Stress. *Front Plant Sci* **10**: 167.

Caudron-Herger, M., Rusin, S.F., Adamo, M.E., Seiler, J., Schmid, V.K., Barreau, E., et al. (2019) R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. *Mol Cell* **75**: 184-199.e10.

Cazalla, D., Xie, M., and Steitz, J.A. (2011) A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol Cell* **43**: 982–992.

Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J., Knowles, D.M., and Moore, P.S. (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**: 1865–1869.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E., Veldhoen, R.A., et al. (2011) Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biol* **8**: 1105–1114.

Chavez-Calvillo, G., Martin, S., Hamm, C., and Sztuba-Solinska, J. (2018) The Structure-To-Function Relationships of Gammaherpesvirus-Encoded Long Non-Coding RNAs and Their Contributions to Viral Pathogenesis. *Non-Coding RNA* **4**: 24.

Chawla, G. and Sokol, N.S. (2014) ADAR mediates differential expression of polycistronic microRNAs. *Nucleic Acids Res* **42**: 5245–5255.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**: 584–589.

Chen, C.-Y.A. and Shyu, A.-B. (2011) Mechanisms of deadenylation-dependent decay: Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* **2**: 167–183.

Chen, J., Fang, X., Zhong, P., Song, Z., and Hu, X. (2019) N6-methyladenosine modifications: interactions with novel RNA-binding proteins and roles in signal transduction. *RNA Biol* **16**: 991–1000.

Chen, T., Brownawell, A.M., and Macara, I.G. (2004) Nucleocytoplasmic shuttling of JAZ, a new cargo protein for exportin-5. *Mol Cell Biol* **24**: 6608–6619.

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., and Smyth, G. edgeR: differential expression analysis of digital gene expression data User's Guide. 111.

Choudhury, N.R., de Lima Alves, F., de Andres-Aguayo, L., Graf, T., Caceres, J.F., Rappsilber, J., and Michlewski, G. (2013) Tissue-specific control of brain-enriched miR-7 biogenesis. *Genes Dev* **27**: 24–38.

Chung, W.-J., Agius, P., Westholm, J.O., Chen, M., Okamura, K., Robine, N., et al. (2011) Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. *Genome Res* **21**: 286–300.

Clemens, M.J. (1997) PKR-a protein kinase regulated by double-stranded RNA. *Int J Biochem Cell Biol* **29**: 945–9.

Cléry, A. (2011) From Structure to Function of RNA Binding Domains. 23.

Clery, A., Sohier, T.J.M., Welte, T., Langer, A., and Allain, F.H.T. (2017) switchSENSE: A new technology to study protein-RNA interactions. *Methods San Diego Calif* **118–119**: 137–145.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., et al. (2015) The RNA Binding Protein Quaking Regulates Formation of circRNAs. *Cell* **160**: 1125–1134.

Conrad, N.K. and Steitz, J.A. (2005) A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. *EMBO J* **24**: 1831–1841.

Contrant, M., Fender, A., Chane-Woon-Ming, B., Randrianjafy, R., Vivet-Boudou, V., Richer, D., and Pfeffer, S. (2014) Importance of the RNA secondary structure for the relative accumulation of clustered viral microRNAs. *Nucleic Acids Res* **42**: 7981–7996.

Cook, K.C. and Cristea, I.M. (2019) Location is everything: protein translocations as a viral infection strategy. *Curr Opin Chem Biol* **48**: 34–43.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**: 2513–2526.

Damianov, A., Ying, Y., Lin, C.-H., Lee, J.-A., Tran, D., Vashisht, A.A., et al. (2016) Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* **165**: 606–619.

Dang, W., Xie, Y., Cao, P., Xin, S., Wang, J., Li, S., et al. (2019) N(6)-Methyladenosine and Viral Infection. *Front Microbiol* **10**: 417.

Das, U., Nguyen, H., and Xie, J. (2019) Transcriptome protection by the expanded family of hnRNPs. *RNA Biol* **16**: 155–159.

Dellaire, G., Eskiw, C.H., Dehghani, H., Ching, R.W., and Bazett-Jones, D.P. (2006) Mitotic accumulations of PML protein contribute to the re-establishment of PML nuclear bodies in G1. *J Cell Sci* **119**: 1034–1042.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014) Assessing the ceRNA Hypothesis with Quantitative Measurements of miRNA and Target Abundance. *Mol Cell* **54**: 766–776.

Denzler, R., McGeary, S.E., Title, A.C., Agarwal, V., Bartel, D.P., and Stoffel, M. (2016) Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Mol Cell* **64**: 565–579.

Dölken, L., Malterer, G., Erhard, F., Kothe, S., Friedel, C.C., Suffert, G., et al. (2010) Systematic analysis of viral and cellular microRNA targets in cells latently infected with human gamma-herpesviruses by RISC immunoprecipitation assay. *Cell Host Microbe* **7**: 324–34.

Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., et al. (2018) Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* **70**: 854-867.e9.

Downie Ruiz Velasco, A., Welten, S.M.J., Goossens, E.A.C., Quax, P.H.A., Rappsilber, J., Michlewski, G., and Nossent, A.Y. (2019) Posttranscriptional Regulation of 14q32 MicroRNAs by the CIRBP and HADHB during Vascular Regeneration after Ischemia. *Mol Ther Nucleic Acids* **14**: 329–338.

Doxtader, K.A., Wang, P., Scarborough, A.M., Seo, D., Conrad, N.K., and Nam, Y. (2018) Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. *Mol Cell* **71**: 1001-1011.e4.

Du, J., Wang, Q., Ziegler, S.F., and Zhou, B. (2018) FOXP3 interacts with hnRNPF to modulate pre-mRNA alternative splicing. *J Biol Chem* **293**: 10235–10244.

Ebert, M.S. and Sharp, P.A. (2012) Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell* **149**: 515–524.

Ederle, H. and Dormann, D. (2017) TDP-43 and FUS en route from the nucleus to the cytoplasm. *FEBS Lett* **591**: 1489–1507.

Ender, C., Krek, A., Friedlander, M.R., Beitzinger, M., Weinmann, L., Chen, W., et al. (2008) A human snoRNA with microRNA-like functions. *Mol Cell* **32**: 519–528.

Esteller, M. (2011) Non-coding RNAs in human disease. *Nat Rev Genet* **12**: 861–874.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* **15**: 346–353.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., et al. (2011) miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nat Struct Mol Biol* **18**: 1211–1217.

Farh, K.K.-H. (2005) The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. *Science* **310**: 1817–1821.

Feldman, E.R., Kara, M., Coleman, C.B., Grau, K.R., Oko, L.M., Krueger, B.J., et al. (2014) Virus-encoded microRNAs facilitate gammaherpesvirus latency and pathogenesis in vivo. *mBio* **5**: e00981-00914.

Fields, B.N., Knipe, D.M., and Howley, P.M. (2013) Fields virology, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Flynt, A.S., Greimann, J.C., Chung, W.-J., Lima, C.D., and Lai, E.C. (2010) MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. *Mol Cell* **38**: 900–907.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2008) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92–105.

Fromm, B., Billipp, T., Peck, L.E., Johansen, M., Tarver, J.E., King, B.L., et al. (2015) A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annu Rev Genet* **49**: 213–242.

Frye, M., Harada, B.T., Behm, M., and He, C. (2018) RNA modifications modulate gene expression during development. *Science* **361**: 1346–1349.

Fuziwara, C.S. and Kimura, E.T. (2015) Insights into Regulation of the miR-17-92 Cluster of miRNAs in Cancer. *Front Med* **2**:.

Ganser, L.R., Kelly, M.L., Herschlag, D., and Al-Hashimi, H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* **20**: 474–489.

Gay, L.A., Sethuraman, S., Thomas, M., Turner, P.C., and Renne, R. (2018) Modified Cross-Linking, Ligation, and Sequencing of Hybrids (qCLASH) Identifies Kaposi's Sarcoma-Associated Herpesvirus MicroRNA Targets in Endothelial Cells. *J Virol* **92**: e02138-17.

Girardi, E., López, P., and Pfeffer, S. (2018) On the Importance of Host MicroRNAs During Viral Infection. *Front Genet* **9**:.

Goncalves, P.H., Ziegelbauer, J., Uldrick, T.S., and Yarchoan, R. (2017) Kaposi sarcoma herpesvirus-associated cancers and related diseases. *Curr Opin HIV AIDS* **12**: 47–56.

Gong, J., Huang, M., Wang, F., Ma, X., Liu, H., Tu, Y., et al. (2017) RBM45 competes with HDAC1 for binding to FUS in response to DNA damage. *Nucleic Acids Res* **45**: 12862–12876.

Gottwein, E. (2012) Kaposi's Sarcoma-Associated Herpesvirus microRNAs. *Front Microbiol* **3**:.

Gottwein, E. and Cullen, B.R. (2010) A human herpesvirus microRNA inhibits p21 expression and attenuates p21-mediated cell cycle arrest. *J Virol* **84**: 5229–5237.

Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W.H., Chi, J.-T.A., et al. (2007) A viral microRNA functions as an orthologue of cellular miR-155. *Nature* **450**: 1096–1099.

Gregory, R.I., Yan, K., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235–240.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007) MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol Cell* **27**: 91–105.

Guil, S. and Caceres, J.F. (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol* **14**: 591–6.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835–840.

Guo, Y.E., Riley, K.J., Iwasaki, A., and Steitz, J.A. (2014) Alternative capture of noncoding RNAs or protein-coding genes by herpesviruses to alter host T cell function. *Mol Cell* **54**: 67–79.

Ha, M. and Kim, V.N. (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**: 509–524.

Haar, J., Contrant, M., Bernhardt, K., Feederle, R., Diederichs, S., Pfeffer, S., and Delecluse, H.-J. (2016) The expression of a viral microRNA is regulated by clustering to allow optimal B cell transformation. *Nucleic Acids Res* **44**: 1326–1341.

Haas, G., Cetin, S., Messmer, M., Chane-Woon-Ming, B., Terenzi, O., Chicher, J., et al. (2016) Identification of factors involved in target RNA-directed microRNA degradation. *Nucleic Acids Res* **44**: 2873–2887.

Haecker, I. and Renne, R. (2014) HITS-CLIP and PAR-CLIP advance viral miRNA targetome analysis. *Crit Rev Eukaryot Gene Expr* **24**: 101–116.

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., et al. (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**: 1697–1712.

Hage, D.S. and Matsuda, R. (2015) Affinity chromatography: a historical perspective. *Methods Mol Biol Clifton NJ* **1286**: 1–19.

Hansen, A., Henderson, S., Lagos, D., Nikitenko, L., Coulter, E., Roberts, S., et al. (2010) KSHV-encoded miRNAs target MAF to induce endothelial cell reprogramming. *Genes Dev* **24**: 195–205.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., and Kay, M.A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA N Y N* **16**: 673–695.

Hayakawa, S., Shiratori, S., Yamato, H., Kameyama, T., Kitatsuji, C., Kashigi, F., et al. (2011) ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nat Immunol* **12**: 37–44.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**: 327–341.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008) Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA. *Mol Cell* **32**: 276–284.

Herbert, A. (2019) ADAR and Immune Silencing in Cancer. *Trends Cancer* **5**: 272–282.

Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., et al. (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics* 15.

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., et al. (2017) Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**: 1198–1204.

Ho, C.K., Wang, L.K., Lima, C.D., and Shuman, S. (2004) Structure and Mechanism of RNA Ligase. *Structure* **12**: 327–339.

Hoshina, S., Sekizuka, T., Kataoka, M., Hasegawa, H., Hamada, H., Kuroda, M., and Katano, H. (2016) Profile of Exosomal and Intracellular microRNA in Gamma-Herpesvirus-Infected Lymphoma Cell Lines. *PloS One* **11**: e0162574.

Huang, H., Zhang, J., Harvey, S.E., Hu, X., and Cheng, C. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the. *Genes Dev* **31**: 2296–2309.

Hussein, H.A.M., Alfhili, M.A., Pakala, P., Simon, S., Hussain, J., McCubrey, J.A., and Akula, S.M. (2019) miRNAs and their roles in KSHV pathogenesis. *Virus Res* **266**: 15–24.

Ipsaro, J.J. and Joshua-Tor, L. (2015) From Guide to Target: Molecular Insights into Eukaryotic RNAi Machinery. 23.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016) ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**: 487–498.

Jankowsky, E. and Harris, M.E. (2015) Specificity and nonspecificity in RNA–protein interactions. *Nat Rev Mol Cell Biol* **16**: 533–544.

Järvelin, A.I., Noerenberg, M., Davis, I., and Castello, A. (2016) The new (dis)order in RNA regulation. *Cell Commun Signal CCS* **14**: 9–9.

Jeffery, C.J. (2015) Why study moonlighting proteins? *Front Genet* **6**:.

Jenner, R.G., Alba, M.M., Boshoff, C., and Kellam, P. (2001) Kaposi's sarcoma-associated herpesvirus latent and lytic gene expression as revealed by DNA arrays. *J Virol* **75**: 891–902.

Jiang, L., Shao, C., Wu, Q.-J., Chen, G., Zhou, J., Yang, B., et al. (2017) NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nat Struct Mol Biol* **24**: 816–824.

Jonas, S. and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**: 421–433.

Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., and Sarnow, P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**: 1577–81.

Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA N Y N* **8**: 426–439.

Juzenas, S., Venkatesh, G., Hübenthal, M., Hoeppner, M.P., Du, Z.G., Paulsen, M., et al. (2017) A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res* **45**: 9290–9301.

Kawai, S. and Amano, A. (2012) BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. *J Cell Biol* **197**: 201–208.

Kent, O.A., Mendell, J.T., and Rottapel, R. (2016) Transcriptional Regulation of miR-31 by Oncogenic KRAS Mediates Metastatic Phenotypes by Repressing RASA1. *Mol Cancer Res MCR* **14**: 267–277.

Kharas, M.G. and Lengner, C.J. (2017) Stem Cells, Cancer, and MUSASHI in Blood and Guts. *Trends Cancer* **3**: 347–356.

Kincaid, R.P., Burke, J.M., and Sullivan, C.S. (2012) RNA virus microRNA that mimics a B-cell oncomiR. *Proc Natl Acad Sci U S A* **109**: 3077–3082.

Kincaid, R.P., Chen, Y., Cox, J.E., Rethwilm, A., and Sullivan, C.S. (2014) Noncanonical microRNA (miRNA) biogenesis gives rise to retroviral mimics of lymphoproliferative and immunosuppressive host miRNAs. *mBio* **5**: e00074.

Knight, S.W. and Bass, B.L. (2002) The role of RNA editing by ADARs in RNAi. *Mol Cell* **10**: 809–17.

Kooshapur, H., Choudhury, N.R., Simon, B., Muhlbauer, M., Jussupow, A., Fernandez, N., et al. (2018) Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. *Nat Commun* **9**: 2479.

Kotaki, R., Higuchi, H., Ogiya, D., Katahira, Y., Kurosaki, N., Yukihira, N., et al. (2017) Imbalanced expression of polycistronic miRNA in acute myeloid leukemia. *Int J Hematol* **106**: 811–819.

Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68-73.

Kudinov, A.E., Karanicolas, J., Golemis, E.A., and Boumber, Y. (2017) Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets. *Clin Cancer Res Off J Am Assoc Cancer Res* **23**: 2143–2153.

Kumar, S., Downie Ruiz Velasco, A., and Michlewski, G. (2017) Oleic Acid Induces MiR-7 Processing through Remodeling of Pri-MiR-7/Protein Complex. *J Mol Biol* **429**: 1638–1649.

Kundu, P., Fabian, M.R., Sonenberg, N., Bhattacharyya, S.N., and Filipowicz, W. (2012) HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Res* **40**: 5088–5100.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–8.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**: 1401–14.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858–62.

Lauressergues, D., Couzigou, J.-M., Clemente, H.S., Martinez, Y., Dunand, C., Bécard, G., and Combier, J.-P. (2015) Primary transcripts of microRNAs encode regulatory peptides. *Nature* **520**: 90–93.

Lee, K.-H., Kim, S.-H., Kim, H.-J., Kim, W., Lee, H.-R., Jung, Y., et al. (2014) AUF1 contributes to Cryptochrome1 mRNA degradation and rhythmic translation. *Nucleic Acids Res* **42**: 3590–3606.

Lee, R.C. (2001) An Extensive Class of Small RNAs in Caenorhabditis elegans. *Science* **294**: 862–864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–854.

Lei, X., Bai, Z., Ye, F., Xie, J., Kim, C.-G., Huang, Y., and Gao, S.-J. (2010) Regulation of NF-kappaB inhibitor IkappaBalpha and viral replication by a KSHV microRNA. *Nat Cell Biol* **12**: 193–199.

Lei, X., Zhu, Y., Jones, T., Bai, Z., Huang, Y., and Gao, S.-J. (2012) A Kaposi's sarcoma-associated herpesvirus microRNA and its variants target the transforming growth factor beta pathway to promote cell survival. *J Virol* **86**: 11698–11711.

Levengood, J.D. and Tolbert, B.S. (2019) Idiosyncrasies of hnRNP A1-RNA recognition: Can binding mode influence function. *Semin Cell Dev Biol* **86**: 150–161.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **120**: 15–20.

Li, C., Hu, J., Hao, J., Zhao, B., Wu, B., Sun, L., et al. (2014) Competitive virus and host RNAs: the interplay of a hidden virus and host interaction. *Protein Cell* **5**: 348–356.

Li, Q., Noel-MacDonnell, J.R., Koestler, D.C., Goode, E.L., and Fridley, B.L. (2018) Subject level clustering using a negative binomial model for small transcriptomic studies. *BMC Bioinformatics* **19**: 474.

Li, W., Jia, X., Shen, C., Zhang, M., Xu, J., Shang, Y., et al. (2016) A KSHV microRNA enhances viral latency and induces angiogenesis by targeting GRK2 to activate the CXCR2/AKT pathway. *Oncotarget* **7**: 32286–32305.

Li, W., Yan, Q., Ding, X., Shen, C., Hu, M., Zhu, Y., et al. (2016) The SH3BGR/STAT3 Pathway Regulates Cell Migration and Angiogenesis Induced by a Gammaherpesvirus MicroRNA. *PLOS Pathog* **12**: e1005605.

Li, Y., Collins, M., An, J., Geiser, R., Tegeler, T., Tsantilas, K., et al. (2016) Immunoprecipitation and mass spectrometry defines an extensive RBM45 protein-protein interaction network. *Brain Res* **1647**: 79–93.

Li, Y., Yamane, D., and Lemon, S.M. (2015) Dissecting the roles of the 5' exoribonucleases Xrn1 and Xrn2 in restricting hepatitis C virus replication. *J Virol* **89**: 4857–4865.

Liang, D., Gao, Y., Lin, X., He, Z., Zhao, Q., Deng, Q., and Lan, K. (2011) A human herpesvirus miRNA attenuates interferon signaling and contributes to maintenance of viral latency by targeting IKKε. *Cell Res* **21**: 793–806.

Lin, S. and Gregory, R.I. (2015) MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer* **15**: 321–333.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **10**: 93–95.

Liu, Y., Sun, R., Lin, X., Liang, D., Deng, Q., and Lan, K. (2012) Kaposi's sarcoma-associated herpesvirus-encoded microRNA miR-K12-11 attenuates transforming growth factor beta signaling through suppression of SMAD5. *J Virol* **86**: 1372–1381.

Louloupi, A. and Orom, U.A.V. (2018) Metabolic Pulse-Chase RNA Labeling for pri-miRNA Processing Dynamics. *Methods Mol Biol Clifton NJ* **1823**: 33–41.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017) Transcriptomics technologies. *PLoS Comput Biol* **13**: e1005457.

Lu, C.-C., Li, Z., Chu, C.-Y., Feng, Jiaying, Feng, Jun, Sun, R., and Rana, T.M. (2010) MicroRNAs encoded by Kaposi's sarcoma-associated herpesvirus regulate viral life cycle. *EMBO Rep* **11**: 784–790.

Lu, F., Stedman, W., Yousef, M., Renne, R., and Lieberman, P.M. (2010) Epigenetic regulation of Kaposi's sarcoma-associated herpesvirus latency by virus-encoded microRNAs that target Rta and the cellular Rbl2-DNMT pathway. *J Virol* **84**: 2697–2706.

Lujan, D.A., Ochoa, J.L., and Hartley, R.S. (2018) Cold-inducible RNA binding protein in cancer and inflammation. *Wiley Interdiscip Rev RNA* **9**:.

Lunde, B.M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8**: 479–490.

Marcinowski, L., Tanguy, M., Krmpotic, A., Rädle, B., Lisnić, V.J., Tuddenham, L., et al. (2012) Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. *PLoS Pathog* **8**: e1002510.

Marco, A. (2019) No evidence of functional co-adaptation between clustered microRNAs. *bioRxiv*.

Marco, A., Ninova, M., Ronshaugen, M., and Griffiths-Jones, S. (2013) Clusters of microRNAs emerge by new hairpins in existing transcripts. *Nucleic Acids Res* **41**: 7745–7752.

Martin, B., Coutard, B., Guez, T., Paesen, G.C., Canard, B., Debart, F., et al. (2018) The methyltransferase domain of the Sudan ebolavirus L protein specifically targets internal adenosines of RNA substrates, in addition to the cap structure. *Nucleic Acids Res* **46**: 7902–7912.

Martin, H.C., Wani, S., Steptoe, A.L., Krishnan, K., Nones, K., Nourbakhsh, E., et al. (2014) Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs. *Genome Biol* **15**: R51.

Massimelli, M.J., Majerciak, V., Kruhlak, M., and Zheng, Z.-M. (2013) Interplay between polyadenylate-binding protein 1 and Kaposi's sarcoma-associated herpesvirus ORF57 in accumulation of polyadenylated nuclear RNA, a viral long noncoding RNA. *J Virol* **87**: 243–256.

de la Mata, M., Gaidatzis, D., Vitanescu, M., Stadler, M.B., Wentzel, C., Scheiffele, P., et al. (2015) Potent degradation of neuronal miRNAs induced by highly complementary targets. *EMBO Rep* **16**: 500–511.

Mcdonald, J.H. HANDBOOK OF BIOLOGICAL STATISTICS. 317.

McHugh, M.L. (2011) Multiple comparison analysis testing in ANOVA. *Biochem Medica* **21**: 203–209.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**: 185–197.

Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**: 343–349.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333–338.

Mendel, M., Chen, K.-M., Homolka, D., Gos, P., Pandey, R.R., McCarthy, A.A., and Pillai, R.S. (2018) Methylation of Structured RNA by the m(6)A Writer METTL16 Is Essential for Mouse Embryonic Development. *Mol Cell* **71**: 986-1000.e11.

Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., et al. (2013) Birth and expression evolution of mammalian microRNA genes. *Genome Res* **23**: 34–45.

Meyer, A., Golbik, R.P., Sanger, L., Schmidt, T., Behrens, S.-E., and Friedrich, S. (2019) The RGG/RG motif of AUF1 isoform p45 is a key modulator of the protein's RNA chaperone and RNA annealing activities. *RNA Biol* **16**: 960–971.

Meyer, K.D. and Jaffrey, S.R. (2017) Rethinking m(6)A Readers, Writers, and Erasers. *Annu Rev Cell Dev Biol* **33**: 319–342.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**: 1635–1646.

Michlewski, G. and Cáceres, J.F. (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol* **17**: 1011–1018.

Michlewski, G. and Caceres, J.F. (2010) RNase-assisted RNA chromatography. *RNA* **16**: 1673–1678.

Michlewski, G., Guil, S., Semple, C.A., and Cáceres, J.F. (2008) Posttranscriptional Regulation of miRNAs Harboring Conserved Terminal Loops. *Mol Cell* **32**: 383–393.

Min, K.-W., Jo, M.H., Shin, S., Davila, S., Zealy, R.W., Kang, S.I., et al. (2017) AUF1 facilitates microRNA-mediated gene silencing. *Nucleic Acids Res* **45**: 6064–6073.

Mohammed, J., Siepel, A., and Lai, E.C. (2014) Diverse modes of evolutionary emergence and flux of conserved microRNA clusters. *RNA* **20**: 1850–1863.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., et al. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**: 123–133.

Moody, R., Zhu, Y., Huang, Y., Cui, X., Jones, T., Bedolla, R., et al. (2013) KSHV MicroRNAs Mediate Cellular Transformation and Tumorigenesis by Redundantly Targeting Cell Growth and Survival Pathways. *PLOS Pathog* **9**: e1003857.

Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008) Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* **15**: 902–909.

Morlando, M., Dini Modigliani, S., Torrelli, G., Rosa, A., Di Carlo, V., Caffarelli, E., and Bozzoni, I. (2012) FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. *EMBO J* **31**: 4502–4510.

Nachmani, D., Stern-Ginossar, N., Sarid, R., and Mandelboim, O. (2009) Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. *Cell Host Microbe* **5**: 376–385.

Noh, J.H., Kim, K.M., Abdelmohsen, K., Yoon, J.-H., Panda, A.C., Munk, R., et al. (2016) HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev* **30**: 1224–1239.

O'Hara, A.J., Chugh, P., Wang, L., Netto, E.M., Luz, E., Harrington, W.J., et al. (2009) Pre-micro RNA signatures delineate stages of endothelial cell transformation in Kaposi sarcoma. *PLoS Pathog* **5**: e1000389.

Ontiveros, R.J., Stoute, J., and Liu, K.F. (2019) The chemical diversity of RNA modifications. *Biochem J* **476**: 1227–1245.

Perez-Perri, J.I., Rogell, B., Schwarzl, T., Stein, F., Zhou, Y., Rettel, M., et al. (2018) Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat Commun* **9**: 4408.

Peters, D.T., Fung, H.K.H., Levdikov, V.M., Irmscher, T., Warrander, F.C., Greive, S.J., et al. (2016) Human Lin28 Forms a High-Affinity 1:1 Complex with the 106~363 Cluster miRNA miR-363. *Biochemistry* **55**: 5021–5027.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grässer, F.A., et al. (2005) Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**: 269–276.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., et al. (2004) Identification of virus-encoded microRNAs. *Science* **304**: 734–6.

Piedade, D. and Azevedo-Pereira, J.M. (2016) The Role of microRNAs in the Pathogenesis of Herpesvirus Infection. *Viruses* **8**: 156.

Ponzo, I., Moller, F.M., Daub, H., and Matscheko, N. (2019) A DNA-Based Biosensor Assay for the Kinetic Characterization of Ion-Dependent Aptamer Folding and Protein Binding. *Mol Basel Switz* **24**:.

Prongidi-Fix, L., Schaeffer, L., Simonetti, A., Barends, S., Ménétret, J.-F., Klaholz, B.P., et al. (2013) Rapid purification of ribosomal particles assembled on histone H4 mRNA: a new method based on mRNA–DNA chimaeras. *Biochem J* **449**: 719–728.

Purushothaman, P., Uppal, T., and Verma, S.C. (2015) Molecular biology of KSHV lytic reactivation. *Viruses* **7**: 116–153.

Rau, F., Freyermuth, F., Fugier, C., Villemin, J.-P., Fischer, M.-C., Jost, B., et al. (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat Struct Mol Biol* **18**: 840–845.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901–906.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002) Prediction of Plant MicroRNA Targets. *Cell* **110**: 513–520.

Ribeiro, D.M., Briere, G., Bely, B., Spinelli, L., and Brun, C. (2019) MoonDB 2.0: an updated database of extreme multifunctional and moonlighting proteins. *Nucleic Acids Res* **47**: D398–D402.

Ringeard, M., Marchand, V., Decroly, E., Motorin, Y., and Bennasser, Y. (2019) FTSJ3 is an RNA 2'-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature* **565**: 500–504.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47–e47.

Rizzuti, M., Filosa, G., Melzi, V., Calandriello, L., Dioni, L., Bollati, V., et al. (2018) MicroRNA expression analysis identifies a subset of downregulated miRNAs in ALS motor neuron progenitors. *Sci Rep* **8**: 10105.

Robb, G.B. and Rana, T.M. (2007) RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol Cell* **26**: 523–537.

Rodriguez, A. (2004) Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Res* **14**: 1902–1910.

Root, D.E., Hacohen, N., Hahn, W.C., Lander, E.S., and Sabatini, D.M. (2006) Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods* **3**: 715–719.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017) Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**: 1187–1200.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**: 83–86.

Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., et al. (2009) The NF90-NF45 Complex Functions as a Negative Regulator in the MicroRNA Processing Pathway. *Mol Cell Biol* **29**: 3754–3769.

Sakamoto, T., Ennifar, E., and Nakamura, Y. (2018) Thermodynamic study of aptamers binding to their target proteins. *Biochimie* **145**: 91–97.

Samols, M.A., Skalsky, R.L., Maldonado, A.M., Riva, A., Lopez, M.C., Baker, H.V., and Renne, R. (2007) Identification of cellular genes targeted by KSHV-encoded microRNAs. *PLoS Pathog* **3**: e65.

Schmidt, T., Friedrich, S., Golbik, R.P., and Behrens, S.-E. (2017) NF90-NF45 is a selective RNA chaperone that rearranges viral and cellular riboswitches: biochemical analysis of a virus host factor activity. *Nucleic Acids Res* **45**: 12441–12454.

Schürmann, N., Trabuco, L.G., Bender, C., Russell, R.B., and Grimm, D. (2013) Molecular dissection of human Argonaute proteins by DNA shuffling. *Nat Struct Mol Biol* **20**: 818–826.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., et al. (2014) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep* **8**: 284–296.

Sheu-Gruttadauria, J. and MacRae, I.J. (2018) Phase Transitions in the Assembly and Function of Human miRISC. *Cell* **173**: 946-957.e16.

Shin, Y. and Brangwynne, C.P. (2017) Liquid phase condensation in cell physiology and disease. *Science* **357**: eaaf4382.

Shun, Z., Silverberg, A., Chang, C.K., and Ouyang, P. (2003) Dunnett's many-to-one test and least square means. *J Biopharm Stat* **13**: 17–28.

Siira, S.J., Spahr, H., Shearwood, A.-M.J., Ruzzenente, B., Larsson, N.-G., Rackham, O., and Filipovska, A. (2017) LRPPRC-mediated folding of the mitochondrial transcriptome. *Nat Commun* **8**: 1532.

Sin, S.-H. and Dittmer, D.P. (2012) Cytokine homologs of human gammaherpesviruses. *J Interferon Cytokine Res Off J Int Soc Interferon Cytokine Res* **32**: 53–59.

Suffert, G., Malterer, G., Hausser, J., Viiliäinen, J., Fender, A., Contrant, M., et al. (2011) Kaposi's sarcoma herpesvirus microRNAs target caspase 3 and regulate apoptosis. *PLoS Pathog* **7**: e1002405.

Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M., and Ganem, D. (2005) SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* **435**: 682–686.

Sun, T., Wu, R., and Ming, L. (2019) The role of m6A RNA methylation in cancer. *Biomed Pharmacother Biomedecine Pharmacother* **112**: 108613.

Tagawa, T., Gao, S., Koparde, V.N., Gonzalez, M., Spouge, J.L., Serquina, A.P., et al. (2018) Discovery of Kaposi's sarcoma herpesvirus-encoded circular RNAs and a human antiviral circular RNA. *Proc Natl Acad Sci U S A* **115**: 12805–12810.

Tan, B. and Gao, S.-J. (2018) RNA epitranscriptomics: Regulation of infection of RNA and DNA viruses by N(6) -methyladenosine (m(6) A). *Rev Med Virol* **28**: e1983.

Tan, B., Liu, H., Zhang, S., da Silva, S.R., Zhang, L., Meng, J., et al. (2017) Viral and cellular N6-methyladenosine and N6,2′-O-dimethyladenosine epitranscriptomes in the KSHV life cycle. *Nat Microbiol*.

Tessier, T.M., Dodge, M.J., Prusinkiewicz, M.A., and Mymryk, J.S. (2019) Viral Appropriation: Laying Claim to Host Nuclear Transport Machinery. *Cells* **8**: 559.

Tili, E., Chiabai, M., Palmieri, D., Brown, M., Cui, R., Fernandes, C., et al. (2015) Quaking and miR-155 interactions in inflammation and leukemogenesis. *Oncotarget* **6**: 24599.

Tompkins, V.S., Valverde, D.P., and Moss, W.N. (2018) Human regulatory proteins associate with non-coding RNAs from the EBV IR1 region. *BMC Res Notes* **11**: 139.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., et al. (2017) A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Mol Cell* **66**: 270-284.e13.

Trobaugh, D.W., Gardner, C.L., Sun, C., Haddow, A.D., Wang, E., Chapnik, E., et al. (2014) RNA viruses can hijack vertebrate microRNAs to suppress innate immunity. *Nature* **506**: 245–248.

Truscott, M., Islam, A.B.M.M.K., and Frolov, M.V. (2016) Novel regulation and functional interaction of polycistronic miRNAs. *RNA* **22**: 129–138.

Ule, J., Hwang, H.-W., and Darnell, R.B. (2018) The Future of Cross-Linking and Immunoprecipitation (CLIP). *Cold Spring Harb Perspect Biol* **10**: a032243.

Umbach, J.L. and Cullen, B.R. (2010) In-depth analysis of Kaposi's sarcoma-associated herpesvirus microRNA expression provides insights into the mammalian microRNA-processing machinery. *J Virol* **84**: 695–703.

Van Treeck, B. and Parker, R. (2018) Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. *Cell* **174**: 791–802.

Varble, A., Benitez, A.A., Schmid, S., Sachs, D., Shim, J.V., Rodriguez-Barrueco, R., et al. (2013) An in vivo RNAi screening approach to identify host determinants of virus replication. *Cell Host Microbe* **14**: 346–356.

Viollet, S., Fuchs, R.T., Munafo, D.B., Zhuang, F., and Robb, G.B. (2011) T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol* **11**: 72.

Wang, D., Sun, X., Wei, Y., Liang, H., Yuan, M., Jin, F., et al. (2017) Nuclear miR-122 directly regulates the biogenesis of cell survival oncomiR miR-21 at the posttranscriptional level. *Nucleic Acids Res*.

Wang, F., Song, W., Zhao, H., Ma, Y., Li, Y., Zhai, D., et al. (2017) The RNA-binding protein QKI5 regulates primary miR-124-1 processing via a distal RNA motif during erythropoiesis. *Cell Res* **27**: 416–439.

Wang, H., Hu, X., Huang, M., Liu, J., Gu, Y., Ma, L., et al. (2019) Mettl3-mediated mRNA m(6)A methylation promotes dendritic cell activation. *Nat Commun* **10**: 1898.

Wang, Y., Luo, J., Zhang, H., and Lu, J. (2016) microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. *Mol Biol Evol* **33**: 2232–2247.

Wang, Y., Vogel, G., Yu, Z., and Richard, S. (2013) The QKI-5 and QKI-6 RNA binding proteins regulate the expression of microRNA 7 in glial cells. *Mol Cell Biol* **33**: 1233–1243.

Weinlich, S., Huttelmaier, S., Schierhorn, A., Behrens, S.-E., Ostareck-Lederer, A., and Ostareck, D.H. (2009) IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3'UTR. *RNA N Y N* **15**: 1528–1542.

White, E.J.F., Matsangos, A.E., and Wilson, G.M. (2017) AUF1 regulation of coding and noncoding RNA. *Wiley Interdiscip Rev RNA* **8**:.

Wieczorek, S., Combes, F., Borges, H., and Burger, T. (2019) Protein-Level Statistical Analysis of Quantitative Label-Free Proteomics Data with ProStaR. In, Brun,V. and Couté,Y. (eds), *Proteomics for Biomarker Discovery*. New York, NY: Springer New York, pp. 225–246.

Wieczorek, S., Combes, F., Lazar, C., Giai Gianetto, Q., Gatto, L., Dorffer, A., et al. (2017) DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **33**: 135–136.

Woo, H.-H., Lee, S.C., Gibson, S.J., and Chambers, S.K. (2017) Expression of the cytoplasmic nucleolin for post-transcriptional regulation of macrophage colony-stimulating factor mRNA in ovarian and breast cancer cells. *Biochim Biophys Acta Gene Regul Mech* **1860**: 337–348.

Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M.-D., Yartseva, V., et al. (2013) Mammalian 5'-capped microRNA precursors that generate a single microRNA. *Cell* **155**: 1568–1580.

Xie, M. and Steitz, J.A. (2014) Versatile microRNA biogenesis in animals and their viruses. *RNA Biol* **11**: 673–681.

Xu, Y., Rodriguez-Huete, A., and Pari, G.S. (2006) Evaluation of the lytic origins of replication of Kaposi's sarcoma-associated virus/human herpesvirus 8 in the context of the viral genome. *J Virol* **80**: 9905–9909.

Yan, L., Majerciak, V., Zheng, Z.-M., and Lan, K. (2019) Towards Better Understanding of KSHV Life Cycle: from Transcription and Posttranscriptional Regulations to Pathogenesis. *Virol Sin* **34**: 135–161.

Yan, Y., Hanse, E.A., Stedman, K., Benson, J.M., Lowman, X.H., Subramanian, S., and Kelekar, A. (2016) Transcription factor C/EBP-β induces tumor-suppressor phosphatase PHLPP2 through repression of the miR-17–92 cluster in differentiating AML cells. *Cell Death Differ* **23**: 1232–1242.

Yang, H., Lan, P., Hou, Z., Guan, Y., Zhang, J., Xu, W., et al. (2015) Histone deacetylase inhibitor SAHA epigenetically regulates miR-17-92 cluster and MCM7 to upregulate MICA expression in hepatoma. *Br J Cancer* **112**: 112–121.

Yang, J., Wang, H., and Zhang, W. (2019) Regulation of Virus Replication and T Cell Homeostasis by N(6)-Methyladenosine. *Virol Sin* **34**: 22–29.

Yang, J.-S. and Lai, E.C. (2011) Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol Cell* **43**: 892–903.

Yang, X.-J., Zhu, H., Mu, S.-R., Wei, W.-J., Yuan, X., Wang, M., et al. (2019) Crystal structure of a Y-box binding protein 1 (YB-1)–RNA complex reveals key features and residues interacting with RNA. *J Biol Chem* **294**: 10998–11010.

Yogev, O., Lagos, D., Enver, T., and Boshoff, C. (2014) Kaposi's Sarcoma Herpesvirus MicroRNAs Induce Metabolic Transformation of Infected Cells. *PLOS Pathog* **10**: e1004400.

Yuan, J. and Muljo, S.A. (2013) Exploring the RNA world in hematopoietic cells through the lens of RNA-binding proteins. *Immunol Rev* **253**: 290–303.

Yue, Y., Liu, Jun, Cui, X., Cao, J., Luo, G., Zhang, Z., et al. (2018) VIRMA mediates preferential m(6)A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. *Cell Discov* **4**: 10.

Zanini, I.M.Y., Soneson, C., Lorenzi, L.E., and Azzalin, C.M. (2017) Human cactin interacts with DHX8 and SRRM2 to assure efficient pre-mRNA splicing and sister chromatid cohesion. *J Cell Sci* **130**: 767–778.

Zhang, T., Wu, Y.-C., Mullane, P., Ji, Y.J., Liu, H., He, L., et al. (2018) FUS Regulates Activity of MicroRNA-Mediated Gene Silencing. *Mol Cell* **69**: 787-801.e8.

Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017) An Atomic Structure of the Human Spliceosome. *Cell* **169**: 918-929.e14.

Zhelkovsky, A.M. and McReynolds, L.A. (2011) Simple and efficient synthesis of 5' pre-adenylated DNA using thermostable RNA ligase. *Nucleic Acids Res* **39**: e117–e117.

Ziegelbauer, J.M., Sullivan, C.S., and Ganem, D. (2009) Tandem array-based expression screens identify host mRNA targets of virus-encoded microRNAs. *Nat Genet* **41**: 130–134.

UNIVERSITE DE STRASBOURG

RESUME DE LA THESE DE DOCTORAT

Discipline : **Science de la vie et de la santé**

Spécialité (facultative) : **Aspects moléculaires et cellulaires de la biologie**

Présentée par : **Creugny Antoine**

Titre : **Identification and functional characterization of proteins involved in the regulation of viral microRNAs**

Unité de Recherche : **CNRS UPR 9002. Architecture et réactivité de l'ARN.**

Directeur de Thèse : **Pfeffer Sébastien – Directeur de Recherche**

Localisation : **Université de Strasbourg. Institut de biologie moléculaire et cellulaire - IBMC**

ECOLES DOCTORALES :
(cocher la case)

☐ ED - Sciences de l'Homme et des sociétés ☐ ED 99 – Humanités ☐ ED 101 – Droit, sciences politique et histoire ☐ ED 182 – Physique et chimie physique ☐ ED 221 – Augustin Cournot ☐ ED 222 - Sciences chimiques	☐ ED 269 - Mathématiques, sciences de l'information et de l'ingénieur ☐ ED 270 – Théologie et sciences religieuses ☐ ED 413 – Sciences de la terre, de l'univers et de l'environnement ☒ ED 414 – Sciences de la vie et de la santé

Introduction

Chez les mammifères, les microARN (miARN) constituent une classe essentielle d'ARN régulateurs qui affectent des processus biologiques fondamentaux comme le contrôle de la division et la différenciation cellulaire ou encore des voies de signalisation, comme la réponse immunitaire. Les miARN chargés dans une protéine Argonaute (AGO) vont inhiber la traduction d'ARN messager (ARNm) par complémentarité de base imparfaite entre la région 3' non codante des ARNm ciblés et le miARN. La biogenèse canonique des miARN est un processus complexe débutant par la transcription d'un long ARN précurseur (pri-miARN). Deux maturations successives dans le noyau puis dans le cytoplasme par les ribonucléases de type III Drosha et Dicer respectivement, permettent de produire un duplexe de miARN dont un des deux brins servira de guide à une protéine AGO. La première étape de maturation réalisée par Drosha et son partenaire DGCR8, formant le complexe du microprocesseur, est une étape clé dans la biogenèse car elle contribue à déterminer la séquence du miARN. L'abondance du miARN mature dépend de sa transcription, de l'efficacité de maturation de son précurseur, de sa stabilité et de sa dégradation. Les phénomènes de régulation liés à ces étapes sont essentiels et font intervenir de nombreuses protéines cofacteurs. On retrouve une dérégulation de l'expression des miARN dans la majorité des processus pathologiques y compris les cancers. Certains miARN sont considérés comme proto-oncogènes, comme par exemple miR-155. D'autres miARN sont considérés comme suppresseurs de tumeur, comme les miARN miR-15 et miR-16. Il est donc essentiel de décortiquer les mécanismes moléculaires contrôlant l'expression des miARN afin de mieux comprendre les pathologies associées à leur dérégulation.

Parmi les facteurs à l'origine du développement de cancer, les virus représentent une part importante. Ainsi, plus de 10% des cancers humains possèdent une étiologie virale. Récemment, il a été montré que certains virus expriment leurs propres miARN et que ceux-ci participent à l'induction des pathologies virales. J'étudie un tel virus oncogène : l'herpèsvirus associé au sarcome de Kaposi (KSHV), un gammaherpesvirus au tropisme lymphocytaire et endothélial. La phase de latence virale est sélectionnée par défaut par le KSHV et est responsable des pathologies et des processus tumoraux tels que le sarcome de Kaposi et plusieurs désordres lymphoprolifératifs agressifs. Durant la latence, le KSHV va exprimer plusieurs protéines oncogènes dont la protéine LANA (Latency Associated Nuclear Antigen) qui induit la dégradation de p53 ainsi que 12 précurseurs de miARN (pré-miR-K1 à pré-miR-K12). Dix des miARN du virus sont regroupés en cluster dans un intron (miR-K1 à -9, et miR-K11), tandis que les miR-K10 et -12 sont présents dans la phase codante et la partie 3' UTR de la protéine oncogène Kaposine respectivement. Ces miARN sont conservés entre différentes souches virales et sont nécessaires pour le maintien de la phase de latence mais également pour les propriétés oncogéniques virales. Ces miARN vont réguler, entre autres, l'apoptose, le cycle cellulaire ou l'immunité innée. De manière intéressante, le miARN miR-K11 est un analogue fonctionnel de l'oncomiR-155, dont on sait qu'il est nécessaire pour l'établissement de la latence des gammaherpesvirus.

Les miARN du KSHV présentent la particularité d'être exprimés sous forme de cluster sur le même transcrit primaire, sous le contrôle d'un promoteur latent. Cependant, ils s'accumulent à des niveaux très différents dans les cellules infectées, suggérant une régulation post-transcriptionnelle de leur expression, une caractéristique que l'on retrouve aussi chez un certain nombre de miARN cellulaires. Différents déterminants de structure et de séquences au

sein des précurseurs de miARN sont déjà connus pour expliquer l'efficacité de maturation des miARN par le microprocesseur. Cependant, l'occurrence ou non de ces déterminants au sein des pré-miARN du KSHV ne suffit pas à expliquer la différence d'accumulation des miARN matures.

Mon projet de thèse vise à comprendre l'implication de cofacteurs cellulaires ou viraux dans la régulation de l'expression des miARN viraux au sein d'un cluster, afin d'expliquer leur abondance différentielle sous leur forme mature. Ceci pourra expliquer comment le virus adapte l'expression de ses miARN sans changer son statut transcriptionnel, favorisant par conséquent son adaptation. Pour ce faire, je me concentre sur l'étape de maturation par le microprocesseur dans le noyau, car en plus de déterminer la cible du miARN, elle va également être l'étape limitante expliquant l'abondance relative des miARN issus d'un même transcrit. Il n'existe à ce jour que très peu d'études de la régulation de l'expression des miARN viraux ou en cluster. Notre contribution dans ce domaine, telle que la découverte de nouveaux déterminants de séquence ainsi que les protéines associées permettront de mieux comprendre la biogenèse des miARN viraux et par extension cellulaires. Ceci apportera aussi des réponses quant aux mécanismes moléculaires sous-jacents à la dérégulation de l'accumulation de miARN cellulaires dans certaines pathologies.

Le travail a débuté par la mise en place d'une méthode de purification de complexes ribonucléoprotéiques (*RNA-pulldown*) innovante, afin d'identifier les protéines cofacteurs interagissant de manière spécifique avec les précurseurs de miARN viraux et modulant potentiellement l'activité de reconnaissance ou de coupure du microprocesseur pour ceux-ci. Pour valider l'importance de la fixation des protéines identifiées sur chacun des précurseurs de miARN viraux, nous avons mis au point un crible à l'aide de rapporteurs luciférase dans le but de caractériser leur rôle négatif ou positif. Pour les protéines les plus intéressantes, des vecteurs d'expression et de *silencing* ont ensuite été générés. Ceux-ci permettent de moduler l'expression de ces protéines afin de déterminer leur rôle dans la biogénèse des précurseurs de miARN viraux.

Résultats

Deux approches de *RNA-pulldown* ont été utilisées pendant ce projet. Pour ces deux approches, j'ai employé des transcrits *in vitro* mimant des précurseurs individuels de miARN viraux substrats du microprocesseur. Les protéines retenues ont été ensuite éluées soit de manière aspécifique par la chaleur soit de manière spécifique par un traitement enzymatique. La chromatographie est réalisée à partir d'extraits nucléaires, issus de lignées endothéliale et lymphocytaire infectées ou non. Les expériences ont été réalisées pour les dix précurseurs de miARN du KSHV issu du cluster (miR-K1 à -K9 et miR-K11) et deux miARN cellulaires (Let-7a-1 et miR-155). L'identification et l'enrichissement de protéines pour un précurseur donné par rapport au contrôle a ensuite été réalisé par spectrométrie de masse en utilisant la méthode de comptage de spectres peptidiques ou en utilisant une méthode de plus haute résolution et sensibilité basé sur l'intensité des peptides (Q-EX Orbitrap). Les données de spectrométrie de masse sont traitées par logiciel Prostar afin d'explorer les données, calculer des ratios d'enrichissement et analyser statistiquement les résultats. Ces analyses m'ont permis d'obtenir une liste exhaustive et fiable de protéines interagissant avec les dix précurseurs de miARN du KSHV et ainsi que les 2 miARN cellulaires. Les expériences chromatographie ont été réalisées en triplicata et représentent plus de 90 analyses en

spectrométrie de masse. Les protéines identifiées sont principalement des protéines de liaison à l'ARN, comme des facteurs d'épissage (hnRNP, RBM, ZNF, SF) ou des hélicases (DDX3, DHX36, DHX37). Certaines protéines identifiées comme interagissant avec tous les différents précurseurs ont déjà été caractérisées comme partenaires du microprocesseur telles que les protéines DDX9, NF90 ou NF45. Certaines protéines identifiées n'ont pas encore été décrites dans la littérature comme étant impliquées dans la biogénèse de miARN comme la protéine mitochondriale LRPPRC capable de se localiser dans différents compartiments cellulaires en contexte infectieux. D'autres protéines connues pour réguler spécifiquement Let-7a, comme LIN28b et KSRP ont été identifiées spécifiquement par mon appât Let-7a. Ces dernières observations ont pu valider mon approche expérimentale, et fournissent une base de travail indispensable pour la suite du projet.

Au final, j'ai réussi à identifier de manière reproductible plus de 20 cofacteurs potentiels par précurseur de miARN étudié. La validation biologique de ces résultats a débuté par la mise en place de différents outils qui permettent de mesurer, suite à un *silencing* ou une surexpression des cofacteurs identifiés : (i) leur rôle positif ou négatif sur le microprocesseur (rapporteur luciférase), (ii) l'accumulation des formes précurseurs et matures des miARN (RT-qPCR, northern blot). Afin de cribler ces différentes protéines candidates, deux types de rapporteurs bi-luciférase ont été généré. Le premier, pmiRGLO, permet d'exprimer un ARNm de firefly luciférase possédant dans sa région 3'UTR un précurseur de miARN substrat du microprocesseur. Ce rapporteur permet d'estimer l'efficacité de coupure du précurseur par le microprocesseur. Le deuxième, psiCHECK-2, permet d'exprimer un ARNm de firefly luciférase possédant un site de fixation du miARN mature étudié dans sa région 3'UTR. Ce rapporteur permet d'estimer l'activité du miARN mature sans pour autant discriminer l'étape de biogénèse qui est modulé par la protéine candidate. Malheureusement, les protéines candidates identifiées ne permettent pas de totalement réprimer ou d'augmenter drastiquement la maturation des précurseurs de miARN. Les variations d'abondance des miRNA matures du KSHV semblent finement régulées pour permettre au virus de s'adapter aux voies de signalisation ainsi qu'aux différentes lignées cellulaires infectées sans que ces modulations lui soit délétère. Le manque de sensibilité des rapporteurs luciférases ainsi que des problèmes de normalisation de transfection du plasmide exprimant le cluster viral de miARN, pour les rapporteurs psiCHECK-2, n'ont pas permis de réaliser efficacement le crible. Ce problème de normalisation de l'expression du cluster viral au sein des différentes conditions analytiques restera présent dans toutes mes expériences réalisées dans un système hétérologue, c'est-à-dire des cellules non infectées naturellement. En effet, la caractérisation de l'accumulation des formes précurseurs et matures des miARN viraux par PCR quantitative ou northern blot ont rencontré la même problématique, m'obligeant à travailler directement dans les lymphocytes infectés par le KSHV. Malgré ces problèmes, une ARN m6A methyl-transférase (METTL16), identifiée spécifiquement pour le précurseur kshv-miRNA-K11 semble impliquée dans la biogénèse du précurseur -K11 ainsi que le précurseur adjacent -K6, en favorisant la coupure de DROSHA pour ceux-ci. Les lymphocytes naturellement infectés par le KSHV étant difficiles à transfecter, j'ai par conséquent du générer différentes constructions de vecteurs lentiviraux pour établir des lignées stables réduisant ou sur-exprimant les protéines candidates. Des vecteurs pLenti6.3 ont été utilisés pour surexprimer les protéines candidates avec un tag Flag/Ha en Nter à l'aide d'un promoteur CMV. Des vecteurs pLKO.1 permettant d'exprimer des *shRNA* dirigé contre l'ARNm des protéines candidates sont également générés afin de réaliser un *silencing* stable. Ces lignées transduites obtenues sont des outils importants

pour le laboratoire car ils permettent de travailler en contexte infectieux, et rendent possible des approches de co-immunoprécipitation ou cross-link immunoprécipitation à l'aide du tag Flag/Ha présent en Nter des protéines candidates. De plus, ces lignées permettent également d'étudier l'importance de la protéine candidate dans le cycle viral ainsi que la relation hôte-virus. La génération de lignées a débuté pour les protéines METTL16, MSI2 et hnRNP-A1. Les modulations de l'expression des miARN viraux ainsi que d'autres transcrit viraux latent ont ensuite été mesuré par northern blot et PCR quantitative. Malheureusement, les premières lignées de lymphocytes transduites ne montrent pas de modulation forte et significative de l'expression des miARN du cluster, mais semblent cependant modifier l'expression de transcrits viraux exprimés en latence, notamment l'ARNm RTA contrôlant le cycle latent du KSHV. Plusieurs analyses complémentaires sont encore nécessaires pour déterminer et valider biologiquement l'action de ces protéines cofacteurs sur le cycle viral ou la modification du protéome de l'hôte.

Conclusion

La modulation de la biogénèse de miARN issu d'un cluster est un phénomène complexe nécessitant l'action séquentielle de nombreux cofacteurs protéiques. L'action d'un cofacteur de maturation spécifique pour un précurseur peut également moduler indirectement la maturation des précurseurs adjacents par une modification structurelle induite par la coupure du transcrit. La structure tertiaire du cluster peut également rapprocher géographiquement des cofacteurs spécifiques d'un précurseur proche d'un autre mais éloigné sur le transcrit ajoutant un degré de complexité à la biogénèse spécifique de miARN issu de cluster. De plus, différents cofacteurs spécifiques de maturation pour un miARN donné peuvent également rentrer en compétition. L'aptitude des miARN à inhiber la traduction de nombreuses cibles peut rendre l'identification de cofacteurs de biogénèse difficile. En effet, une augmentation ou diminution même faible de l'expression d'un miARN, dans un laps de temps plus ou moins long, peut modifier l'équilibre de protéines en compétition et par conséquent la réponse de la cellule à différents stimuli. Concernant les virus, le protéome cellulaire disponible lors de l'infection peut favoriser ou non l'établissement efficace du virus. L'action de ces cofacteurs de maturations peut par conséquent être fondamentale uniquement pour l'établissement de l'infection ou la production de néo-particules virales.

Durant ce projet, j'ai réussi à identifier des dizaines de protéines fixant spécifiquement ou non des précurseurs de miARN, substrats du microprocesseur, pour les 10 miARN présents sur le cluster du KSHV. J'ai également mis en place dans le laboratoire différents outils de crible ainsi que des approches expérimentales nécessaires à l'étude de l'implication de ces protéines dans la biogénèse des miARN que ce soit hors ou en contexte infectieux qui s'avèreront indispensables pour la suite de l'étude.

Publication :

Creugny, A., Fender, A., and Pfeffer, S. (2018). Regulation of primary microRNA processing. FEBS Letters 592(12):1980-1996. doi: 10.1002/1873-3468.13067

Communications orales :

RNA Stras Salon - IBMP Université de Strasbourg – 11 juin 2018
Identification of factors regulating Kaposi sarcoma herpesvirus-encoded miRNA biogenesis.
Antoine Creugny[a], Mélanie Messmer[a], Béatrice Chane-Woon-Ming[a], François Delalande[b], Aurélie Hirschler[b], Sarah Cianferani[b], Aurélie Fender[a], Sébastien Pfeffer[a]

Doctoral School Days 2019 – ED414 Université de Strasbourg – 21-22 mars 2019
Identification of factors regulating Kaposi sarcoma herpesvirus-encoded miRNA biogenesis.
Antoine Creugny[a], Mélanie Messmer[a], Béatrice Chane-Woon-Ming[a], François Delalande[b], Aurélie Hirschler[b], Sarah Cianferani[b], Aurélie Fender[a], Sébastien Pfeffer[a]

Communication par affiche :

The complex life of RNA – EMBL Heidelberg – 3-6 octobre 2018
Identification of factors regulating Kaposi sarcoma herpesvirus-encoded miRNA biogenesis.
Antoine Creugny[a], Mélanie Messmer[a], Béatrice Chane-Woon-Ming[a], François Delalande[b], Aurélie Hirschler[b], Sarah Cianferani[b], Aurélie Fender[a], Sébastien Pfeffer[a]

Auteurs :

Antoine Creugny[a], Mélanie Messmer[a], Béatrice Chane-Woon-Ming[a], François Delalande[b], Aurélie Hirschler[b], Sarah Cianferani[b], Aurélie Fender[a], Sébastien Pfeffer[a]
[a] Architecture et Réactivité de l'ARN – UPR 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS, Université de Strasbourg, Strasbourg, France
[b] Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, Strasbourg, France

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

Institut de biologie moléculaire et cellulaire de Strasbourg

Architecture et réactivité de l'ARN – UPR9002

THÈSE présentée par :

Antoine CREUGNY

soutenue le : **15 octobre 2019**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaire et cellulaire de la biologie

Identification et caractérisation fonctionnelle de protéines impliquées dans la régulation de la biogénèse de miARN viraux

[Identification and functional characterization of proteins involved in the regulation of viral microRNAs biogenesis]

THÈSE dirigée par :

Mr PFEFFER Sébastien	Directeur de recherche, Université de Strasbourg

RAPPORTEURS externes :

Mr DIEDERICHS Sven	Professeur, DKFZ – Heidelberg; ZTZ - Freiburg
Mr DECROLY Etienne	Directeur de recherche, Université Aix Marseille

RAPPORTEUR interne :

Mme MEIGNIN Carine	Professeur, Université de Strasbourg

Le manuscrit de thèse est rédigé en anglais. Ce document ne représente qu'un résumé en français. J'encourage le lecteur à préférer la lecture du manuscrit complet (en anglais) pour lequel des liens hypertextes ont été inséré afin de faciliter la lecture.

I. INTRODUCTION

Chez les mammifères, les microARN (miARN) constituent une classe essentielle d'ARN régulateurs et affectent des processus biologiques fondamentaux comme le contrôle de la division et la différenciation cellulaire ou encore des voies de signalisation, comme la réponse immunitaire (Meister and Tuschl, 2004). Les miARN chargés dans une protéine Argonaute (AGO) vont inhiber la traduction d'ARN messager (ARNm) par complémentarité de base imparfaite entre la région 3' non codante des ARNm ciblés et le miARN (Ipsaro and Joshua-Tor, 2015). La biogenèse canonique des miARN est un processus complexe débutant par la transcription d'un long ARN précurseur (pri-miARN) (Ha and Kim, 2014; Creugny *et al.*, 2018). Deux maturations successives dans le noyau puis dans le cytoplasme par les ribonucléases de type III Drosha et Dicer respectivement, permettent de produire un duplexe de miARN dont un des deux brins servira de guide à une protéine AGO. La première étape de maturation réalisée par Drosha et son partenaire DGCR8, formant le complexe du microprocesseur, est une étape clé dans la biogenèse car elle contribue à déterminer la séquence du miARN. Différentes études ont également démontré l'existence de miARN fonctionelle indépendant de l'activité de Drosha et DGCR8 (Figure 1)

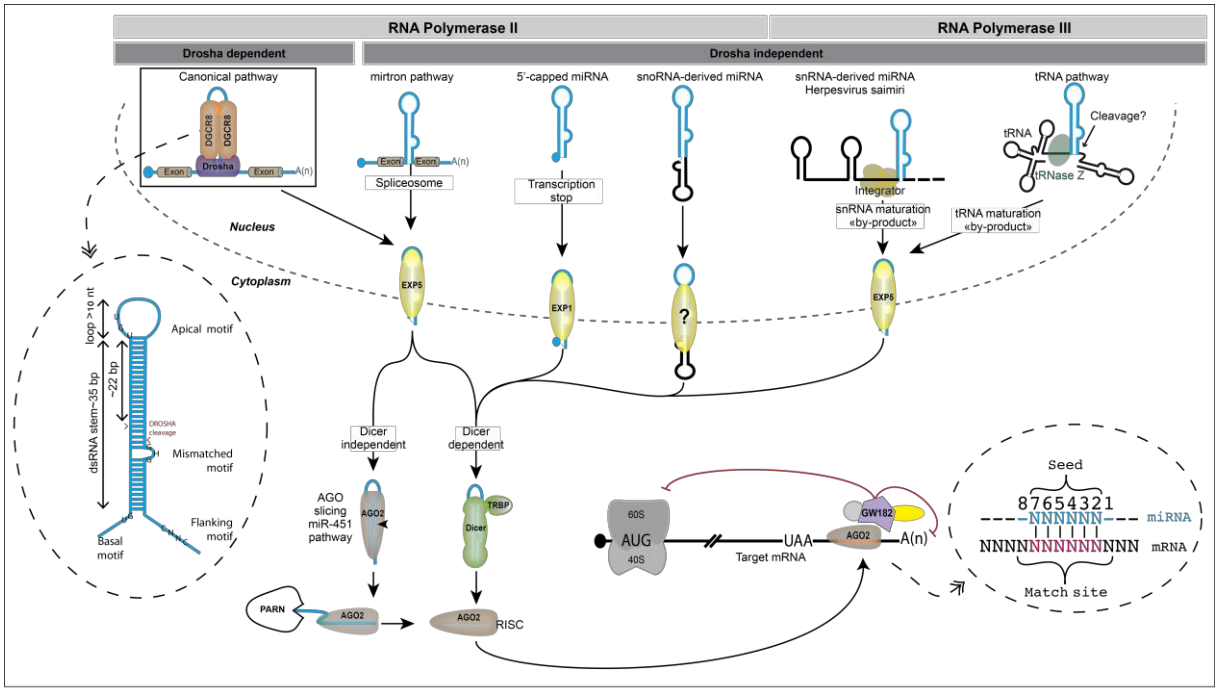


Figure 1. Biogenèse des miARN. Dans la biogenèse canonique, le transcrit primaire contenant une tige boucle précurseur unique ou en cluster est transcrit par la polymérase II. Le microprocesseur (DROSHA/DGCR8) coupe le transcrit primaire pour libérer le pre-miARN qui sera exporter dans le cytoplasme par l'eexportine-5. Les tiges boucles présentes dans le transcrit primaire doivent posséder des motifs de structure et de séquence bien particulier afin d'être reconnu efficacement par le microprocesseur (voir cercle de gauche). Le pre-miARN est

ensuite coupé par DICER pour générer un duplexe de miARN mature pour lequel un des deux brins sera préférentiellement chargé dans une proteine AGO pour former le complexe RISC (RNA induced silencing complex). RISC est dirigé vers des ARNm via l'appariement partiel du miARN avec le messager (voir cercle de droite). Mirtron, miARN 5'-coiffé, snoRNA-derived miRNA correspondent à des précurseurs de miARN indépendant de la maturation du microprocesseur. L'épissage et les arrets de transcriptions sont responsible respectivement de la génération des mirtrons et des miARN 5'coiffé. snRNA-derived and tRNA-derived pre-miARN sont généré lors de la maturation des snoARN et des ARNt. Quasiment tous les miARN sont dépendant de la maturation de DICER. Pre-miR-451 n'est pas maturé par DICER mais directement par l'activité catalytique de AGO2 pour générer un seul miARN.

L'abondance du miARN mature dépend de sa transcription, de l'efficacité de maturation de son précurseur, de sa stabilité et de sa dégradation (Denzler *et al.*, 2016). Les phénomènes de régulation liés à ces étapes sont essentiels et font intervenir de nombreuses protéines cofacteurs (Ha and Kim, 2014; Creugny *et al.*, 2018) (Figure 2).

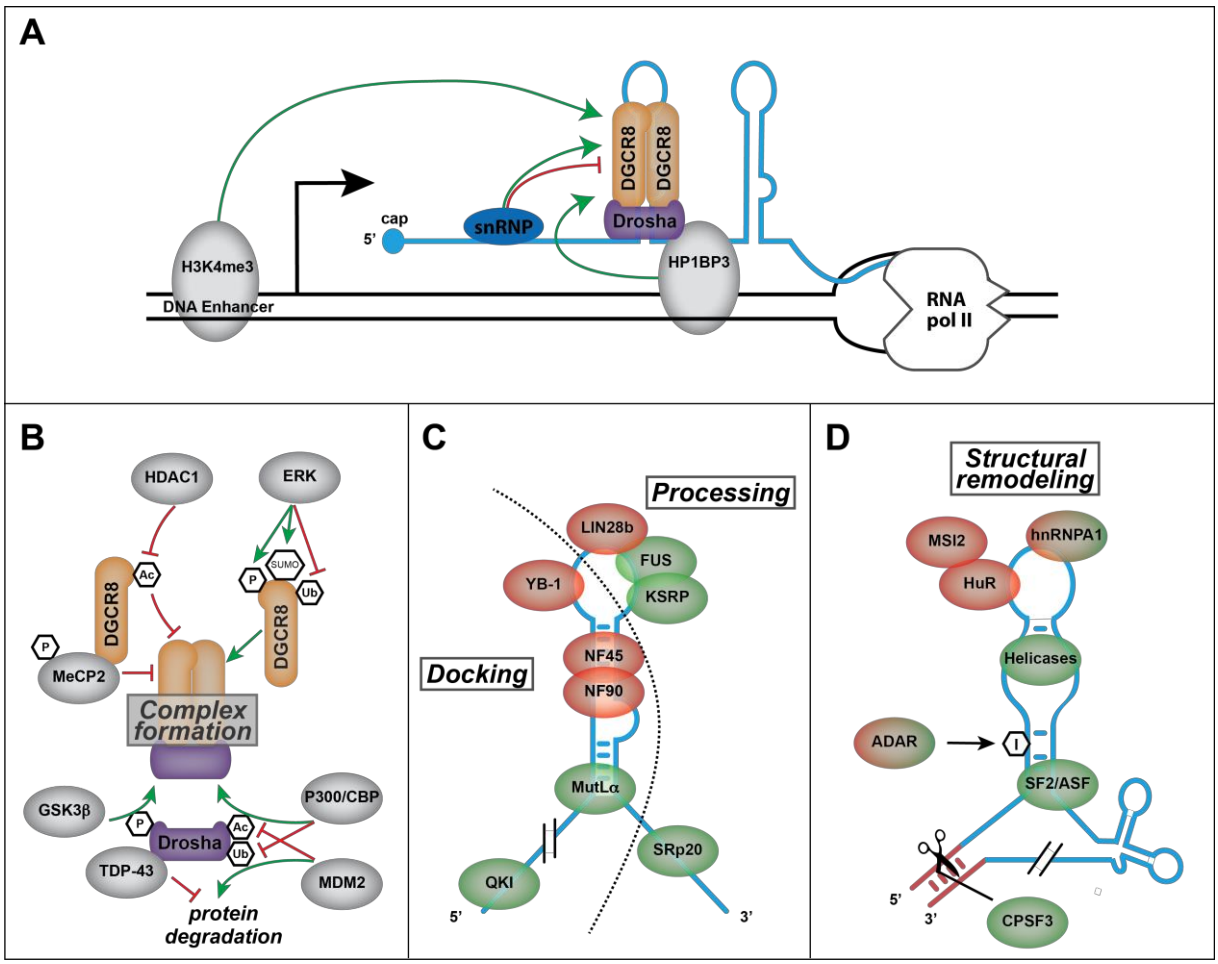


Figure 2. (A) Représentation schématique des processus co-transcriptionels de la biogenèse des miARN. Les réseaux de régulation positif (vert) et négatif (rouge) sont affichés. Le microprocesseur peut être régulé par des snRNP, et/ou des protéines de modifications des histones et marque épigénétiques. (B) Action directe des protéines accessoires sur le complexe

du microprocesseur. Les modifications post-traductionnelles sont représentées par des hexagones (P : phosphate ; Ac : acétyle ; Ub : ubiquitine). Les flèches vertes et rouges représentent l'action positive ou négative. (C) Protéines accessoires interagissant avec le transcrit primaire pour réguler l'ancrage (côté gauche) ou l'activité de coupure (côté droit) du microprocesseur. (D) Protéines accessoires interagissant avec le transcrit primaire induisant un remodelage structural.

On retrouve une dérégulation de l'expression des miARN dans la majorité des processus pathologiques y compris les cancers. Certains miARN sont considérés comme proto-oncogènes, comme hsa-miR-155. D'autres miARN sont considérés comme suppresseurs de tumeur, comme le miARN hsa-miR-15 et hsa-miR-16. Il est donc essentiel de décortiquer les mécanismes moléculaires contrôlant l'expression des miARN afin de comprendre les pathologies associées à leur dérégulation.

Parmi les facteurs à l'origine du développement de cancer, les virus représentent une part importante. Ainsi, plus de 10% des cancers humains possèdent une étiologie virale. Récemment, il a été montré que certains virus expriment leurs propres miARN et que ceux-ci participent à l'induction des pathologies virales (Pfeffer *et al.*, 2004; Girardi *et al.*, 2018). J'étudie un tel virus oncogène : l'herpèsvirus associé au sarcome de Kaposi (KSHV), un gammaherpesvirus au tropisme lymphocytaire et endothélial (Fields *et al.*, 2013). La phase de latence virale est sélectionnée par défaut par le KSHV et est responsable des pathologies et des processus tumoraux tels que le sarcome de Kaposi et plusieurs désordres lymphoprolifératifs agressifs (Feldman *et al.*, 2014). Durant la latence, le KSHV va entre autres exprimer plusieurs protéines oncogènes dont la protéine LANA (Latency Associated Nuclear Antigen) qui induit la dégradation de p53 ainsi que 12 précurseurs de miARN (pré-miR-K1 à pré-miR-12) (Piedade and Azevedo-Pereira, 2016). Dix miARN sont regroupés en cluster dans un intron (kshv-miR-K12-1 à -9, et kshv-miR-K12-11), tandis que kshv-miR-K12-10 et -12 sont présents dans la phase codante et la partie 3' UTR de la protéine oncogène Kaposine respectivement (Cai and Cullen, 2006). Ces miARN sont conservés entre différentes souches virales et sont nécessaires pour le maintien de la phase de latence mais également pour les propriétés oncogéniques virales. Ces miARN vont réguler, entre autres, l'apoptose via kshv-miR-K12-4, le cycle cellulaire via kshv-miR-K12-1, ou l'immunité innée via kshv-miR-K12-11 (Haecker and Renne, 2014) (figure 3).

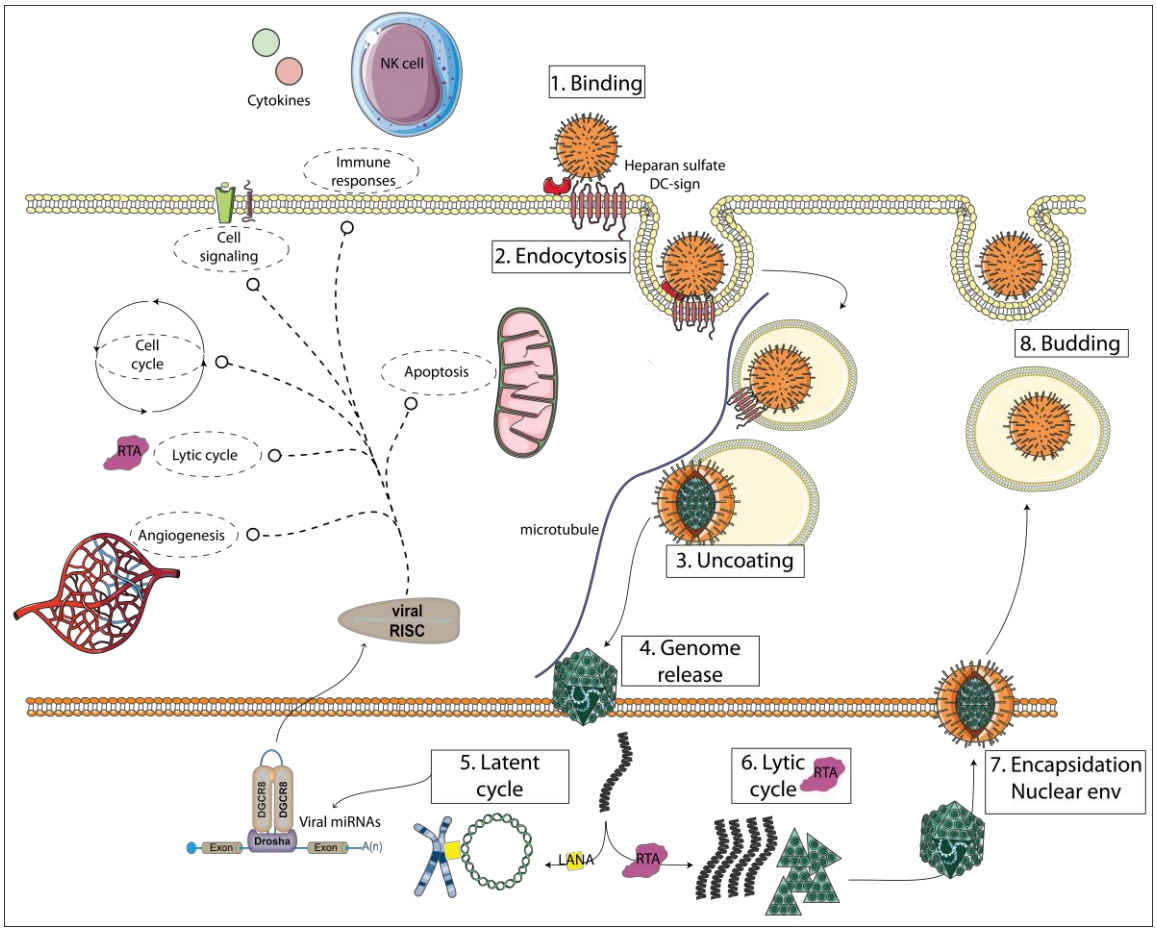


Figure 3. Cycle viral du KSHV et miARN viral. e KSHV reconnaît sa cellule hôte en se liant aux récepteurs. Après endocytose, l'enveloppe virale fusionne avec l'endosome pour libérer la capside. La capside est transportée vers le noyaux à l'aide des microtubules. Le génome du KSHV est libéré dans le noyau. Pendant le cycle latent, le génome viral est circularisé et lié à des chromosomes hôtes via la protéine virale LANA. Lors de la réactivation, le génome viral et les protéines structurelles sont fortement générés. La formation de la capside virale a lieu dans le noyau. L'enveloppe virale est obtenue par bourgeonnement du virus sur la membrane nucléaire. Enfin, la particule virale est libérée de la cellule par bourgeonnement. Pendant la latence, le KSHV exprime de multiples miARNs qui déstabilisent des processus biologiques fondamentaux de l'hôte afin de favoriser un environnement cellulaire avantageux pour le virus.

Les miARN du KSHV présentent la particularité d'être exprimés sous forme de cluster sur le même transcrit primaire, sous le contrôle d'un promoteur latent. Cependant, ils s'accumulent à des niveaux très différents dans les cellules infectées, suggérant une régulation post-transcriptionnelle de leur expression, une caractéristique que l'on retrouve aussi chez un certain nombre de miARN cellulaires (Contrant *et al.*, 2014). Différents déterminants de structure et de séquences au sein des précurseurs de miARN sont déjà connus pour expliquer l'efficacité de maturation des miARN par le microprocesseur (Auyeung *et al.*, 2013).

Cependant, l'occurrence ou non de ces déterminants au sein des pré-miARN du KSHV ne suffit pas à expliquer la différence d'accumulation des miARN matures.

II. OBJECTIFS DE THESE

Mon projet de thèse vise à comprendre l'implication de cofacteur cellulaire ou virale dans la régulation de l'expression des miARN viraux au sein d'un cluster, afin d'expliquer leur abondance différentielle sous leur forme mature (Creugny *et al.*, 2018). Ceci pourra expliquer comment le virus adapte l'expression de ses miARN sans changer son statut transcriptionnel, favorisant par conséquent son adaptation. Pour ce faire, je me concentre sur l'étape de maturation par le microprocesseur dans le noyau, car en plus de déterminer la cible du miARN, elle va également être l'étape limitante expliquant l'abondance relative des miARN issus d'un même transcrit. Il n'existe à ce jour que très peu d'études de la régulation de l'expression des miARN viraux ou en cluster. Notre contribution dans ce domaine, telle que la découverte de nouveaux déterminants de séquence ainsi que les protéines associées permettront de mieux comprendre la biogenèse des miARN viraux et par extension cellulaires. Ceci apportera aussi des réponses quant aux mécanismes moléculaires sous-jacents à la dérégulation de l'accumulation de miARN cellulaires dans certaines pathologies.

Le travail a débuté par la mise en place d'une méthode de purification de complexes ribonucléoprotéiques (*RNA-pulldown*) innovante, afin d'identifier les protéines cofacteurs interagissant de manière spécifique avec les précurseurs de miARN viraux et modulant potentiellement l'activité de reconnaissance ou de coupure du microprocesseur pour ceux-ci. Les protéines identifiées par chacun des précurseurs de miARN viraux, ont ensuite été criblé à l'aide de rapporteur luciférases dans le but de caractériser leur rôle négatif ou positif. Pour les protéines les plus intéressantes, des vecteurs d'expression et de *silencing* ont ensuite été généré. Le rôle de ces protéines, après la modulation de leurs expressions, dans la biogénèse des précurseurs de miARN viraux est ensuite déterminé par northern blot et PCR quantitative.

III. RESULTATS

Deux approches de *RNA-pulldown* ont été utilisé pendant ce projet. Pour ces deux approches, j'ai employé des transcrit *in vitro* mimant des précurseurs individuels de miARN viraux substrats du microprocesseur. Les protéines retenues ont été ensuite élué soit de manière aspécifique par la chaleur soit de manière spécifique par un traitement enzymatique. La chromatographie est réalisée dans des extraits nucléaires, issus de lignée endothéliale et lymphocytaire infecté ou non (Figure 4).

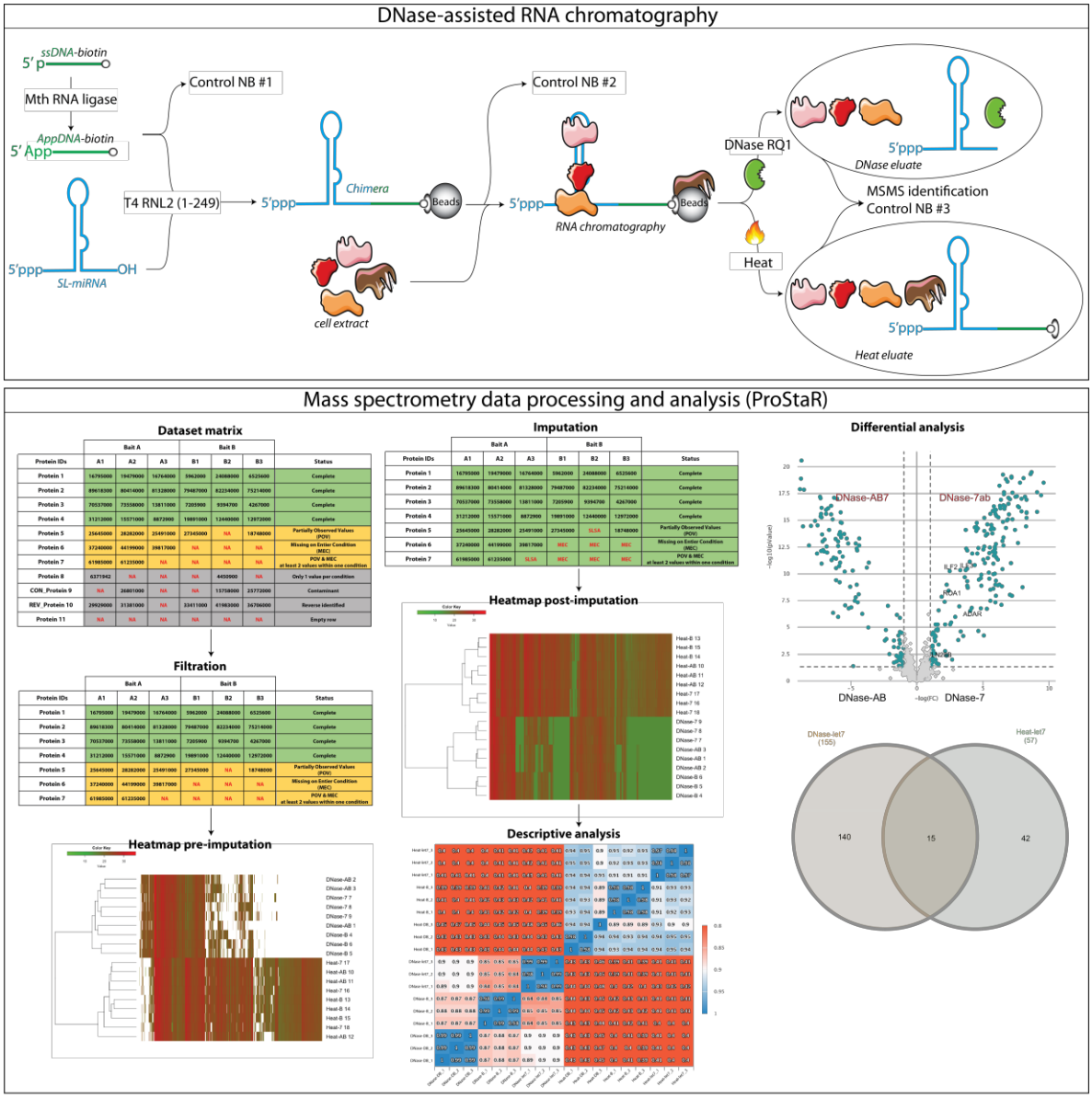


Figure 4. Résumé graphique de l'approche de chromatographie d'affinité à l'ARN et du traitement analytique des données de spectrométrie de masse. Brièvement, la chromatographie utilise des appâts chimériques composés d'un ARN d'intérêt couplé à un oligonucléotide ADN permettant une élution de l'appât à l'aide d'un traitement DNase. Le jeu de donné obtenu par spectrométrie de masse est exploré à l'aide de ProStaR suivant les différentes étapes indiquées.

Les données sont filtrées, et les données manquantes sont insérées. Une analyse descriptive est générée (matrice de corrélation et analyse en composante principale). Puis l'analyse différentielle, basé sur un test de Limma, est réalisée.

Les expériences ont été réalisées pour les dix précurseurs de miARN du KSHV issu du cluster (kshv-miR-K2-1 à -9 et kshv-miR-K12-11) et deux miARN cellulaires (hsa-let-7a-1 et hsa-miR-155). L'identification et l'enrichissement de protéines pour un précurseur donné par rapport au contrôle a ensuite été réalisé par spectrométrie de masse en utilisant la méthode de comptage de spectres peptidiques ou en utilisant une méthode de plus haute résolution et sensibilité basé sur l'intensité des peptides (Q-EX Orbitrap). Les data de spectrométrie de masse sont traitées par logiciel ProStaR (Wieczorek *et al.*, 2017, 2019) afin de, entre autres, explorer les données, calculer des ratios d'enrichissement et analyser statistiquement les résultats. Ces analyses m'ont permis d'obtenir une liste exhaustive et fiable de protéines interagissant avec les dix précurseurs de miARN du KSHV et ainsi que les 2 miARN cellulaires (Figure 5 : miARN -K1, -2, -3, -5, -let7).

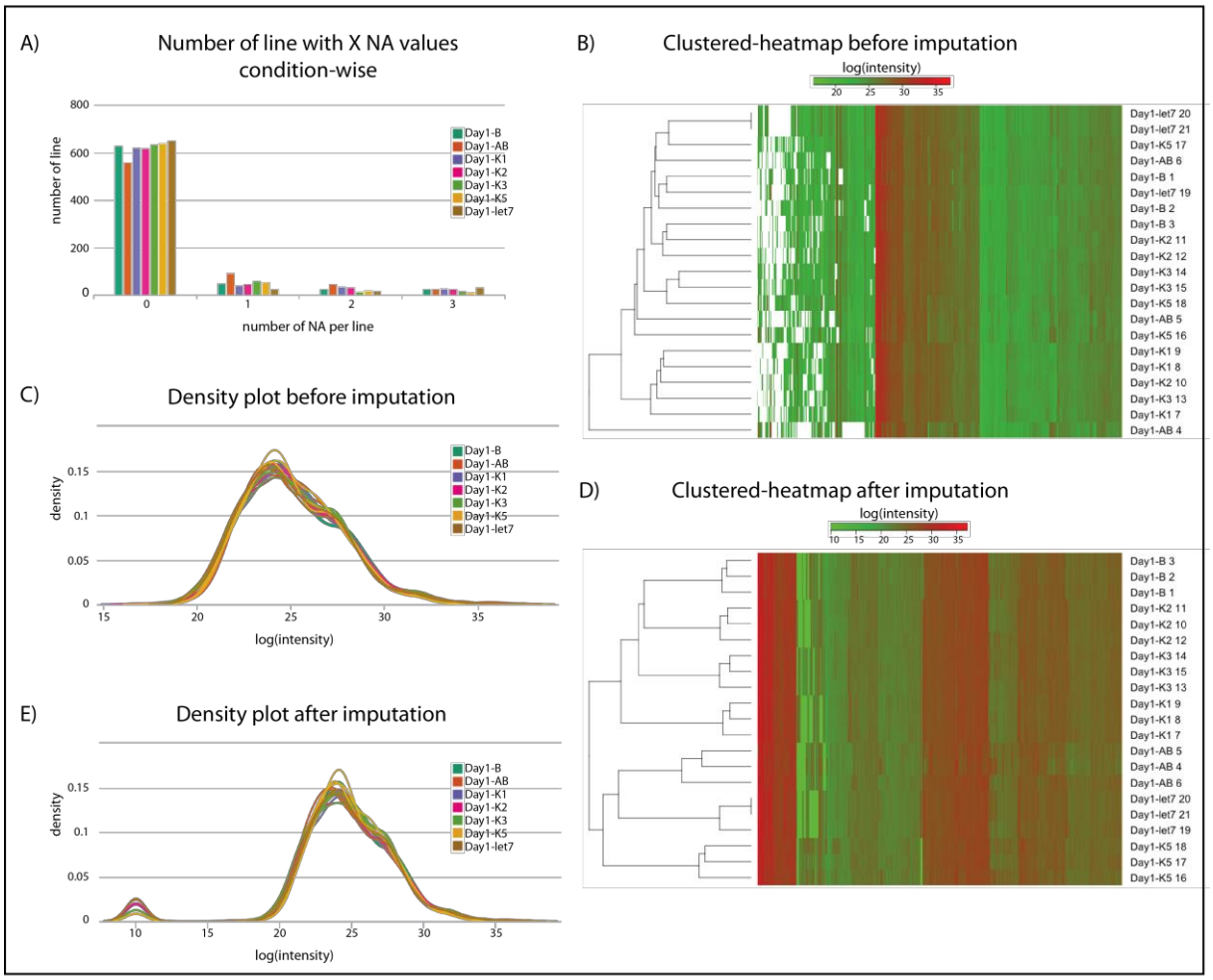


Figure 5. Vue d'ensemble de la chromatographie d'affinité réalisé avec les appats -K1, -K2, -K3, -K5 et -let7 avant et après ajout des données manquantes. Avant l'ajout des données manquantes, (A) représente les données non disponibles parmis la matrice, (B) représente une Euclidian clustered heatmap, (C) montre la densité de l'intensité globale des protéines au sein d'une condition analytique. Après l'ajout des données manquantes, (D) représente une Euclidian clustered heatmap, (E) montre montre la densité de l'intensité globale des protéines au sein d'une condition analytique.

Les expériences chromatographie ont été réalisées en triplicata et représentent plus de 90 analyses en spectrométrie de masse. Les protéines identifiées sont principalement des protéines de liaison à l'ARN, comme des facteurs d'épissage (hnRNP, RBM, ZNF, SF) ou des hélicases (DDX3, DHX36, DHX37) (Figure 6).

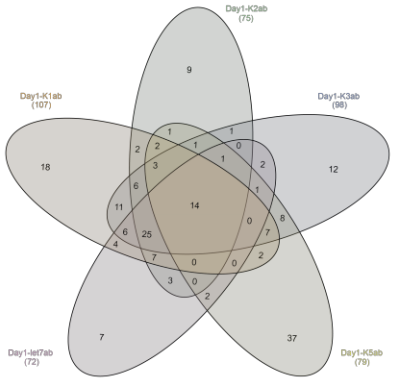
A)

SL-K1	SL-K2	SL-K3	SL-K5	SL-let7	Entry name
Enriched	Enriched	Enriched	NA	NA	3BP1_HUMAN
NA	Enriched	NA	Enriched	NA	ABC3A_HUMAN
NA	NA	NA	Enriched	NA	ABC3C_HUMAN
Enriched	NA	Enriched	NA	NA	ABCE1_HUMAN
NA	NA	NA	Enriched	NA	ADT2_HUMAN
NA	NA	NA	Enriched	NA	ANXA1_HUMAN
Enriched	NA	NA	NA	NA	AP2B1_HUMAN
NA	NA	NA	NA	Enriched	AP3B1_HUMAN
NA	NA	Enriched	Enriched	NA	AP3S1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	BAG2_HUMAN
NA	Enriched	Enriched	Enriched	Enriched	CAH1_HUMAN
Enriched	NA	Enriched	NA	NA	CBX3_HUMAN
Enriched	NA	Enriched	NA	Enriched	CFA20_HUMAN
Enriched	NA	NA	NA	Enriched	CFA58_HUMAN
Enriched	NA	Enriched	NA	Enriched	CIRBP_HUMAN
Enriched	NA	Enriched	NA	NA	CMTR1_HUMAN
NA	NA	Enriched	NA	NA	COF1_HUMAN
NA	NA	Enriched	NA	NA	CPSF4_HUMAN
NA	NA	Enriched	NA	NA	CPSF5_HUMAN
NA	NA	NA	Enriched	NA	CREB1_HUMAN
NA	NA	NA	Enriched	NA	CSK23_HUMAN
NA	NA	NA	Enriched	NA	CTNB1_HUMAN
Enriched	Enriched	Enriched	NA	NA	CWC27_HUMAN
NA	Enriched	NA	NA	Enriched	DAZP1_HUMAN
NA	NA	NA	Enriched	NA	DDX5_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	DHX15_HUMAN
Enriched	NA	NA	NA	NA	DHX36_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	DHX9_HUMAN
Enriched	NA	NA	NA	NA	DJC10_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	DRG1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	DSRAD_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	DYN2_HUMAN
Enriched	NA	Enriched	NA	Enriched	E2AK2_HUMAN
NA	NA	NA	Enriched	NA	EBP_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	EF1A3_HUMAN
Enriched	NA	Enriched	NA	Enriched	EI2BB_HUMAN
NA	Enriched	NA	NA	Enriched	ELAV1_HUMAN
NA	Enriched	Enriched	NA	NA	ELMO1_HUMAN
NA	NA	Enriched	NA	Enriched	ELOB_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	ERH_HUMAN
NA	NA	Enriched	NA	Enriched	ERP29_HUMAN
Enriched	NA	NA	NA	NA	F16A1_HUMAN
NA	NA	NA	Enriched	NA	FABP5_HUMAN
NA	NA	Enriched	Enriched	NA	FEN1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	FKB11_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	FUBP2_HUMAN
NA	NA	NA	Enriched	NA	G3P_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	GCH1_HUMAN
NA	NA	NA	Enriched	NA	H2A1J_HUMAN
NA	NA	NA	Enriched	NA	H2B1L_HUMAN
NA	NA	NA	Enriched	NA	H32_HUMAN
NA	NA	NA	Enriched	NA	H4_HUMAN
NA	NA	NA	Enriched	Enriched	HBA_HUMAN
NA	NA	NA	Enriched	Enriched	HBB_HUMAN
NA	NA	NA	NA	Enriched	HBD_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	HBS1L_HUMAN
Enriched	NA	NA	NA	Enriched	HNRH1_HUMAN
NA	NA	NA	NA	Enriched	HNRH3_HUMAN
Enriched	Enriched	Enriched	NA	NA	HNRL1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	HNRLL_HUMAN
NA	Enriched	NA	NA	NA	HNRPD_HUMAN
Enriched	NA	Enriched	NA	NA	HNRPK_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	HNRPL_HUMAN
Enriched	Enriched	NA	Enriched	NA	HNRPQ_HUMAN
Enriched	Enriched	Enriched	NA	NA	HSP7E_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	IF2A_HUMAN
Enriched	NA	Enriched	Enriched	NA	IF2B_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	IF2G_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	ILF2_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	ILF3_HUMAN
Enriched	NA	Enriched	Enriched	NA	IMA1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	IRF4_HUMAN
Enriched	NA	Enriched	NA	Enriched	ISG20_HUMAN
NA	NA	NA	Enriched	NA	K22O_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	LA_HUMAN
NA	NA	NA	Enriched	NA	LEG7_HUMAN
NA	NA	NA	Enriched	NA	LMNA_HUMAN
Enriched	Enriched	NA	NA	Enriched	LN28B_HUMAN
Enriched	NA	NA	NA	NA	LSM2_HUMAN
NA	Enriched	NA	NA	NA	MAOM_HUMAN
Enriched	NA	NA	NA	NA	MAP2_HUMAN
NA	Enriched	NA	NA	NA	MBNL1_HUMAN
NA	Enriched	NA	NA	NA	MCM3_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	MCTS1_HUMAN
NA	NA	Enriched	NA	NA	MGN2_HUMAN
NA	Enriched	Enriched	Enriched	NA	MGST1_HUMAN
NA	NA	NA	Enriched	NA	MPCP_HUMAN
Enriched	NA	Enriched	Enriched	NA	MRGBP_HUMAN
NA	NA	NA	NA	Enriched	MSI2H_HUMAN
NA	Enriched	NA	NA	NA	NEPR1_HUMAN
Enriched	NA	NA	NA	Enriched	NIPBL_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	NOSIP_HUMAN
Enriched	NA	NA	NA	NA	NUCL_HUMAN
Enriched	Enriched	NA	NA	Enriched	PABP1_HUMAN
Enriched	NA	NA	NA	NA	PACS1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	PARVB_HUMAN

B)

SL-K1	SL-K2	SL-K3	SL-K5	SL-let7	Entry name
Enriched	NA	Enriched	NA	NA	PCBP1_HUMAN
Enriched	NA	Enriched	Enriched	NA	PCBP2_HUMAN
Enriched	NA	NA	NA	NA	PDIA6_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	PHB_HUMAN
Enriched	NA	NA	NA	NA	PMS2_HUMAN
NA	NA	NA	Enriched	NA	PPIE_HUMAN
NA	NA	Enriched	Enriched	NA	PRAF3_HUMAN
NA	NA	NA	NA	Enriched	PRDX3_HUMAN
Enriched	NA	NA	NA	Enriched	PROF1_HUMAN
NA	NA	NA	Enriched	NA	PRP8_HUMAN
NA	NA	NA	NA	Enriched	PRS6A_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	PTBP1_HUMAN
Enriched	NA	Enriched	Enriched	NA	QKI_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RAB21_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RAB2A_HUMAN
NA	NA	Enriched	NA	NA	RAB7A_HUMAN
NA	NA	Enriched	NA	NA	RAGP1_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RAN_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RAP1A_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RB11A_HUMAN
NA	NA	Enriched	NA	NA	RBBP6_HUMAN
NA	NA	NA	NA	Enriched	RBM3_HUMAN
Enriched	NA	Enriched	NA	NA	RBM39_HUMAN
Enriched	NA	Enriched	NA	NA	RBM45_HUMAN
Enriched	Enriched	NA	NA	NA	RBMS1_HUMAN
Enriched	Enriched	NA	NA	NA	RBMS2_HUMAN
NA	NA	NA	Enriched	NA	RBMX_HUMAN
Enriched	Enriched	Enriched	NA	NA	RBP2_HUMAN
Enriched	NA	NA	Enriched	NA	RBP56_HUMAN
NA	Enriched	NA	NA	NA	RCC1_HUMAN
Enriched	NA	NA	NA	NA	RECQ1_HUMAN
NA	NA	NA	Enriched	NA	RFA1_HUMAN
NA	NA	NA	Enriched	NA	RFA2_HUMAN
NA	Enriched	NA	NA	Enriched	ROA0_HUMAN
Enriched	Enriched	NA	NA	Enriched	ROA1_HUMAN
Enriched	Enriched	NA	NA	Enriched	ROA2_HUMAN
Enriched	Enriched	NA	NA	Enriched	ROA3_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RPAB1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RPAC1_HUMAN
NA	NA	Enriched	NA	NA	RPB1_HUMAN
NA	NA	Enriched	NA	NA	RPB2_HUMAN
Enriched	NA	Enriched	NA	Enriched	RPB3_HUMAN
NA	NA	Enriched	NA	NA	RPB9_HUMAN
Enriched	NA	Enriched	NA	NA	RPP40_HUMAN
NA	NA	NA	Enriched	NA	RS11_HUMAN
NA	NA	Enriched	Enriched	NA	RS16_HUMAN
NA	NA	Enriched	Enriched	NA	RS18_HUMAN
NA	NA	NA	Enriched	NA	RS2_HUMAN
NA	NA	NA	Enriched	NA	RS23_HUMAN
Enriched	Enriched	Enriched	Enriched	NA	RS3A_HUMAN
Enriched	NA	Enriched	Enriched	NA	RS5_HUMAN
NA	NA	NA	Enriched	NA	RS7_HUMAN
NA	NA	NA	Enriched	NA	RS9_HUMAN
Enriched	Enriched	NA	NA	Enriched	RT27_HUMAN
Enriched	NA	NA	NA	NA	RTF2_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	RU2A_HUMAN
NA	NA	Enriched	NA	NA	RU2B_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	RUN3B_HUMAN
NA	NA	Enriched	Enriched	NA	RUXE_HUMAN
NA	NA	Enriched	Enriched	NA	RUXF_HUMAN
Enriched	Enriched	NA	NA	Enriched	SCFD1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	SF01_HUMAN
Enriched	NA	Enriched	NA	NA	SF3A2_HUMAN
NA	NA	Enriched	NA	NA	SMD1_HUMAN
NA	NA	NA	Enriched	NA	SPR2B_HUMAN
Enriched	NA	Enriched	Enriched	NA	SRP09_HUMAN
NA	NA	NA	Enriched	NA	SRP14_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	SRSF9_HUMAN
NA	NA	NA	Enriched	NA	SSRD_HUMAN
NA	Enriched	NA	NA	NA	STAG2_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	STRBP_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	STXB2_HUMAN
Enriched	NA	NA	NA	NA	SUGP1_HUMAN
Enriched	Enriched	NA	Enriched	NA	SYDC_HUMAN
Enriched	NA	NA	NA	NA	SYFB_HUMAN
Enriched	NA	NA	NA	NA	SYYC_HUMAN
Enriched	Enriched	Enriched	NA	NA	T2EA_HUMAN
Enriched	NA	NA	NA	NA	TADBP_HUMAN
Enriched	NA	Enriched	NA	NA	TASOR_HUMAN
NA	NA	NA	Enriched	NA	TBA4A_HUMAN
NA	NA	Enriched	Enriched	NA	TCEA1_HUMAN
NA	NA	NA	Enriched	NA	TCP4_HUMAN
NA	NA	NA	Enriched	NA	THOC4_HUMAN
NA	NA	NA	Enriched	NA	THOC6_HUMAN
NA	Enriched	NA	NA	NA	THYN1_HUMAN
Enriched	NA	NA	NA	NA	TLN1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	TLN2_HUMAN
NA	NA	Enriched	Enriched	Enriched	TMEDA_HUMAN
NA	NA	NA	Enriched	NA	U2AF1_HUMAN
Enriched	NA	Enriched	NA	NA	U2AF2_HUMAN
Enriched	NA	NA	Enriched	NA	U520_HUMAN
NA	NA	NA	Enriched	NA	VDAC1_HUMAN
Enriched	NA	NA	NA	NA	WWP2_HUMAN
NA	Enriched	NA	NA	NA	XRCC6_HUMAN
Enriched	Enriched	Enriched	Enriched	Enriched	YBOX1_HUMAN
Enriched	Enriched	Enriched	NA	Enriched	ZFR_HUMAN

C)



D)

Common proteins (minus -K5)		
BAG2_HUMAN	IF2G_HUMAN	RAP1A_HUMAN
DHX15_HUMAN	ILF2_HUMAN	RB11A_HUMAN
DHX9_HUMAN	ILF3_HUMAN	RPAB1_HUMAN
DSRAD_HUMAN	IRF4_HUMAN	RPAC1_HUMAN
DYN2_HUMAN	LA_HUMAN	RU2A_HUMAN
EF1A3_HUMAN	MCTS1_HUMAN	RUN3B_HUMAN
ERH_HUMAN	NOSIP_HUMAN	SF01_HUMAN
FKB11_HUMAN	PARVB_HUMAN	SRSF9_HUMAN
FUBP2_HUMAN	PHB_HUMAN	STRBP_HUMAN
HBS1L_HUMAN	PTBP1_HUMAN	STXB2_HUMAN
HNRLL_HUMAN	RAB21_HUMAN	TLN2_HUMAN
HNRPL_HUMAN	RAB2A_HUMAN	YBOX1_HUMAN
IF2A_HUMAN	RAN_HUMAN	ZFR_HUMAN

Figure 6. Analyse différentielle de l'ensemble des données issues des chromatographies réalisées avec les appâts -K1, -K2, -K3, -K5 et -let7. (A) et (B) représentent les protéines enrichies de manière significative par rapport au témoin (enrichissement supérieur à 2, p-values inférieure à 5%). (C) Diagramme de Venn des protéines enrichies. (D) Protéines enrichies significativement par tous les appâts de la chromatographie.

Certaines protéines identifiées comme interagissant avec tous les différents précurseurs ont déjà été caractérisées comme partenaires du microprocesseur telles que les protéines DDX9, NF90, NF45. Certaines protéines identifiées n'ont pas encore été décrite dans la littérature comme étant impliqué dans la biogénèse de miARN comme la protéine mitochondrial LRPPRC capable de se localiser dans différents compartiments cellulaires en contexte infectieux. D'autres protéines connues pour réguler spécifiquement hsa-let-7a-1, comme LIN28b et KSRP ont été identifiées spécifiquement par mon appât hsa-let-7a-1. Ces dernières observations ont pu valider mon approche expérimentale, et fournissent une base de travail indispensable pour la suite de mon projet.

Au final, j'ai réussi à identifier de manière reproductible plus de 20 cofacteurs potentiels par précurseurs de miARN étudié. La validation biologique de ces résultats a débuté par la mise en place de différents outils qui permettent de mesurer, suite à un *silencing* ou une surexpression des cofacteurs identifiés : (i) leur rôle positif ou négatif sur le microprocesseur (rapporteur luciférase), (ii) l'accumulation des formes précurseurs et matures des miARN (RT-qPCR, northern blot). Afin de cribler ces différentes protéines candidates, deux types de rapporteurs bi-luciférases ont été généré. Le premier, pmiRGLO, permet d'exprimer un ARNm de firefly luciférase possédant dans sa région 3'UTR un précurseur de miARN substrat du microprocesseur. Ce rapporteur permet d'estimer l'efficacité de coupure du précurseur par le microprocesseur (Figure 7).

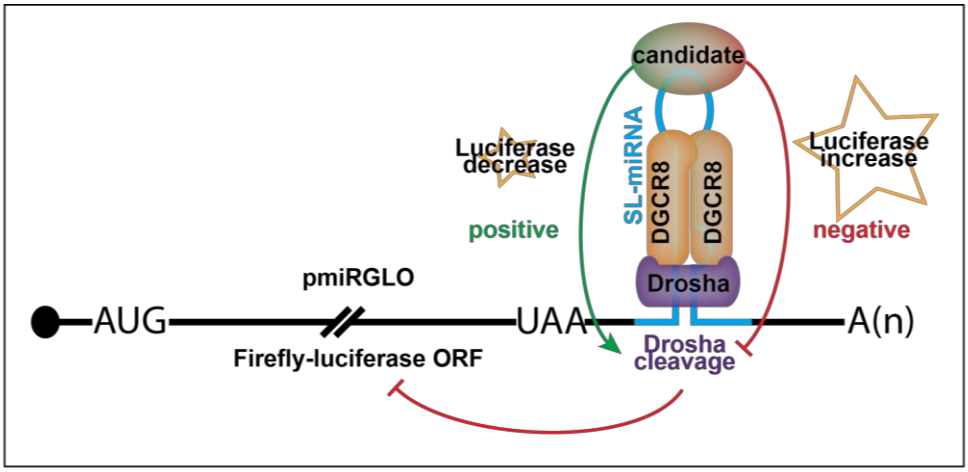


Figure 7. Représentation schématique du rapporteur luciférase pmiRGLO. La modulation de la protéine candidate est réalisée avant la transfection du rapporteur. Le candidat peut avoir

un effet positif ou négatif sur le clivage de DROSHA induisant respectivement une augmentation ou une diminution du clivage et par conséquent une déstabilisation ou stabilisation de l'ARNm de la luciférase.

Le deuxième, psiCHECK-2, permet d'exprimer un ARNm de firefly luciférase possédant un site de fixation du miARN mature étudié dans sa région 3'UTR. Ce rapporteur permet d'estimer l'activité du miARN mature sans pour autant discriminer l'étape de biogénèse qui est modulé par la protéine candidate (Figure 8).

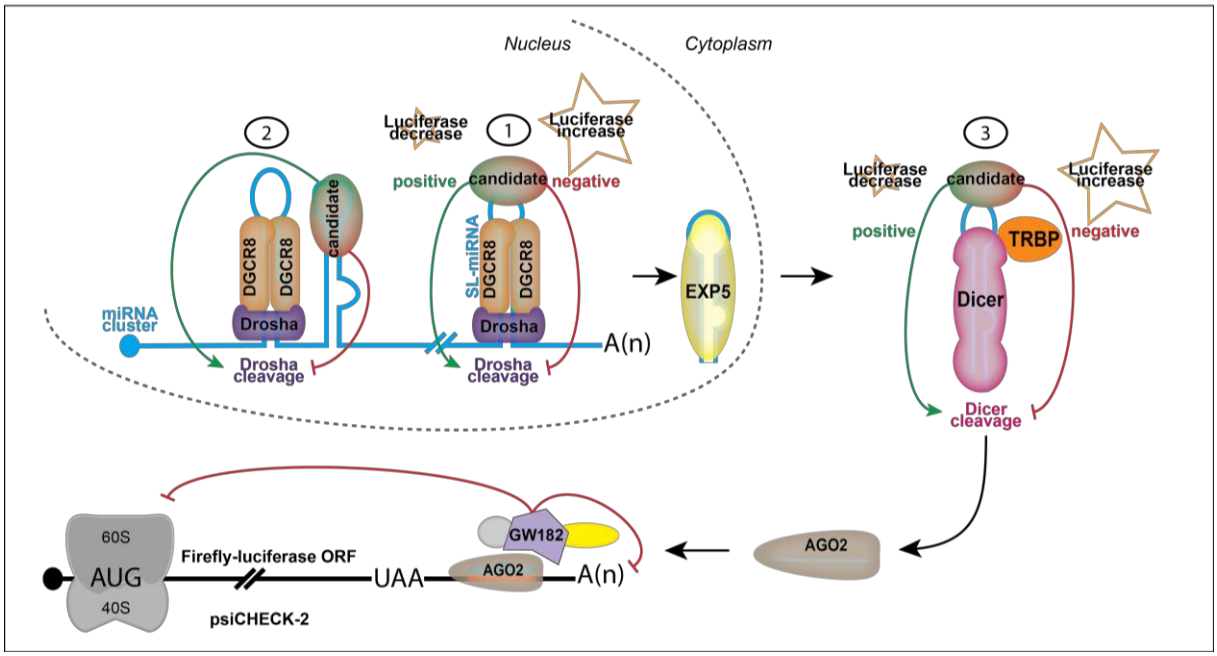


Figure 8. Représentation schématique du système rapporteur psiCHECK-2. La modulation du candidat est réalisé avant la transfection du psiCHECK-2 ainsi que le plasmide exprimant les miARN du KSHV. Le candidat peut agir positivement ou négativement sur toutes les étapes de maturation du miARN. Un candidat à action positive va réduire le signal de luciférase. Chaque candidat peut également agir de manière direct (1) ou indirect (2) sur les précurseurs voisins.

Malheureusement, les protéines candidates identifiées ne permettent pas de totalement réprimer ou d'augmenter drastiquement la maturation des précurseurs de miARN. Les variations d'abondance des miRNA matures du KSHV semblent finement régulées pour permettre au virus de s'adapter aux voies de signalisation ainsi qu'aux différentes lignées cellulaires infectées sans que ces modulations lui soit délétère. Le manque de sensibilité des rapporteurs luciférases ainsi que des problèmes de normalisation de transfection du plasmide exprimant le cluster viral de miARN, pour les rapporteurs psiCHECK-2, n'ont pas permis de réaliser efficacement le crible (Figure 9).

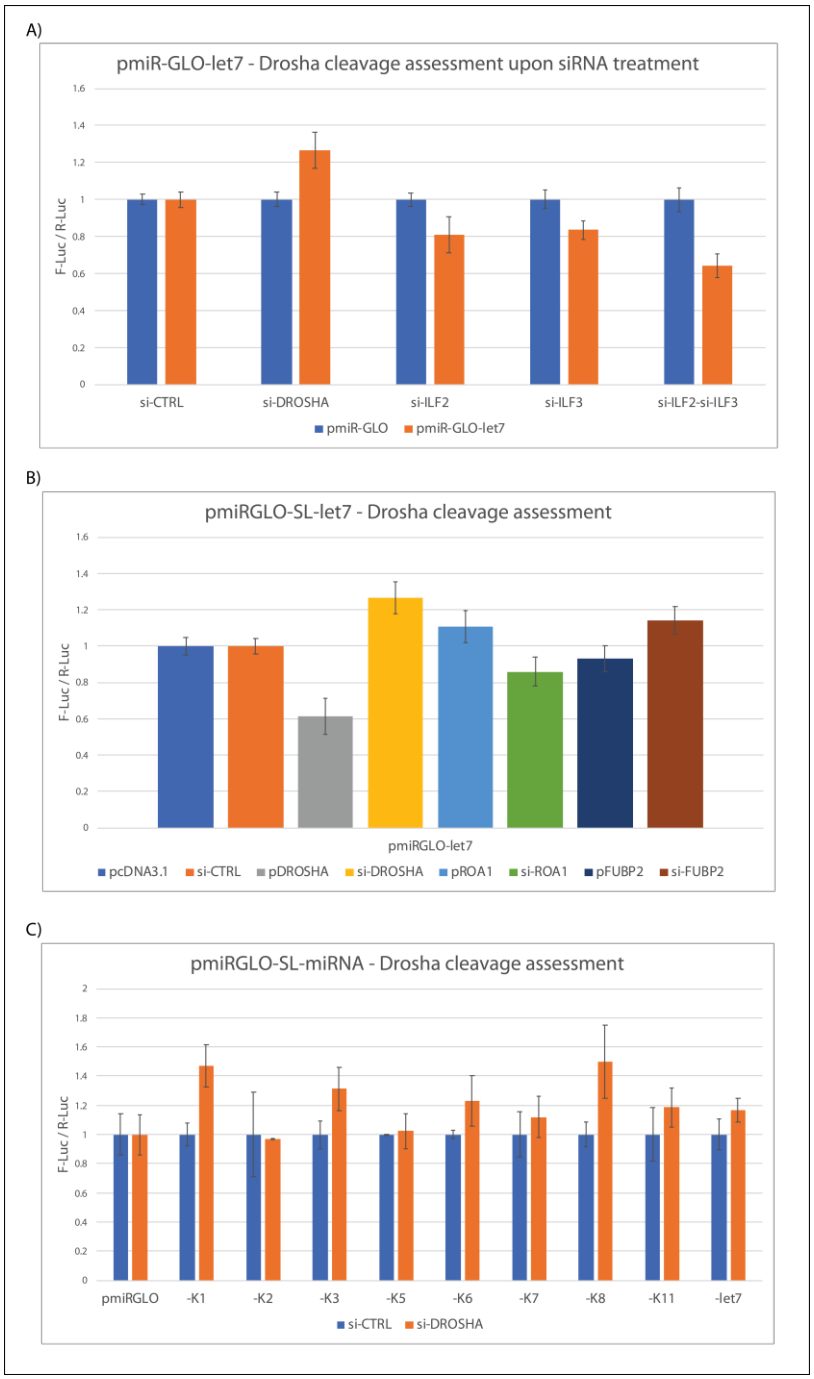


Figure 9. Résultat du rapporteur pmiRGLO. Le signal de firefly luciférase (F-Luc) est médié par l'étape de clivage DROSHA. La luciférase de Renilla (R-Luc) agit comme un normalisateur de transfection. L'affichage du signal de luciférase correspond au rapport F-Luc/R-Luc. Les données ont fait l'objet d'une double normalisation d'abord sur le contrôle pmiRGLO vide, puis sur le contrôle si-ARN (siCTRL) pour l'expérience de silencing ou sur le vecteur vide (pcADN3.1) pour l'expérience de surexpression. A) pmiRGLO-let7 est transfecté dans des cellules HEK293-GripTite durant un traitement par ARN interférent dirigé contre des composant du microprocesseur. B) pmiRGLO-let7 est transfecté dans des cellules HEK293-GripTite en présence de silencing ou sur-expression de protéine cofacteur du microprocesseur. C) Les différents pmiRGLO portant les précurseurs de miARN du KSHV sont transfectés individuellement dans des cellules HEK293-GripTite lors du silencing de DROSHA.

Ce problème de normalisation de l'expression du cluster virale au sein des différentes conditions analytiques restera présent dans toutes mes expériences réalisées dans un système hétérologues, c'est-à-dire des cellules non infectées naturellement. En effet, la caractérisation de l'accumulation des formes précurseurs et matures des miARN viraux par PCR quantitative ou northern blot ont rencontrées la même problématique, m'obligeant à travailler directement dans les lymphocytes infectés par le KSHV (Figure 10).

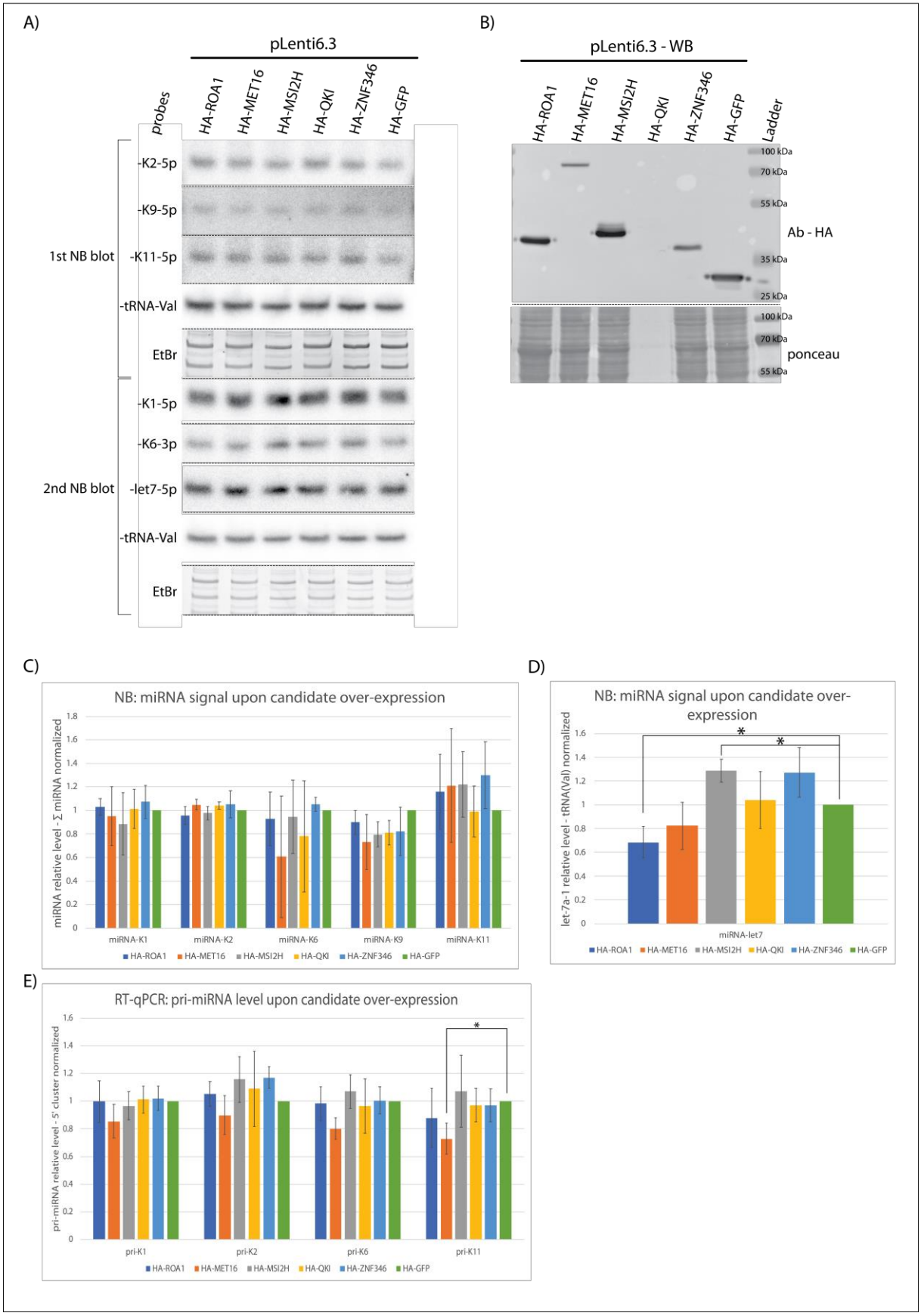


Figure 10. Surexpression des candidats dans lignées non infectées. A) Northern blot à partir d'extrait d'ARN totaux issue de lignée sur-exprimant les candidats. B) Western blot de

validation de la sur-expression. C) Quantification des triplicats biologiques par northern blot. D) Quantification du signal de let-7a-1 par northern blot pour les différentes sur-expression. E) Quantification du pri-miARN par RT-qPCR.

Malgré cette problématique, une ARN m6A methyl-transférase (METTL16), identifié spécifiquement par le précurseur kshv-miRNA-K12-11 semble impliqué dans la biogénèse du précurseur -K12-11 ainsi que le précurseur adjacent -K12-6, en favorisant la coupure de DROSHA pour ceux-ci. Les lymphocytes naturellement infectés par le KSHV étant excessivement difficiles à transfecter, j'ai par conséquent du générer différentes constructions de vecteurs lentivrale pour établir des lignées stables réduisant ou sur-exprimant les protéines candidates. Des vecteurs pLenti6.3 ont été utilisé pour sur-exprimer les protéines candidates avec un tag Flag/Ha en N-ter à l'aide d'un promoteur CMV. Des vecteurs pLKO.1 permettant d'exprimer des *short-hairpinRNA* dirigé contre l'ARNm des protéines candidates sont également générer afin de réaliser un *silencing* stable. Ces lignées transduites obtenues sont des outils importants pour le laboratoire car ils permettent de travailler en contexte infectieux, rendent possible des approches de co-immunoprécipitation ou cross-link immunoprécipitation à l'aide du tag Flag/Ha présent en N-ter des protéines candidates. De plus, ces lignées permettent également d'étudier l'importance de la protéine candidate dans le cycle viral ainsi que la relation hôte-virus. La génération de lignée a débuté pour les protéines METTL16, MSI2 et hnRNP-A1. Les modulations de l'expression des miARN viraux ainsi que d'autres transcrit viraux latent ont ensuite été mesuré par northern blot et PCR quantitative. Malheureusement, les premières lignées de lymphocytes transduites ne montrent pas de modulation forte et significative de l'expression des miARN du cluster, mais semblent cependant modifier l'expression de transcrits viraux exprimés en latence, notamment l'ARNm RTA contrôlant le cycle latent du KSHV (Figure 11).

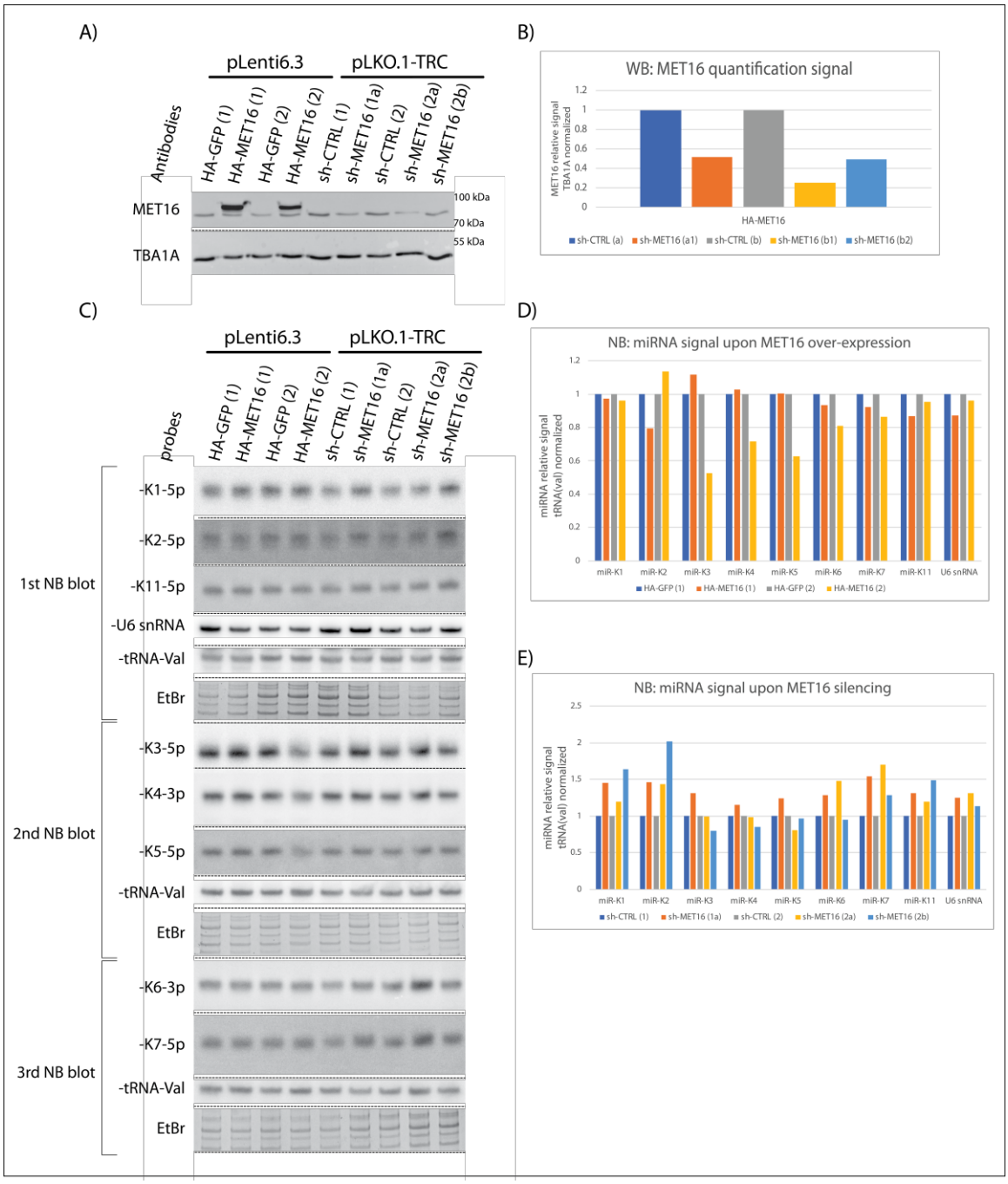


Figure 11. Surexpression et silencing de MET16 dans des lignées de lymphocyte naturellement infecté par le KSHV. A) Western blot permettant la vérification de la surexpression et du silencing. La tubuline (TBA1A) est utilisé comme normalisateur. B) Quantification des signaux de western blot à l'aide de ImageQuant-TL. C) Northern blot des extraits d'ARN totaux issue des lymphocytes transduits. D) Quantification des signaux de northern blot issue de lignée surexprimant MET16. L'histogramme représente le signal moyen obtenu avec les différents réplicas biologiques. E) Quantification des signaux de northern blot issue de lignée sous-exprimant MET16.

Plusieurs analyses complémentaires sont encore nécessaires pour déterminer et valider biologiquement l'action de ces protéines cofacteurs sur le cycle virale ou la modification du protéome de l'hôte.

Conclusion

La modulation de la biogénèse de miARN issu d'un cluster est un phénomène complexe nécessitant l'action séquentielle de nombreux cofacteurs protéiques. L'action d'un cofacteur de maturation spécifique pour un précurseur peut également moduler indirectement la maturation des précurseurs adjacents par une modification structurelle induite par la coupure du transcrit. La structure tertiaire du cluster peut également rapprocher géographiquement des cofacteurs spécifiques d'un précurseur proche d'un autre mais éloigné sur le transcrit ajoutant un degré de complexité à la biogénèse spécifique de miARN issu de cluster. De plus, différents cofacteurs spécifiques de maturation pour un miARN donné peuvent également rentrer en compétition. L'aptitude des miARN à inhiber la traduction de nombreuses cibles rends peut rendre l'identification de cofacteurs de biogénèse difficile. En effet, une augmentation ou diminution même faible de l'expression d'un miARN, dans un laps de temps plus ou moins long, peut modifier certain équilibre de protéine en compétition et par conséquent la réponse de la cellule a différent stimuli. Concernant les virus, le protéome cellulaire disponible lors de l'infection peut favoriser ou non l'établissement efficace du virus. L'action de ces cofacteurs de maturations peut par conséquent être fondamentale uniquement pour l'établissement de l'infection ou la production de néo-particules virale.

Durant ce projet, j'ai réussi à identifier des dizaines de protéines fixant spécifiquement ou non des précurseurs de miARN, substrats du microprocesseur, pour les 10 miARN présents sur le cluster du KSHV. J'ai également mis en place dans le laboratoire différents outils de crible ainsi que des approches expérimentales nécessaire à l'étude de l'implication de ces protéines dans la biogénèse des miARN que ce soit hors ou en contexte infectieux qui s'avèreront indispensable pour la suite de l'étude.

De multiples expériences (résumé Figure 12) sont nécessaire pour valider l'implication biologique du candidat dans la biogenèse spécifique de miARN.

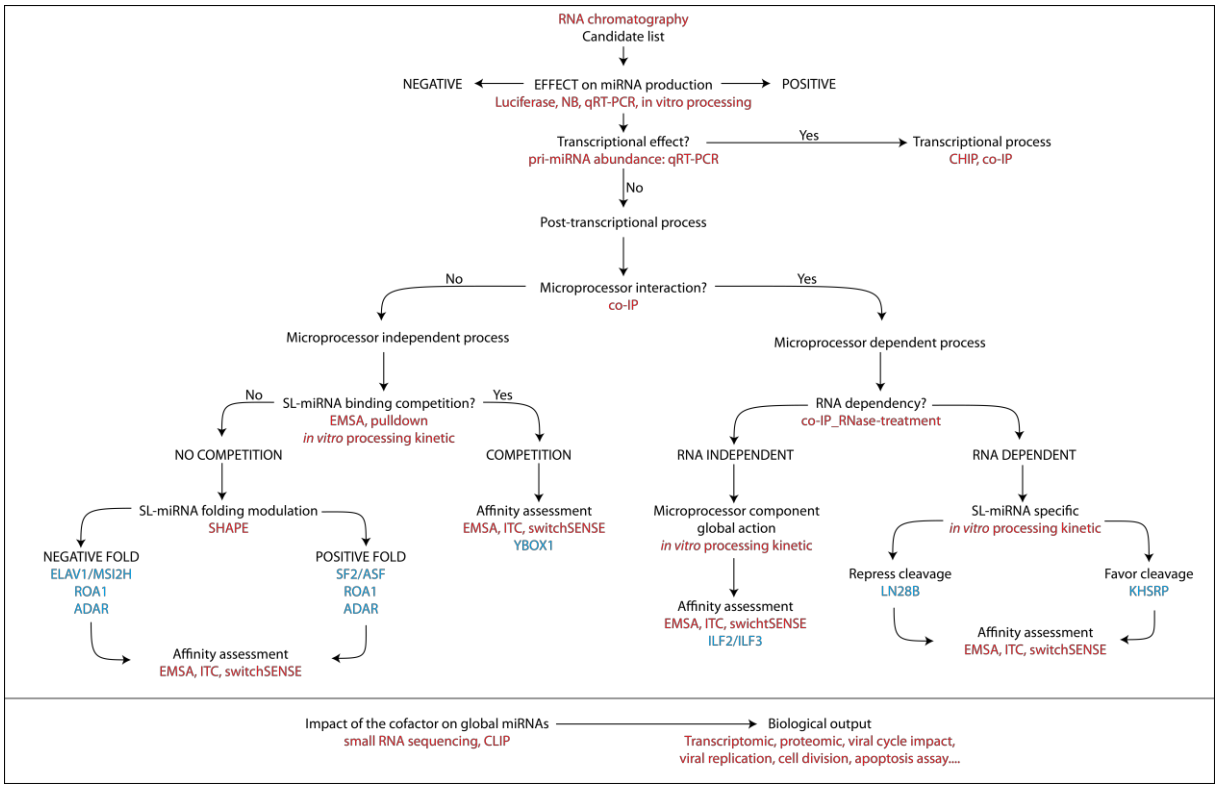


Figure 12. Représentation des expériences nécessaires pour évaluer l'implication des cofacteurs dans la biogenèse spécifique de miARN. Les approches techniques sont notées en rouge. Les exemples de protéine sont notés en bleu.

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Antoine CREUGNY

Identification and functional characterization of proteins involved in the regulation of viral microRNAs

Les microARNs sont des régulateurs post-transcriptionnelles de l'expression géniques. Partagé par tous les eucaryotes et certains virus, la compréhension de leur régulation est indispensable pour comprendre la physiologie cellulaire. Les précurseurs de microARN peuvent être exprimé individuellement ou en cluster au sein d'un transcrit dit primaire. Les caractéristiques structurales et de séquences de ces précurseurs ne sont cependant pas suffisante pour comprendre leur maturation différentielle. Ce projet porte sur la recherche de cofacteur protéique impliqué dans la maturation spécifique de miRNA exprimés par l'herpesvirus humain de type 8 au sein d'un transcrit primaire possédant 10 précurseurs de miARN. Des chromatographies à l'ARN ont été réalisé pour identifier de potentielles protéines impliquées dans la maturation spécifique des miARN du virus. Des outils de cribles, d'expressions et d'analyses ont été mis en place pour commencer à décrire le mécanisme moléculaire sous-jacent à la maturation spécifique de précurseur de microARN.

Mots-clés : microARN cluster, herpesvirus, régulation post-transcriptionnelle

MicroRNAs are post-transcriptional regulators of gene expression. Shared by all eukaryotes and some virus, deciphering their regulation is essential to understand cellular physiology. microRNA precursors can be expressed individually or in cluster within a so-called primary transcript. However, structural and sequence features of these precursors are not sufficient to describe their differential maturation. This project involves the search for a protein cofactor involved in the specific maturation of miRNA expressed by the human herpesvirus type 8 within a primary transcript possessing 10 microRNA precursors. RNA chromatographies have been carried out to identify potential proteins involved in the specific maturation of virus micoRNAs. Screening, expression and analysis tools have been put in place to begin describing the molecular mechanism underlying specific miRNA precursor maturation.

Keywords: microRNA cluster, herpesvirus, post-transcriptional regulation