CG7009 and CG5220 - novel tRNA methyltransferases linking tRNA biogenesis to the regulation of the sncRNA pathways

Margarita Angelova

Sorbonne Université

École doctorale 515 - Complexité du Vivant

IBPS -UMR7622 Laboratoire de biologie du développement SU CNRS

Génétique et Epigénétique de la Drosophile

CG7009 and CG5220 – novel tRNA methyltransferases linking tRNA biogenesis to the regulation of the sncRNA pathways

Par Margarita Angelova

Thèse de doctorat de Biologie

Dirigée par Clément Carré

Présentée et soutenue publiquement le 17 septembre 2018

Devant un jury composé de :

Dr. Severine CHAMBEYRON	Directrice de recherche	Rapporteur
Pr. Phillip D. ZAMORE	Professeur	Rapporteur
Dr. Evelyn HOULISTON	Directrice de recherche	Présidente du jury
Dr. Matthias SCHAEFER	Maître de Conférences	Examinateur
Dr. Jean-Yves ROIGNANT	Chercheur	Examinateur
Dr. Christophe ANTONIEWSKI	Directeur de recherche	Membre invité
Dr. Clément CARRÉ	Maître de Conférences	Directeur de thèse

ACKNOWLEDGMENTS

I would like to spend a line in order to acknowledge all the flies that were sacrificed for me to have a PhD.

I also want to thank all the jury member for accepting to evaluate my work.

Most personally I want to thank my mom Krasi and my Dad Peter who give me the example who and how I should be. If it weren't for their guidance I wouldn't have made it to the PhD defense. I also want to thank my other dad Todor, as he is the reason I was able to come to France and to fulfill my dreams. Thank you for being a great dad.

I would like to thank to my supervisors Clément and Christophe for giving me the opportunity to work on this project.

Clément, thank you for guiding me throughout these 4 years and for your precious advises. In our hard moments I learned a lot of things I like to do and also things I don't think are good to do. This allowed me to acquire knowledge through this PhD experience not only on biology and experiments but also on communication and character management.

Christophe, thank you for your advices and for your extremely exact and consistent comments, which made me think a lot.

I want also to mention Bruno for being a positive an easy to work with colleague, as well as for the tremendous amount of work he performed on the project: flies' crosses and dissections, survival tests, gypsy-LacZ staining, and many other.

Dilyana, thank you for being a good listener and a wise adviser. Thank you also for the experimental help during our co-working on the project and for your help with RNA extraction for Mass spectrometry analysis.

Cyrinne, thank you for being the most motivated and amongst the brightest students I ever met as well as for your work on the RTL-P experiments and for the co-IP experiments of CG7009 and CG33172.

I want to also acknowledge the work of Souraya and Bruno on the automiW experiments, and Souraya for being such a good friend and for all the moral support when I had hard times during my PhD.

Guillaume Brulère, my first student for the second hit mutation characterization and for cloning experiments of CG33172.

Marius, I want to acknowledge all the super interesting discussions we had, you are a truly deep source of ideas. Thank you also for your precious advises in the beginning of my PhD and more specifically for your help with Galaxy.

Safia, my first mentor during the whole first year of the PhD experience. Thank you for your patience, thank you for staying objective, thank you for all the experiments you thought me.

I want to mention all our collaborators and especially Jean-Yves Roignant, Damien Brégeon and Matthias Schaefer with whom I had a more important communication.

Jean-Yves, thank you for travelling three times to France for my PhD: for two thesis committees and for the defense. Thank you also for the experiments you and your co-workers performed in the terms of the project. Thank you also for the opportunity you gave me to participate in the writing of a review article. But most importantly thank you for the discussions and for the time you spent to share your experience with me, which thought me many things.

Matthias, thank you for your precious advices and for your guidance during my PhD on every occasion we met. Thanks for accepting to be part of my professional development. Thank you of course for all the experiments on the project, all the CRISPR mutagenesis and of course the precious Northern blots. I also want to thank you for always being so reactive and for your willingness to help and correct me.

Damien, thank you for accepting to work on the project and for providing us with the first proof of the molecular activity of the proteins of interest. Thank you also for precious discussions and advises during my PhD.

Yuri and Virginie, thank you for the easy communication and especially for this great and very fruitful collaboration on the RiboMeth seq experiments which allowed us to answer so many questions and to unravel so many new unexpected findings at the very end of my PhD.

Jozef Gecz, whose kind gift of LCL derived from patients of the human ortholog of CG7009 and CG5220 are now allowing us to test if the processes regulated by CG7009 and CG5220 and uncovered in the terms of my PhD are evolutionary conserved in humans.

I want to thank my tutor Jean Antoine Lepesant, and also Laure Teysset, Antoine Boivin and Stéphane Ronsseray for all their help and guidance during my PhD.

I also want to acknowledge all the Drosophila's community at the UMR7622, with special thanks to Hélène Thomassin for her precious help on Northern blot experiments.

I also want to thank our secretaries Isabelle and Hafida for saving me in so many occasions, as well as Sylvie, the director of our unit for her implication.

Last but not least, I also want to thank Valérie Biou, my first internship mentor, who was the first person to believe in me and to give me an inspiring example of academic research.

ABSTRACT

Small non-coding RNA-mediated silencing is a widespread mechanism of genetic repression amongst eukaryotes. Deregulations of the different pathways of small RNA silencing have been linked to severe phenotypes and pathologies in many species. The three best-characterized small RNA silencing pathways are the small interfering RNA (siRNA), the microRNA (miRNA) and the piwi-interacting RNA (piRNA) pathways.

It has been shown that some post-transcriptional RNA modifications, including on small non-coding RNAs and on tRNAs, are important for their activity and/or stability. During my PhD I characterized the molecular function of two 2′-O tRNA methyltransferases encoded by the genes *CG7009* and *CG5220* in *Drosophila melanogaster* which seem to be involved in the different small RNA silencing pathways and which are conserved from yeast (Trm7) to human (FTSJ1). Mutant flies have an ovarian size reduction phenotype, decreased resistance to DCV viral infection, and differential transposable element expression levels in somatic and germinal tissues. Mutations in their yeast ortholog Trm7 are linked to a severe growth defect phenotype and in their human ortholog, FTSJ1, to non-syndromic X-linked intellectual disability (NSXLID).

I confirmed that *CG7009* is involved in both the miRNA Ago-2-dependent and in the siRNA pathways in the *Drosophila melanogaster* S2 cell line and in developing flies. I also discovered that CG5220, a paralog of CG7009, is involved in the miRNA Ago-2-dependent pathway in flies, similarly to CG7009. Furthermore, we also showed that CG7009 and CG5220 have a role in the somatic piRNA pathway. Last but not least, I was able to characterize that CG7009 lack of function, leading to loss of 2′-O methylation (Nm) on $tRNA^{Phe}$, is accompanied by differential tRNA fragmentation profiles and tRNA fragments (tRFs) accumulation. tRFs are stable RNAs derived from the cleavage of mature tRNAs, that have been detected in bacteria, plants, yeast, flies and humans and linked to a variety of functions. Recent research point to tRFs as a novel class of sncRNAs. Our results link, for the first time, the loss of 2′-O tRNA modifications with tRFs accumulation and the regulation of the sncRNA pathways.

Index

AC: anticodon
Ago-1: Argonaute 1 (protein)
Ago-2: Argonaute 2 (protein)
Ago-2: Argonaute 2 (gene)
Ago-2: *Ago-2* (gene, mutated)
Ago-3: Argonaute 3 (protein)
Ap: *Apterous (*ap[Xa]: Apterous-Xa linked chromosome)
Aub: Aubergine (protein)
BAC: Bacterial Artificial Chromosome
bp: base pair
cDNA: complementarity DNA
CDS: Coding DNA Sequence
C. elegans: Caenorhabditis elegans
CyO: CurlyO
D: dihydrouridine
Da: dalton
Dcr-1: Dicer-1 (protein)
Dcr-2: Dicer-2 (protein)
Dcr-2: Dicer-2 (gene)
dcr-2: Dcr-2 (gene, mutated)
D. melanogaster: *Drosophila melanogaster*
DmHen1: *Drosophila melanogaster* HEN1 (protein)
DmHen1: Drosophila melanogaster HEN1 (gene)
Dmhen1: DmHen1 (gene, mutated)
DNA: deoxyribonucleic acid
DNase: DNA nuclease
Dnmt2: DNA methyltransferase 2
dNTP: Deoxynucleotide triphosphate
dsRNA: double-stranded RNA
DTT: dithiothreitol
E. coli: *Escherichia coli*
EDTA: Ethylenediaminetetraacetic acid
endo-siRNA: endogenous small interfering RNA
EtBr: Ethidium bromide
EtOH: ethanol
FW: forward
FTSJ1: FtsJ RNA methyltransferase homolog 1
Ftz: fushi tarazu
gDNA: Genomic DNA
GDP: The Gene Disruption Project
GFP: Green Fluorescent Protein
GST: Glutathione S-transferase
hc-siRNA: heterochromatic small interfering RNA
HEN1: HEN1 (protein)
Hen1: HEN1(gene, mutated)
H2O: water
IP: immunoprecipitation
iPCR: Inverse Polymerase Chain Reaction
Kb: kilobase
KD: knock down
KDa: kilodalton
LB: Luria Bertani broth
Loqs: Loquacious protein
LTR: long terminal repeat
mg: milligram
MgCl2: magnesium chloride
mg/ml: milligram per milliliter
min: minute
miRNA: microRNA
mL: milliliter
mM: millimolar
mRNA: messenger RNA
MW: molecular weight
ncRNA: non-coding RNA
nm: nanometer
Nm: 2'-O-methylation
Nsun2: NOP2/Sun RNA methyltransferase family member 2
nt: nucleotide
OD: optical density
ON: over night
ORF: open reading frame
PBS: Phosphate buffered saline
PCR: Polymerase Chain Reaction
piRNA: PIWI-interacting RNA
Piwi: PIWI protein
pre-mRNA: precursor messenger RNA
pre-miRNA: precursor miRNA
pri-miRNA: primary miRNA
PTGS: post-transcriptional gene silencing
RdRP: RNA-dependent RNA polymerase
Rev: reverse
RISC: RNA-induced silencing complex
RLC: RISC loading complex
RNA: ribonucleic acid
RNAi: RNA interference
RNase: ribonuclease
RpL17: Ribosomal Protein L17 (gene)
rpm: rotation per minute
rRNA: ribosomal RNA
RTLP: Reverse Transcription at Low concentration of dNTP followed by PCR
RT-PCR: Reverse transcription PCR
s: second
SAH: S-adenosyl homocysteine
SAM: S-adenosyl methionine
Sb: Stubble
SDS: Sodium dodecyl sulfate
SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser: serrate
siRNA: small interfering RNA
sncRNA: small non-coding RNA
snoRNA: small nucleolar RNA
ssRNA: single-stranded RNA
S2 cells: Drosophila Schneider 2 cells
TAE-Tris-Acetate EDTA electrophoresis buffer
Tb: Tubby
TGS: transcriptional gene silencing
Tris: tris(hydroxymethyl)aminomethane
Trm7: tRNA methyltransferase 7
tRF: tRNA-derived fragment
tRNA: transfer RNA
T7: bacteriophage T7
tRNA: transfer RNA
UV: ultra violet
U/ml: unit per milliliter
V: Volt
v/v: volume per volume
WT: wild type
w/v: weight per volume
X-gal:5-bromo-4-chloro-3-indolyl-**β**-D-galactopyranoside
X-linked: X chromosome linked
3'UTR: 3'- untranslated region
µg: microgram
µg/ml: microgram per milliliter
µL: microliter
µM: micromolar
°C: degree Celsius
Ψ: pseudouridine

Table of Contents

INTRODUCTION: Coding or non-coding? Layers of information

What is non-coding RNA? Today we acknowledge that the simple distinction between *coding*, referring to protein coding RNA, and *non-coding*, referring to the rest of the RNA classes, is far from sufficient to give an accurate idea on the complexity of the RNA world. Bacterial, fungal, plant, archaea, insect and human genomes are composed mainly of non-protein coding transcripts (Palazzo and Lee, 2015). Amongst the first well characterized non-coding RNAs (ncRNAs) were transfer RNAs (tRNA), ribosomal RNAs (rRNA), small nuclear RNAs (snRNA) and small nucleolar RNAs (snoRNA). More recently, multiple other classes of non-coding RNAs were characterized, including the small non-coding silencing RNAs. In a fascinating way, the complexity of functions of RNA in the cell cannot be fully explained solely by the discovery of these novel RNA species. Discoveries were also made of *non-coding* functions carried by parts of *coding* RNAs (*e.g.* circular RNAs built from exonic mRNA sequences (Memczak et al., 2013)). A concept of "levels of utilization" of the genetic information has been introduced.

Furthermore, several levels of information within the same *non-coding RNA* can also exist. A good example is tRNAs, as they serve at least once as adaptors during translation, but some of them can also serve a second time after fragmentation to tRFs (tRNA-derived fragments), molecules having a variety of possible functions that are distinct from the function of the mature tRNA they derive from (see below). A striking example of 3 different molecules, arising from the same tRNA, each with unique function, was described recently by Schorn and colleagues (Schorn et al., 2017; Schorn and Martienssen, 2018). The authors showed that in mouse embryonic fibroblasts (MEFs) the same tRNA molecule can serve in translation, as well as in the production of 2 tRFs with distinct sizes and functions. As of today, multiple studies witness that the information encoded in the same DNA sequence can be used in different fashions. It has become very clear in the last several years that another layer of information exists in RNA. Like a secondary information code, similarly to the genetic code, or the histone and DNA modification epigenetic code, the field of *Epitranscriptomics* which is expanding very fast, has certified that and enormous and yet to be discovered layer of information is hidden in RNA modifications. In the first chapter of the introduction I will introduce the three main small non-coding RNA (sncRNA) pathways, implicated in gene regulation. In the second chapters I provide more information on RNA modifications and on tRFs, as a novel class of small non-coding regulatory RNAs, as they are central to my research and at the core of the studied topic: regulation of the sncRNA pathways and epitranscriptomics connection *via* tRFs.

Chapter I: Gene expression repression by small non-coding RNAs

Three classes of small non-coding RNAs (sncRNAs) emerged as core components of a vast gene silencing framework in metazoan. They fulfil critical roles as transcriptional and post-transcriptional regulators of many RNA targets and are involved in numerous biological processes, including defense against pathogens, repression of transposable element, chromatin dynamics, development, cell proliferation and longevity. SncRNAs, namely miRNAs (microRNA), siRNAs (small interfering RNA) and piRNAs (*piwi*-interacting RNA), have distinct biogenesis pathways but they all associate with members of the Argonaute protein family. The repression by small-non-coding RNAs in the way we know it today is accomplished by nucleotide complementarity between the target and the small RNA regulator, which allows the small RNA to serve as a guide for its associated protein complexes. Argonaute proteins form the core of central enzyme complexes, the RNA-induced silencing complexes (RISC), which mediate target regulation in small RNA silencing pathways (Hammond et al., 2000). sncRNA pathways trigger post-transcriptional repression which operates either through endonucleolytic cleavage (si- and piRNA), or by repression of translation or direct degradation (miRNA) of the target RNA. However, it has been shown that *Drosophila* Ago-2 loaded small RNAs can also trigger translational repression, similarly to miRNAs under certain conditions (Iwasaki et al., 2009). Moreover, si- and piRNA-loaded RISCs are also involved in transcriptional repression on nascent RNAs, by guiding chromatin-modifying complexes to targeted DNA loci (for review, see (Huang et al., 2013).

Small RNAs can be classified according to their size, their substrates, the proteins involved in their biogenesis, their effector protein partners and their modes of target regulation. These RNAs have several common characteristics: a short length (~20 to 30 nucleotides), a terminal 5'-phosphate and a 3'-hydroxyl groups, albeit the 3’-end can sometimes be modified (see below) and their specific association with members of the Argonautes protein family. As the three pathways are present in flies, *Drosophila melanogaster* is a good model organism to study the sncRNA pathways.

1. Discovery and function of small non-coding regulatory RNAs

A. microRNAs

The first small non-coding RNA was discovered in 1993 by Victor Ambros and colleagues in a screen designed to uncover genes implicated in the post-embryonic development of *Caenorhabditis elegans* (*C. elegans*) (Lee et al., 1993). The authors discovered that the *lin-4* locus, which is implicated in the temporal control of the transition between pupal stages L1 and L2, does not encode a protein but produces two small non-coding RNAs with an approximate length of 22 and 61 nucleotides (nt), respectively. As they shared the same nucleotide sequence at their 5'-end, it seemed that the longer RNA was the precursor of the shorter one. Moreover, the small RNA of 22nt was complementary to multiple sites in the 3'-untranslated region (3'UTR) of the *lin-14* gene. Interestingly, a loss-of-function of *lin-4* induced developmental defects comparable to a gain-of-function of *lin-14*. These observations led to the proposal of a model in which the small RNA *lin-4* is inhibiting the expression of the *lin-14* mRNA in a sequence-specific manner by base pairing with the 3'UTR of this regulatory target. This work and the discovery of another small RNA, *let-7* (Reinhart et al., 2000), revealed for the first time the existence of a genetic regulatory mechanism performed by small non-coding RNAs. In the years that followed, this kind of small non-coding RNAs were discovered in other organisms, in particular in *Drosophila* and humans (Lagos-Quintana et al., 2001). They were baptized microRNAs (miRNAs).

B. small-interfering RNAs

In parallel to the discovery of miRNA, Fire and Mello made a discovery in 1998, which reversed the, by then, largely accepted view that antisense RNAs exert their silencing effects by base pairing with their mRNA counterparts, thereby inhibiting their translation. This led to the discovery of double-stranded RNAs (dsRNA) as the trigger for silencing in *C. elegans* (Fire et al., 1998). A Nobel Prize acknowledged their work in 2006. It was then discovered in plants that dsRNAs of exogenous origin, like viral RNAs, are sliced into small interfering RNAs (siRNAs) with a length of approximately 21 nt (Hamilton and Baulcombe, 1999). This work also marks the discovery of the first siRNA. The first siRNAs of endogenous origin (endo-siRNAs) were also detected in plants (Hamilton et al., 2002, 2015) and later discovered in *C. elegans* (Ambros et al., 2003), flies (Czech et al., 2008; Ghildiyal et al., 2008) and mammals (Maillard et al., 2013) suggesting their ubiquity among higher eukaryotes (reviewed in (Ghildiyal and Zamore, 2009)).

siRNAs function on complementarity principle as guides for their bound Argonaute protein partner towards the target mRNA for its endonucleolytic degradation (Hammond et al., 2000; Zamore et al., 2000). This mechanism of post-transcriptional gene silencing (PTGS) was initially baptized RNA interference (RNAi).

C. piwi-interacting RNAs

A third class of small non-coding RNAs, *piwi*-interacting RNAs (piRNAs) with a length of 23-29 nt was discovered more recently in flies (Aravin et al., 2001; Vagin et al., 2006). As their name implies, piRNAs bind to the Piwi clade of Argonaute proteins (see below). Thus far, piRNAs and Piwi family proteins have been extensively studied in *Drosophila* species and in mice. Several genome-wide screens in flies led to the discovery of multiple factors involved in piRNA biogenesis or piRNA-mediated silencing steps of the pathway (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). piRNAs have sequence complementary to transposable elements (TEs) and their expression in the germlines repress TE expression. Therefore, piRNAs ensure the maintenance of genomic stability and the preservation of genomic integrity in the germ line for the future generations *via* the repression of transposable elements (Brennecke et al., 2007). *Drosophila* mutants for any of the PIWI proteins are characterized by TEs derepression, suggesting that the three PIWI proteins are not redundant in their functions (reviewed in (Yamashiro and Siomi, 2018)).

A recent study on piRNAs in arthropods revealed that their germline-restricted function might not be as universal as previously thought, as piRNAs with somatic functions have been described in arthropods (Lewis et al., 2018). piRNA have been studied in other species from basal metazoans like the distant relatives of the *Bilateria* (*i.e.* the poriferan *Amphimedon queenslandica*, a sponge) and close relatives (*i.e.* the cnidarian *Nematostella vectensis*, starlet sea anemone), to bilaterian animals such as worms, zebrafish, mice, rats and humans. piRNAs have been shown to play a role in the development in planarians. In Aplysia, piRNAs have been linked to long-term memory development. In mosquitos, the piRNA pathway has a function in antiviral defense. These supplementary functions of piRNAs have been recently reviewed in (Yamashiro and Siomi, 2018).

2. General characteristic of the small RNA silencing pathways

Repression events occurring in the three small RNA silencing pathways in flies can be summed up in the following consecutive steps: (i) small RNA biogenesis, (ii) loading into an Argonaute protein (iii) strand selection, (iv) target recognition and effector function. On the mechanistic molecular level however, there are some notable differences between the three pathways. Unlike miRNA and siRNA, which are produced from double-stranded precursors, piRNAs derive from single-stranded precursors (Vagin et al., 2006). piRNAs are produced through alternative biogenesis mechanism of miRNAs and siRNAs, independently of RNase III enzymes (Brennecke et al., 2007; Vagin et al., 2006). Indeed, in contrast to the two other small non-coding RNA pathways, the piRNA pathway looks completely Dicer-independent. As piRNAs derive from ssRNA precursors, not duplexes but ssRNAs are loaded in PIWI-proteins. In flies, the miRNA and the siRNA pathways have members of the Ago clade of Argonaute proteins – Ago-1 and Ago-2 respectively - for effectors, whereas piRNAs are loaded into Piwi clade proteins -PIWI (here named Piwi), Argonaute 3 (Ago-3) and Aubergine (Aub).

The 5'-end of mi-, si- and piRNAs contains a monophosphate at their 5'-end, a modification that is characteristic for the three sncRNAs. However, piRNAs often start with uridine (1U bias) (Brennecke et al., 2007; Gunawardane et al., 2007; Malone et al., 2009; Saito et al., 2006; Siomi et al., 2011). The 3'-end of piRNAs is 2'-O-methylated (Nm) by the methyltransferase DmHen-1/ Pimet (Horwich et al., 2007; Saito et al., 2007). This is similar to siRNAs and Ago-2-loaded miRNAs, but not to Ago-1-loaded miRNAs in flies (Saito et al., 2007; Vagin et al., 2006). Nm addition increases the stability of the small RNAs (Horwich et al., 2007; Yu et al., 2005) and differently to *Drosophila*, plants miRNA also bear an Nm at their 3'-extremity (Li et al., 2005; Yu et al., 2005).

It was initially discovered in mammals and in *C. elegans* that the same dsRNA-specific RNase III family ribonuclease, the protein Dicer (Dcr) is required for the maturation of si- and miRNAs (Bernstein et al., 2001; Knight and Bass, 2001). However, *Drosophila* species have evolved two Dicer proteins: Dcr-1 is involved in miRNAs biogenesis, whereas Dcr-2 is specialized in siRNA production (Lee et al., 2004b). The fly siRNA pathway has an important role in the antiviral defense and the divergence of Dicer into two isoforms could serve to decrease the competition for Dicer between precursors of miRNAs (pre-miRNAs) and dsRNAs of viral origin (reviewed in (Ghildiyal and Zamore, 2009)). Alternatively, this specialization of factors that are specific for the siRNA pathway (Dcr-2, as well as the Argonaute protein family member Ago-2) suggests the existence of a selection pressure on the siRNA pathway to respond to rapidly evolving

anti-RNAi viral strategies. This is supported by the fact that *Dcr-2* and *Ago-2* belong to the most rapidly evolving genes in *Drosophila* (Obbard et al., 2006).

3. Argonaute proteins, universal components of RNA silencing related mechanisms

Drosophila has five different Argonaute proteins; Ago-1 and Ago-2 belong to the Ago clade of Argonaute proteins and are ubiquitously expressed, whereas Piwi, Aub and Ago-3 are members of the PIWI clade and are expressed specifically in the germline (Cox et al., 2000; Harris and Macdonald, 2001; Siomi et al., 2011; Vagin et al., 2006). In the fly ovaries, Piwi family members have distinct expression patterns: Piwi is expressed in germline cells as well as in somatic ovarian follicle cells that surround the developing egg chamber, whereas Aub and Ago-3 are expressed only in germline cells (Brennecke et al., 2007). The three proteins also differ in their subcellular localization. The Piwi protein has a nuclear subcellular localization in *Drosophila* germ cells and adjacent somatic cells, whereas Aub and Ago-3 in the germline are cytoplasmic and found in a specialized perinuclear ring called *nuage*, which is associated with germline function. The N-terminus of the Piwi protein, as well as its loading with piRNAs, are involved in its nuclear localization, and unloaded Piwi proteins have been detected in the cytoplasm, where the biogenesis of piRNAs take place (Olivieri et al., 2010; Saito et al., 2010). A third clade of Argonaute proteins, the WAGO clade, is specific for *C. elegans* (Gu et al., 2009).

Argonaute proteins have RNA-binding domains (PAZ and MID domains) and a RNase H similar PIWI domain, which performs the endonucleolytic cleavage of the target mRNAs (Figure 1) (recently reviewed in (Olina et al., 2018)).

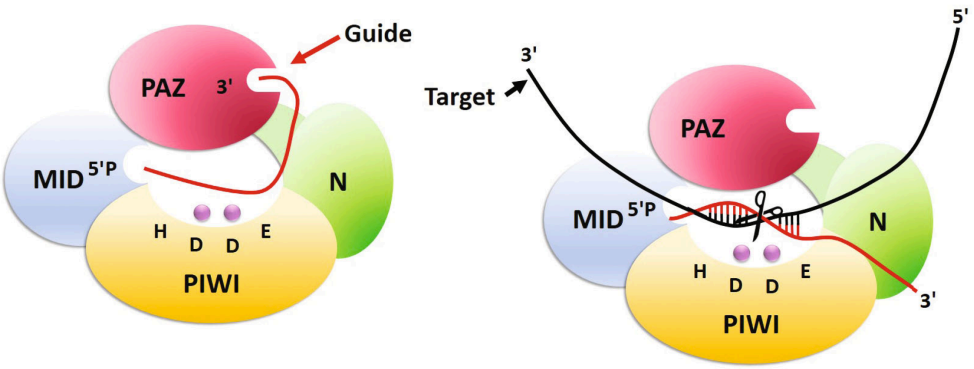

Figure 1. Schematic representation of an Argonaute protein. Argonaute protein loaded with a small RNA (left part) or a small RNA bound to its target (right part). Represented are the four distinct domains of Argonaute proteins (MID, PIWI, PAZ and N-domains), with the catalytic tetrad in the RNase H similar PIWI domain. Purple dots indicate two coordinated magnesium ions, important for the catalytic activity.

A difference between PIWI and AGO proteins is an arginine-glycine enriched region near the N-terminus of PIWI proteins (reviewed in (Yamashiro and Siomi, 2018)). Some of the arginines get symmetrically dimethylated, to produce symmetrically dimethylated arginines (sDMAs). These sDMAs are required in Aub and Ago-3 for binding with Tudor domain-containing piRNA factors, some of which are required for piRNA biogenesis (reviewed in (Yamashiro and Siomi, 2018)).

Typically, it is considered that Ago-1-loaded miRNAs exert their function by mRNA destabilization and translational repression of their target mRNAs, whereas Ago-2 loaded siRNAs induce target degradation. However, target regulation by small RNAs seems to be much more complicated: (i) Ago-2 is the only mammalian Ago-clade protein (out of 4 Argonautes) thought to mediate target cleavage, whereas in flies Ago-1 also has a slicer activity, although less pronounced that the one of Ago-2 (Miyoshi et al., 2005) (ii) a part of the miRNAs in flies can be loaded alternatively on Ago-2, the typical effector of the siRNA pathway (Forstemann et al., 2007), providing a proof of the interconnection between these two small RNA silencing pathways.

4. Biogenesis and mechanism of action of the sncRNAs

A. Biogenesis and mechanism of action of siRNAs

dsRNAs are cleaved into double-stranded siRNAs by Dicer (Bernstein et al., 2001). In flies, the Dcr-2 protein is responsible for the processing of long dsRNA (Fig.1). siRNA comprise two ~21nt strands (a duplex) with a 5'-phosphate and a 3′-hydroxyl group and two-nucleotide overhangs at the 3′-ends, a hallmark of RNase III processing (Zamore et al., 2000). The guide strand is used to guide the silencing machinery while the other strand, the passenger strand, is cleaved by Ago-2 (Matranga et al., 2005). The relative thermodynamic stabilities of the 5′-ends of the two siRNA strands in the duplex determines the identity of the guide and passenger strands (Khvorova et al., 2003). In flies, the sensor of this thermodynamic difference is R2D2, a dsRNA-binding partner of Dcr-2 restricting the substrate specificity of Dcr-2 (Cenik et al., 2011). R2D2 forms the RISC loading complex (RLC) with Dcr-2, which loads siRNA duplexes into Ago-2 (Liu et al., 2003).

A general principle for Argonaute-mediated, small RNA-guided silencing, is that Ago-2 cleaves its target at the phosphodiester bond between the nucleotides that are paired to nucleotides 10 and 11 of the guide strand (Elbashir et al., 2001). The same authors also described the minimal RISC: a recombinant Ago-2 protein associated with siRNA is sufficient for slicer activity. After the release of the passenger

strand, pre-RISC is converted to mature RISC (ssRNA-loaded guide strand). In flies, the guide strand is 2′-O-methylated at its 3′-end by the DmHen1/Pimet methyltransferase in an S-adenosyl methionine (SAM) dependent manner, completing RISC assembly (Horwich et al., 2007; Saito et al., 2007). As mentioned, in plants, both miRNAs and siRNAs are terminally methylated, a modification that is crucial for their stability (Yu et al., 2005). This paper is one of the first reports of small silencing RNA methylation, raising the idea that small RNA stability and function could be controlled by post-transcriptional modification.

In plants, as in worms, endo-siRNAs are generally produced through the action of RNA-dependent RNA polymerases (RdRPs). However, endo-siRNAs have been identified in flies and in mammals even though their genomes do not encode such enzymes. In the germline and in somatic tissues of *Drosophila* species the endo-siRNA population comprises small RNAs of ~21nt length (Czech et al., 2008). They are present in sense and antisense orientations, have modified 3′-ends and are produced in a Dcr-2-dependent manner, although in the absence of Dcr-2 a remnant of the population persists. Fly endo-siRNAs derive from transposons, heterochromatic sequences, intergenic regions, transcripts with extensive structures and from some mRNAs. Expression of some transposon mRNAs increases in both *dcr-2* and *ago-2* mutants, implying an endogenous siRNA pathway in the silencing of transposons in flies. A subset of fly endo-siRNAs derives from RNA transcripts that fold into structured intramolecularly paired hairpins (Czech et al., 2008). Accumulation of these siRNAs requires as dsRNA-binding partner, an isoform of the protein Loquacious (Loqs-PD) and not R2D2, the usual partner of Dcr-2.

The most abundant endo-siRNAs in plants are hc-siRNAs (heterochromatic siRNAs): ~24 nt long and methylated by HEN1. They originate from transposons, repetitive elements, tandem repeats and promote heterochromatin formation by directing DNA methylation and histone modification at the loci from which they originate (Hamilton et al., 2002, 2015).

B. Biogenesis and mechanism of action of miRNAs

miRNAs are encoded in the genome and are transcribed as independent units, many of which consist of clusters containing multiple miRNAs that are transcribed as polycistronic units (Lagos-Quintana et al., 2001). Transcription of miRNA genes by DNA-dependent RNA polymerase II yields primary miRNA (pri-miRNA) transcripts, which are capped and polyadenylated and have local stem–loop structures, as well as sometimes introns (Lee et al., 2004a). Generation of ~ 22 nt mature miRNAs from pri-miRNAs requires the sequential action of two RNase III endonucleases, assisted by their dsRNA-binding partner proteins (Figure 2). Processing begins in the nucleus where Pasha (in flies) or DGCR8 (in mammals) bind to the pri-

miRNA and recruit the RNase III enzyme Drosha to form the Microprocessor multiprotein complex. The pri-miRNA is cleaved by Drosha to liberate a ~ 60–70-nt short precursor miRNA (pre-miRNA) with a hairpin structure and a 2-nt single-stranded 3′-overhang, indicative of its RNase III-mediated production (Lee et al., 2003). Alternatively, miRNAs can be maturated in a Drosha-independent manner. pre-miRNAs of mirtrons, miRNAs that are present in the introns of pre-mRNAs, are generated by the cellular splicing machinery (Okamura et al., 2007).

Once generated, pre-miRNAs are exported to the cytoplasm *via* the nuclear pore complex, were the processing of pre-miRNAs to mature miRNAs occurs. As mentioned above, in flies Dcr-1 is responsible for the processing of the pre-miRNA into a ~22–23-nt miRNA:miRNA* duplex corresponding to the two sides of the base of the stem (Bernstein et al., 2001; Lee et al., 2004b; Zamore et al., 2000). miRNA:miRNA* duplexes generated by Dcr-1 and its dsRNA binding partner Loqs-PB, an isoform of the protein Loquacious exhibit the RNase III cleavage characteristic 2-nt single-stranded 3′-overhangs at both ends. The two duplex strands correspond to the guide and passenger (star *) strands of an siRNA and similar thermodynamic criteria influence the choice of miRNA versus miRNA* (Khvorova et al., 2003). miRNAs can arise from either arm of the pre-miRNA stem, and some pre-miRNAs produce mature miRNAs from both arms, whereas others show such pronounced asymmetry that the miRNA* is rarely detected, even in high-throughput sequencing experiments.

The majority of miRNAs bind to their targets only through a small sequence located at the miRNA 5'-end, the "seed region" which contributes most of the energy for target binding. Argonaute proteins display nucleotides 2–8 of the seed, defining this region as the primary specificity determinant for target recognition. The small size of the seed, as well as the presence of central mismatches within this region, increases the number of different genes that a single miRNA can regulate and thus make the prediction of targets still difficult today.

In flies, the extensive double-stranded character of a small RNA duplex, as found in siRNAs, directs it into Ago-2, whereas central mismatches, like those found in the majority of miRNA/miRNA* duplexes, direct duplexes into Ago-1. Central to this sorting decision is the affinity of the small RNA duplex for the Dcr-2/R2D2 heterodimer, which loads small RNAs with perfect complementarity into Ago-2 (Liu et al., 2003). The function of the Dcr-2/R2D2 heterodimer in Ago-2 loading is separate and distinct from its role in dicing siRNAs from long dsRNA. Indeed, a mutant form of the protein that lacks its helicase activity is unable to dice but can still load siRNA into Ago-2 (Lee et al., 2004b).

The mechanism of target regulation by a miRNA, translational repression by mRNA destabilization or degradation of target mRNAs, is influenced by the specific Argonaute partner into which the miRNA is loaded. The majority of *Drosophila* miRNAs are loaded into Ago-1. However, as mentioned above, it has been found that miRNA/miRNA* duplex strands can alternatively be sorted towards Ago-2 in a process that requires both Dcr-2 and R2D2 (Forstemann et al., 2007). Like siRNAs, mature miRNA bound to Ago-2 typically present perfect complementarity, begin with cytidine and cleave their targets. In contrast, Ago-1 bound miRNAs usually begin with a uridine and avoid the translation of the target without systematic cleavage. Unlike most of the animal miRNAs, plant miRNAs are 2′-O-methylated at their 3′-ends by HEN1 (Park et al., 2002; Yu et al., 2005). However, it has been shown that *Drosophila* Ago-2 bound miRNAs are methylated by the homolog of HEN1, the protein DmHen1 (Horwich et al., 2007), also called Pimet (Saito et al., 2007).

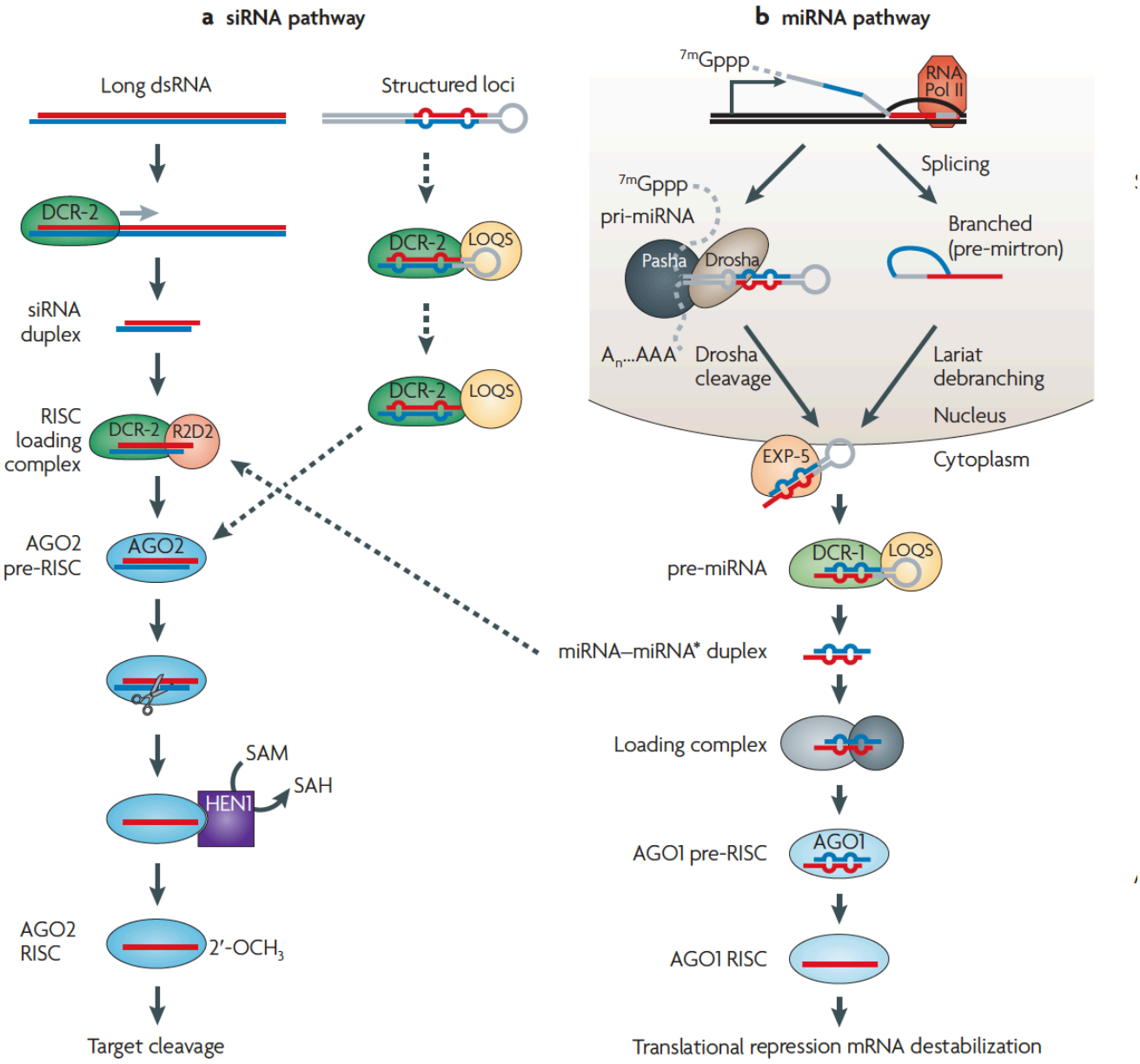

Figure 2. Schematic representation of the *Drosophila* small interfering RNA (siRNA) and microRNA (miRNA) pathways. A. dsRNA precursors are processed by Dicer-2 (Dcr-2) to siRNA duplexes composed of a guide and a passenger strand. Dcr-2 and its dsRNA-binding protein partner R2D2 form the RISC-loading complex (RLC), which loads the duplex into Argonaute2 (Ago-2). Accumulation of a subset of endogenous siRNAs (endo-siRNAs) requires another dsRNA-binding partner, the protein Loquacious (Loqs-PD**).** The passenger strand is cleaved and the guide strand, which is 2'-O-methylated by DmHen-1 directs Ago-2 to the target RNA. B. miRNAs encoded in the genome are transcribed to primary miRNAs (pri-miRNAs), which are cleaved by Drosha to yield shorter precursor miRNAs (pre-miRNAs). A subset of miRNAs, termed mirtrons, originate from introns and are processed to pre-miRNAs by the splicing machinery. Mature pre-miRNAs are exported from the nucleus to the cytoplasm and processed by Dicer-1 (Dcr-1) to mature miRNA:miRNA* duplexes which are loaded into Ago-1. One of the two strands, the miRNA strand, triggers translational repression of the target mRNAs. SAH, S-adenosyl homocysteine; SAM, S-adenosyl methionine. Image adapted from (Ghildiyal and Zamore, 2009).

C. Biogenesis and mechanism of action of piRNAs

The piRNA pathway includes three major steps: (i) piRNA cluster determination, (ii) piRNA biogenesis, and (iii) piRNA-mediated silencing. piRNAs biogenesis occurs through three mechanisms: (i) the primary piRNA pathway, (ii) the piRNA amplification loop or ping-pong cycle, and (iii) the ping-pong-cycle-dependent phased piRNA biogenesis. piRNAs show a remarkable sequence diversity but collectively the majority of *Drosophila* piRNAs map to a few hundred genomic clusters. piRNA clusters are distinct loci, comprising a complex mixture of sequences spanning large portions of transposons (Brennecke et al., 2007). However, the transposon sequences contained in piRNA clusters are fragmented and cannot transpose, this is why those clusters are often seen as transposon graveyards or catalogs of inactive transposable elements sequences. The clusters are transcribed by RNA-Pol II, they are often located in the heterochromatin of pericentromeric and telomeric regions and are divided in two categories: uni-strand and dual-strand clusters (Senti and Brennecke, 2010; Yamanaka et al., 2014). So far, all the described piRNA clusters, except the clusters in the *Drosophila* germline, are uni-strand clusters. In mammals the uni-strands clusters transcription has been characterized as uni- or bidirectional (Aravin et al., 2006; Aravin et al., 2007; Girard et al., 2006).

a. Biogenesis and mechanism of action of piRNAs in the gonadal somatic cells

In flies, the gonadal somatic cells form an envelope surrounding the ovaries and the testes. Piwi is expressed in both germ and somatic gonadal cells and as mentioned before Piwi has a nuclear localization (Brennecke et al., 2007). In the gonadal somatic cells only Piwi from the three PIWI clade proteins is expressed and it generates piRNAs by a mechanism of primary piRNA biogenesis (Figure 3).

In the gonadal somatic cells piRNAs are almost exclusively produced from uni-strand clusters (Malone et al., 2009). Uni-strand clusters contain a promoter at one of their extremity and undergo unidirectional transcription and mRNA similar processing, including 5’ capping, splicing, and addition of a poly(A)-tail. After processing, the cluster transcript is exported to the cytoplasm and processed to mature piRNAs. One of the best studied and major locus in this system is the flamenco locus. Flamenco is localized on the X-chromosome, near the centromeres and contains sequences in antisense orientation of parts of the transcripts of several transposons including *gypsy*, *ZAM*, and *idefix*. It has been shown that piRNAs processed by this locus are able to target the aforementioned transposons by complementarity and to repress their expression (Desset et al., 2003; Malone et al., 2009; Mevel-Ninio et al., 2007; Pelisson et al., 2007; Pelisson et al., 1994; Prud'homme et al., 1995). The locus is transcribed by RNA-Pol II under the

control of a non-piRNA specific transcriptional factor regulated by the Hedgehog signaling pathway Ci (Cubitus interruptus) and then processed similarly to a mRNA of protein coding genes (capped, spliced and poly-adenylated) (Goriaux et al., 2014a).

piRNA cluster transcripts must be exported from the nucleus to the cytoplasm for mature piRNA biogenesis to occur. Amongst the known export factors, Nxf1 and Nxt1 have been linked to the export of these piRNA precursors. In ovarian somatic cells, aberrant accumulation of flamenco cluster transcripts in the nucleus (Dennis et al., 2016), as well as TEs deregulation in ovaries (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013) have been observed following a KD of those two export factors. Moreover, flamenco transcript and its encoding DNA foci co-localized only in nxf1 mutant ovaries (Dennis et al., 2016), which was not the case in WT ovaries. The same was observed for mutations in proteins from the exon junction complex (EJC), implicating the splicing machinery in the piRNA pathway regulation (Dennis et al., 2016).

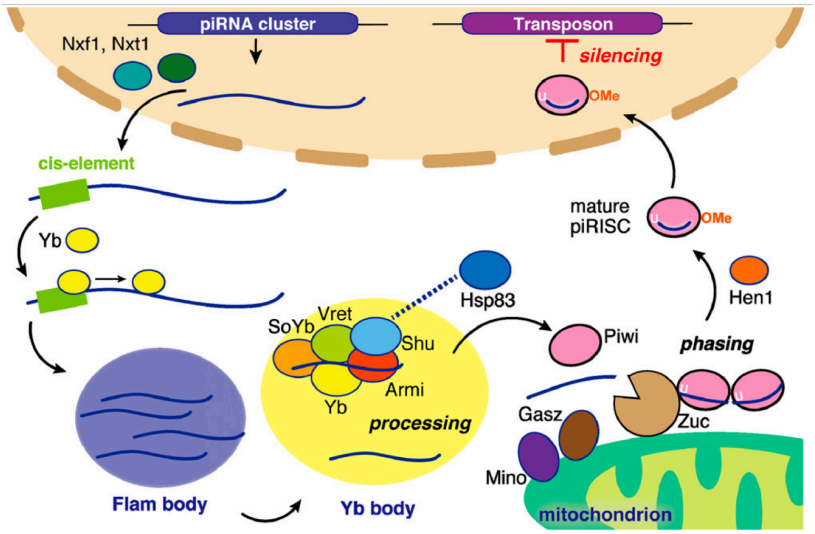

Figure 3. Schematic representation of the *Drosophila* somatic piRNA pathway. The piRNA precursor is exported from the nucleus to the cytoplasm *via* interaction with the Nxf1 Nxt1 export factors. By interacting with a cis-element on the precursor piRNA transcript, Yb associates with the precursor and participates in its sequestering in a cytoplasmic structure (Flam body), where accumulation of precursors has been described. Yb also participates in the transfer of the precursor from the Flam body to the Yb-body where piRNA biogenesis takes place. Other important for piRNA biogenesis factors in the Yb body are Armitage (Armi), Vreteno (Vret) and Sister of Yb (SoYb) and Shutdown (Shu). The interaction of the last with HSP83 is important for piRNA loading into Piwi. An important protein for piRNA biogenesis is the nuclease Zucchini (Zuc), localized on the mitochondrial membrane. On the mitochondrial surface the mitochondrial proteins Mino and Gasz are implicated in piRNA biogenesis. piRNAs in the ovarian gonadal soma have a phasing pattern, which depends on Zuc activity, and Zuc is responsible for piRNAs ends formation. Mature piwi-loaded piRNAs are 3'-terminally 2'-O methylated (OMe) by DmHen-1 (Hen-1) and translocated to the nucleus to mediate transcriptional gene silencing at transposon loci by recognition of nascent transposon RNA and induction of repressive heterochromatin histone modifications. OMe: 2'-O-methylation. Image from (Yamashiro and Siomi, 2018)

In the cytoplasm, the piRNA precursors are bound by a Tudor-domain and a DEAD-box RNA helicase domain containing protein called female sterile (1) or Yb (Figure 3). This protein forms the core of a perinuclear structure of the cytoplasm, characteristic for the ovarian somatic cells, the Yb body (Olivieri et al., 2010; Qi et al., 2011; Saito et al., 2010). The Yb body is also composed of the following proteins, necessary for piRNA biogenesis: the two Tudor-domain proteins Vreteno (Vret) and Sister of Yb (SoYb), a DEAD-box protein, Armitage (Armi), and Shutdown (Shu), a cochaperone interacting with HSP83 that is important for Piwi-piRNA loading (Handler et al., 2011; Huang et al., 2014; Olivieri et al., 2012; Olivieri et al., 2010; Preall et al., 2012; Saito et al., 2010; Zamparini et al., 2011).

Apart piRNA clusters like flamenco, some genic piRNAs were detected in OSC cells as they mapped to the 3′-UTR of some protein-coding genes, *i.e.* traffic jam (tj) (Saito et al., 2009). Binding of Yb on a specific tj and flam transcripts cis-elements was shown to be important for piRNA production (Homolka et al., 2015; Ishizu et al., 2015). Yb binds the piRNA precursors at sites where mature piRNAs have been mapped and sequesters them to the Flam body (Murota et al., 2014) that is supposed to be a place of storage of piRNA precursors and their intermediates before they are translocated by Yb to the Yb body for processing and mature piRNA production. Another protein, required for piRNA biogenesis in this system is the endonuclease Zucchini (Zuc) (Malone et al., 2009; Pane et al., 2007; Saito et al., 2009). Zuc is located on the surface of the mitochondria (Olivieri et al., 2010; Saito et al., 2010) and it processes piRNA precursors and piRNA intermediates to mature piRNAs. This Zuc-dependent endonucleolytic cleavage RNAs products contain a mono-phosphate at the 5′-end, like do mature piRNAs (Nishimasu et al., 2012). This has led to establish a model in which Zuc is responsible for piRNAs 5′-end formation. Zuc is thought to form both the 5′- and 3′-ends of piRNAs. However, Piwi-bound piRNAs have a strong 1U bias piRNAs (Nishimasu et al., 2012) and Zuc showed little nucleotide bias for RNA cleavage. Furthermore, in ovarian somatic cells piRNAs have a phasing pattern, similar to the phased piRNAs from the *Ping-Pong-Cycle-Dependent Phased piRNA Biogenesis* of germ cells (Han et al., 2015; Mohn et al., 2015). It is currently unknown how Zuc generates the 5′-end rich in uridine (1U-biased) and phased piRNAs in the ovarian somatic cells. Two others mitochondrial proteins necessary for primary piRNA biogenesis in ovarian somatic cells and shown in Figure 3 are Gasz and Minotaur (Mino), their function in the piRNA pathway remains however currently unknown.

piRNAs are then 2′-O-methylated at their 3′-termini by the methyltransferase DmHen-1/Pimet, similarly to siRNAs and Ago-2-loaded miRNAs in flies (Horwich et al., 2007; Saito et al., 2007). After piRNA biogenesis and loading of Piwi with mature piRNAs the protein translocates in the nucleus. There, it guides

transcriptional gene silencing (TGS) on TEs nascent transcripts, indicating that for silencing to take place, the TEs should be transcribed. It has been shown that the Slicer activity of Piwi is not required for the functioning of the piRNA pathway (Saito et al., 2009). In TGS, silencing is accomplished by heterochromatinization *via* deposition of histone3 lysine9 tri-methylation (H3K9me3) at *loci* of active TEs (Brennecke et al., 2007; Gunawardane et al., 2007; Kalmykova et al., 2005; Li et al., 2009; Malone et al., 2009; Sabin et al., 2013; Saito and Siomi, 2010). The aforementioned review of Yamashiro and colleagues gives a very detailed characterization of the piRNA-mediated TGS (Yamashiro and Siomi, 2018). Briefly, several important cofactors of Piwi in this mechanism have been described. Piwi interacts directly with Panoramix/Silencio (Panx) (Sienski et al., 2015; Yu et al., 2015). Panoramix then recruits the H3K9 methyltransferase Eggless (Egg) (Rangan et al., 2011; Yoon et al., 2008) and its cofactor Windei (Wde) (Czech et al., 2013; Handler et al., 2013; Koch et al., 2009), and other factors required for heterochromatinization, resulting in the repression of transposons. Asterix is another factor that interacts with Piwi, but probably mediates its functions upstream of Panoramix (Donertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013) however its function in the pathway is yet unknown.

Piwi also interacts with the linker histone H1 and this interaction targets H1 at DNA loci containing TEs (Iwasaki et al., 2016). The group discovered that the accumulation of the H3K9me3 repressive chromatin mark at the loci is not influenced by the presence of H1 and *vice versa* (Iwasaki et al., 2016). Nevertheless, this interaction of Piwi with the linker histone H1 implies a role of the protein in chromatin remodeling. Whether TGS is accomplished only by histone modifications, or if it also includes remodeling of the chromatin and if these processes are linked remains to be uncovered.

b. Biogenesis and mechanism of action of piRNAs in the germ cells

Drosophila germ cells express the three PIWI proteins: Piwi, Aub and Ago-3 (Brennecke et al., 2007). In the germline piRNA-mediated silencing of TEs is accomplished *via* the primary piRNA pathway, but also *via* two other mechanisms: the piRNA amplification loop or ping-pong cycle, and the ping-pong-cycle-dependent phased piRNA biogenesis.

In the *Drosophila* germline, most of the piRNAs derive from dual-strand clusters, which are probably unique to *Drosophila*. These clusters are transcribed on both genomic strands and thus their transcription products are complementary (Brennecke et al., 2007; Malone et al., 2009). These clusters are targets of the piRNAs they produce. Moreover, dual-strand clusters lack promoter sequences and are enriched in transcriptional repression signals, such as H3K9me3 (Le Thomas et al., 2014; Mohn et al., 2014). Thus, one

of the major questions in the field was to determine how dual-strand clusters are transcribed. Answers to this question came from the study of the *Drosophila* clusters 42AB and 80F. The core of the components shown to define dual-strand clusters in the *Drosophila* ovaries is a complex formed between the proteins Rhino, Deadlock and Cutoff, the RDC complex (Mohn et al., 2014). As of today, several functions of Rhino and Cutoff in the production of dual-strand clusters-derived piRNAs have been described (see below). Much less is known about the role of Deadlock of the RDC complex. Two hybrid experiments showed the interaction of this protein with Rhino and Cutoff but, as it doesn't have any conserved functional domains, deciphering its function is more complicated (Mohn et al., 2014).

It is important to mention that the RDC-complex mediated transcription of piRNA dual-strand clusters is not conserved in other species. Rhino, is a heterochromatin protein-1 (HP1) paralog, enriched at piRNAs bidirectional loci (Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2014), that anchors the RDC complex *via* its chromodomain to the H3K9me3 chromatin marks on the clusters. Moreover, a regulatory feedback loop between piRNAs producing loci and Piwi protein was discovered, as loss of Piwi provoked loss of small RNAs and of RDC from some piRNA clusters. In their study, Mohn and colleagues (Mohn et al., 2014) described the RDC as licenser of noncanonical transcription of dual-strand piRNA clusters, supposing that: "*this process involves 5′ end protection of nascent RNAs and suppression of transcription termination*". They proposed two models for the mechanism by which RNA-Pol II transcribes dual-strand clusters. One of the mechanisms consists in transcriptional read-through of the clusters that has been initiated at the promoters of the cluster-flanking genes. This model was further supported by the known anti-termination of transcription function of Cutoff from the RDC complex (Chen et al., 2016; Mohn et al., 2014). The second model implicates dual-strand clusters transcription by Pol II, mediated by pervasive internal transcription initiation. In the study of Andersen and colleagues (Andersen et al., 2017), the authors tested the read-through model by deleting the promoter of a gene flanking the 42AB cluster, the largest bidirectional piRNA cluster. It seemed that not only the cluster transcription was not affected by the lack of a nearby gene promoter but the cluster propagated on this promoter-less region. Furthermore, Rhino occupancy was increased and small ectopic RNAs from the region were detected. The authors concluded that: "*bidirectional piRNA cluster expression does not rely on read-through transcription. Instead, flanking transcription units delimit piRNA clusters*". This work (Andersen et al., 2017) validated the second model and uncovered a mechanism of widespread RNA-Pol II-mediated transcriptional internal initiation within the heterochromatic piRNA dual-strand clusters. Dual-strand clusters contain pyrimidine/purine dinucleotides motifs (YR motifs) on both strands. The basal transcription factor TFIID

complex binds to this motif during transcription initiation (Kaufmann and Smale, 1994; Purnell et al., 1994). The authors showed that Moonshiner, a paralog of TFIIA-L, mediates the transcription initiation in dual-strand clusters by helping to recruit the TFIID complex. Moonshiner binds at the same time Rhino and the TFIID complex protein TRF2 (TATA-box-binding protein-related factor) (Andersen et al., 2017).

Unlike the uni-strand clusters in somatic ovarian cells, the piRNA precursors from dual-strand clusters are unspliced. Two proteins have been implicated in protecting dual-strand cluster transcripts from splicing, Cutoff and the DEAD-box containing protein UAP56, as loss of both of these proteins lead to splicing of the piRNA cluster transcript (Chen et al., 2016; Zhang et al., 2014). UAP56 is also recruited to the dual-strand piRNA precursors by Rhino (Zhang et al., 2012).

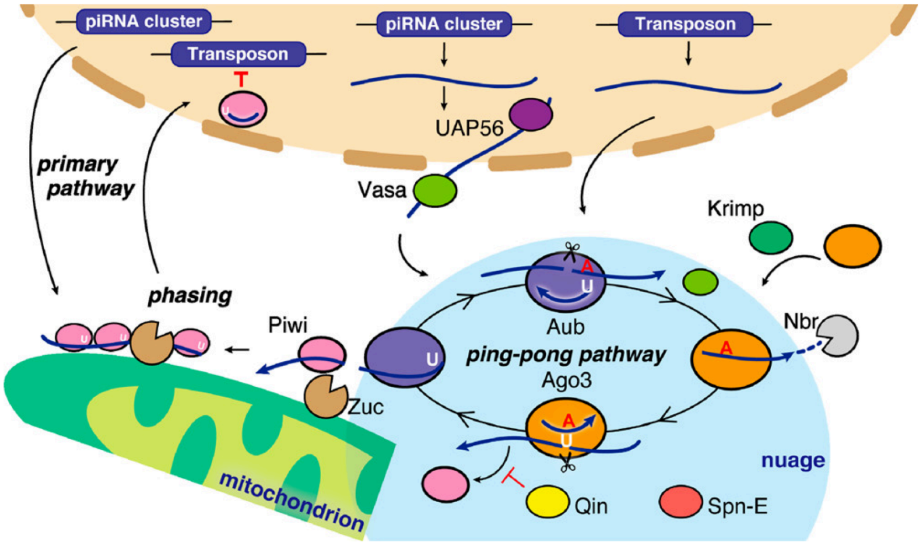

Figure 4. Schematic representation of the *Drosophila* germinal piRNA pathway. piRNA cluster transcripts are exported to the cytoplasm *via* their interaction with UAP56 in the nucleus and their subsequent transfer from UAP56 to the helicase Vasa, localized in the cytoplasm. Vasa then transports the piRNA precursor to the perinuclear cytoplasmic structure of germ cells, the nuage. In the nuage a ping-pong cycle of amplification of piRNAs take place. Antisense piRNAs derived from the piRNA cluster are loaded into Aubergine (Aub). Aub-loaded piRNA silence transposon transcripts *via* Aub's slicer activity. The derived sense piRNAs are loaded in Ago-3 and drives production of novel anti-sense piRNAs prom the piRNA precursor transcript (anti-sense). These can be again loaded in Aub, leading to s amplification of the piRNAs simultaneously with the post-transcriptional silencing of the TE transcripts. Vasa, Krimper (Krimp), Qin, and Spindle-E (Spn-E) also participate in the ping-pong cycle. Nibbler is involved in the piRNAs 3'-end formation. In the germline a second mechanism for piRNA biogenesis exists. Some of the Ago-3-loaded piRNAs are passed to Zuc on the mitochondrial membrane to trigger ping-pong dependent phasing piRNA production, which like in the soma is dependent on Zuc. The piRNAs generated are loaded in Piwi. A primary piRNA pathway also generates piRNAs loaded in Piwi, which silence transposable elements transcriptionally. Mature piRNAs are 2'-O-methylated at their 3'-terminus by DmHen-1(not shown). Adapted from (Yamashiro and Siomi, 2018).

In piRNA biogenesis in the germline, UAP56 is involved in the nuclear export of the dual-stranded piRNA cluster transcripts (Figure 4). UAP56 makes a link between the cluster transcript and the cytoplasmic piRNA biogenesis factor Vasa (Liang et al., 1994; Malone et al., 2009). UAP56 is the protein responsible for the transfer of the cluster transcript through the nuclear pore to the cytoplasmic helicase Vasa (Meignin and Davis, 2008; Zhang et al., 2012). Vasa is localized on the cytoplasmic side of the nuclear pores and is responsible for the piRNA precursors transfer to the "nuage". In the germ cells, nuage is a perinuclear structure functionally similar to the Yb body of the ovarian somatic cells (Brennecke et al., 2007; Gunawardane et al., 2007; Huang et al., 2014; Lim and Kai, 2007; Nishida et al., 2007).

The current model for piRNA biogenesis was inferred from the sequences of piRNAs that are bound to Piwi, Aub and Ago-3. piRNAs bound to Piwi and Aub are typically antisense to transposon mRNAs, whereas Ago-3 is loaded with piRNAs corresponding to the transposon mRNAs themselves (sense sequences). Moreover, the first 10 nucleotides of antisense piRNAs are frequently complementary to the sense piRNAs found in Ago-3 (Brennecke et al., 2007). According to the current model of piRNA pathway target regulation, antisense piRNAs direct cleavage of transposon mRNA or chromatin modifications at transposon loci (Sienski et al., 2012). In the germ line, apart from the primary biogenesis mechanism, Aub collaborates with Ago-3 to mediate an amplification cycle that generates additional piRNAs, the ping-pong cycle. The ping-pong cycle relies on the slicer activities of Aub and Ago-3 (Figure 4). The cleavage takes place on the target RNAs between the 10^{th} and 11^{th} nt relative to piRNAs, similarly to AGO clade proteins activity (Brennecke et al., 2007; Gunawardane et al., 2007; Malone et al., 2009). These piRNAs are capable to target the transposable elements transcripts. Thus, in the germline the repression is accomplished by TGS but also by PTGS (post transcriptional gene silencing), whereas, in the follicular cells the TGS seems to be the main functional mechanism for transposon silencing.

The ping-pong cycle takes place in nuage and involves the proteins Aub, loaded with antisense to TE transcripts piRNAs and Ago-3, binding sense transcripts (Brennecke et al., 2007; Malone et al., 2009). It is thought that some of the Aub-loaded piRNAs are derived from the piRNA cluster transcripts. Then Aub-bound antisense to TEs piRNAs direct the silencing of transposons by cleaving their complementary transcripts. Unlike in the RNAi pathway, where Ago-2 loaded with siRNA degrades all of its target RNA, the TE mRNAs, that are cut by Aub, are loaded in Ago-3 generating a second wave of piRNAs derived from the cut TE transcript. These sense piRNA from Ago-3 are complementary and able to target antisense TE sequences contained in the piRNA cluster transcripts. This leads to the production of a new set of antisense piRNAs, capable to target TE transcripts that get loaded in Aub. This cycle of consecutive

cleavage of TE transcripts by Aub-loaded piRNAs and amplification of the available piRNAs from the piRNAs clusters, by use of Ago-3-loaded sense degradation products of the action of Aub, leads to efficient PTGS of TE through amplification of piRNAs. Aub-bound piRNA, similarly to PIWI-bound piRNAs have a strong 1U bias, Ago-3-loaded piRNAs do not have a 1U bias but, as products of the Slicer activity of Aub, a 10A bias. Aub and Ago-3-loaded piRNAs are often complementary at the 10 first nt from their 5'-end, the so-called ping-pong signature (Brennecke et al., 2007; Gunawardane et al., 2007), which is conserved amongst different species (Grimson et al., 2008). In the ping-pong generated piRNAs, the 5'-end of these piRNAs is determined by the Slicer activity of the two proteins involved, Ago-3 and Aub. For some of the piRNAs once the 5'-end is generated their 3'-end must be trimmed (Kawaoka et al., 2011; Zhang et al., 2015). Amongst the nucleases shown to be involved in the determination of the 3'-end of mature piRNAs in this mechanism are Zuc and Nibbler (Feltzin et al., 2015; Han et al., 2015; Hayashi et al., 2016; Mohn et al., 2015; Wang et al., 2016).

Other factors required for the ping pong cycle are Vasa and three Tudor-domain proteins, Krimper (Krimp), Qin, and Spindle-E (Spn-E) (Anand and Kai, 2012; Czech et al., 2013; Handler et al., 2013; Malone et al., 2009; Muerdter et al., 2013; Ott et al., 2014; Patil and Kai, 2010). Vasa helicase activity is thought to increase the efficiency of the ping-pong cycle by displacing the cleaved target transcripts from Aub-piRISC. Krimper binds Ago-3 and its activity is important for the proper formation of the complex between Ago-3 and Aub (Sato et al., 2015; Webster et al., 2015). In Krimp KD in fly ovaries Ago-3 has been detected as loaded with Aub/Piwi-piRNAs (Sato et al., 2015). Qin's assures that the Ago-3-cleaved small RNAs are loaded onto Aub and not in Piwi (Wang et al., 2015). Spn-E's exact molecular function in the piRNA biogenesis is not yet elucidated. Further details on the pathway are reviewed in (Yamashiro and Siomi, 2018).

A third mechanism of piRNA biogenesis is the ping-pong-cycle-dependent phased piRNA biogenesis (reviewed in (Czech and Hannon, 2016; Hirakata and Siomi, 2016)). Similarly, to the siRNAs phasing pattern, arising at each 21 nt from their dsRNA precursors due to Dcr-2 consecutive cleavage of the precursor, Piwi-loaded piRNA have also a phasing pattern. The phasing is Zuc dependent, as it disappears in zuc mutants (Han et al., 2015; Mohn et al., 2015; Wang et al., 2015). Thus, in the germline as well as in the ovarian somatic cells of *Drosophila,* Zuc functions in production of phased piRNAs loaded in Piwi. In the germ cells a proportion of piRNAs generated by Ago-3, instead of going in the ping-pong cycle are loaded onto Piwi (known as trailer piRNAs). As Zuc is responsible for the production of the phased Piwi-

bound piRNAs, not only the 5'- but also the 3'-ends of the phased piRNAs are defined by the endonuclease activity of Zuc.

As piRNAs are present in the oocyte they get maternally deposited in the early embryos (Brennecke et al., 2008; Grentzinger et al., 2012). It has been found that Piwi induced H3K9me3 accumulation on piRNA clusters occurs in the early stages of embryogenesis, as zygotic Piwi KD led to reduction of the repressive mark on the clusters that persisted in the adult ovaries and triggered TE derepression and reduction in piRNAs (Akkouche et al., 2017; Wang and Elgin, 2011). Thus, Piwi function in the piRNA pathway is crucial in the earlier embryonic stages. Moreover, subsequent inactivation of Piwi in the later development stages did not suppress the deposed H3K9me3 histone marks (Akkouche et al., 2017; Le Thomas et al., 2014). This unraveled a role of maternally deposited piRNAs and their Piwi-associated complexes in specifying the piRNA clusters which persists in the adult ovaries, uncovering more insights into the mechanisms of the epigenetic function of this small RNA silencing pathway (Brennecke et al., 2008; de Vanssay et al., 2012).

c. Repression of protein-coding genes by piRNAs

Apart silencing TEs, piRNAs have been shown to regulate the expression of protein coding genes, similarly to miRNAs. For instance, in the male germline of *Drosophila* piRNAs target the Stellate locus, located on the X chromosome. Expression of Stellate, a subunit of casein kinase II, leads to accumulation of aggregates during spermatogenesis and infertility (Pandey and Pillai, 2014). Similarly, to TEs sequences, in the Stellate locus the sequence coding for casein kinase II is repeated on the X chromosome. In order to overcome the effects of the expression of Stellate a repression mechanism has evolved in the male germline. On the Y chromosome, a piRNA cluster type locus, named Suppressor of Stellate [Su(Ste)], contains Stellate fragments which are antisense to the Stellate locus. Interestingly, chromosome Y contains a locus rich in Ste fragments, and produces small Aub-loaded RNAs (Nishida et al., 2007; Schmidt et al., 1999; Vagin et al., 2006). Currently the small RNAs are referred to as Su(Ste)-piRNAs and the Su(Ste) locus is considered a piRNA cluster implicated in protein-coding silencing.

d. Regulation of maternal mRNAs localization by piRNAs

It has been shown that in the female germline of flies maternally deposited Aub-loaded piRNAs regulate the posterior localization of some maternal mRNAs in the embryos (Barckmann et al., 2015; Rouget et al., 2010; Vourekas et al., 2016). Such mRNAs accumulation at the posterior pole is important

for proper development and the pioneer study, linking piRNAs function to maternal mRNA localization regulation, showed that Aub-piRNA complexes destabilized the morphogens nanos and oskar at the anterior pole of embryos, allowing their accumulation only at the posterior pole (Barckmann et al., 2015; Rouget et al., 2010).

It has been shown that in the embryos Aub-piRNAs complexes bind to numerous, predominantly posterior localized mRNAs (Vourekas et al., 2016). This function is dependent on the extent of complementarity between the small RNA and its target. If the complementarity is very low, the complex might not be stable enough. If very high complementarity exists, this might induce the Slicer activity of Aub, leading to cleavage of the associated maternal mRNAs and thus fail in their maternal deposition and/or localization. A third possibility is that supplementary co-factors inhibit Aub's Slicer activity in these complexes, allowing the accumulation of these specific transcripts at the posterior pole. A similar function of the piRNA pathway in embryogenesis has also been identified in mice (Goh et al., 2015; Gou et al., 2014; Zhang et al., 2015).

5. Pimet (DmHen1) and the role of methylation for small RNA stability

Plant small-RNA methylation is catalyzed by HEN1 (Li et al., 2005; Park et al., 2002; Yu et al., 2005). Unlike in animals, in plants both miRNAs and siRNAs are 3' terminally methylated, a modification that is crucial for their stability (Park et al., 2002; Yu et al., 2005). HEN1-mediated methylation protects plant miRNAs from further modifications, such as 3'-uridylation, which marks miRNAs for degradation by exonucleases (Li et al., 2005). The *D. melanogaster* homolog of HEN1 (DmHen1, also named Pimet) bears an evolutionarily conserved methyltransferase domain, closely resembling the *Arabidopsis* HEN1 catalytic domain (Park et al., 2002). In flies, DmHen1 catalyzes the 2'-*O*-methylation at the 3'-termini of siRNAs and piRNAs as well as Ago-2-loaded miRNAs (Horwich et al., 2007). Modification of siRNA *in vitro* is dependent on SAM, consistent with DmHen1 transferring a methyl group from SAM to the terminal 2' hydroxyl group of the RNA (Horwich et al., 2007). Unlike *Arabidopsis* HEN1, the *Drosophila* ortholog does not contain a dsRNA-binding motif and acts on ssRNA. In contrast to its *D. melanogaster* homologue, plant HEN1 is nuclear and probably acts before miRNAs are loaded into Ago-1, because both strands of the miRNA:miRNA* duplex have modified 3'ends (Yu et al., 2005).

All 2'-*O*-methylated small RNAs (pi, mi or siRNA) that have been identified to date associate with a RISC complexes having an efficient endonucleolytic activity, like Ago-1-associated plant miRNAs, fly Ago-2-associated miRNAs, animal PTGS-mediating PIWI-clade associated piRNAs and Ago-2-associated

siRNAs. Conversely, *Drosophila* miRNAs, which are typically loaded on Ago-1-RISC, characterized with a very weak endonucleolytic activity, are unmodified. It was then speculated that the association of DmHen1 with RISC complexes is depending on the identity of their Argonaute protein (Horwich et al., 2007).

Results of experiments in somatic *Drosophila* S2 cell lines and in germline-tissues suggested that the modification of the 3' terminus of siRNAs and Ago-2-loaded miRNAs is coupled to the assembly of Ago-2-RISC and occurs following the cleavage of the passenger strand and thus the conversion of pre-RISC to mature RISC (Horwich et al., 2007). 2'-*O*-methylation of the single-stranded small RNA is considered to be the final step in mature RISC assembly (Horwich et al., 2007; Saito et al., 2007).

In *Drosophila*, piRNA methylation occurs after the piRNAs are loaded onto PIWI proteins. The length and abundance of piRNAs as well as piRNA function are decreased in *Dmhen1* mutants. In the female germline it was shown that *HeT-A* transposon expression increases in *Dmhen1* null mutants, suggesting that modification of the 3'- ends of piRNAs by DmHen1 is required for piRNA-directed silencing (Horwich et al., 2007).

It was also shown that DmHen1 physically interacts with the Argonaute PIWI proteins (Saito et al., 2007). The authors also showed that a GST-DmHen1 recombinant protein was able to methylate piRNAs immunoprecipitated with the PIWI proteins from DmHen1 mutant ovaries.

The methylation of small RNAs by DmHen1 seems to be influenced by the accessibility of the 3'-ends of the small RNA substrates. Structural analysis of Argonaute proteins suggested that the 5'-end of the small RNA is anchored in a highly conserved pocket in the PIWI domain, whereas the 3'-end of the small RNA is embedded in the PAZ domain (Figure 1). It was proposed that the 3'-ends of PIWI-clade protein-associated piRNAs are more flexible and are exposed to the surface of the protein, whereas the 3'-ends of Ago-1-associated miRNAs are more likely to be embedded in the PAZ domain and therefore not exposed to the surface of the protein for DmHen1 methylation.

Surprisingly, fly mutant for *DmHen1* are viable and fertile and, strikingly, the purified recombinant GST-DmHen1 protein is 50-fold more active on siRNAs methylation when supplemented with ovary lysate from *Dmhen1* null mutant homozygous flies (Czech et al., 2008). The authors hypothesized that this phenomenon reflects the requirement of the Ago-2 RISC machinery for DmHen1 function in flies, although they did not exclude the possibility that the lysate contains supplementary factor(s) required for regulating

DmHen1 activity. However, we cannot exclude that this observation might be due to redundant protein(s) function(s).

6. Implication of the sncRNAs in pathology

This variety of functional and mechanistic outputs of sncRNA pathways might explain why their misregulation have been correlated with the emergence and development of cancer, neurodegenerative diseases and fertility troubles (for a review see (Huang et al., 2013)).

In plants and animals, the early study of miRNA pathway mutants showed a higher occurrence of developmental defects and lethality in mutants of miRNA-associated Argonaute proteins (Knight and Bass, 2001; Park et al., 2002). Presently, miRNAs are recognized as a large class of endogenous post-transcriptional regulators of numerous biological processes, including early development, stem cell differentiation, the immune response, cellular homeostasis, cell-cycle control and viral replication. The conservation of multiple miRNAs across different species during evolution indicates their importance. Deregulation of miRNA expression or function has been linked to many human diseases, including diverse cancer types and a number of neurological disorders including Alzheimer's disease, multiple sclerosis and schizophrenia (reviewed in (Dong et al., 2013; Hebert and De Strooper, 2009; Liu, 2016). Thus, targeting miRNAs opens the possibility of a novel class of therapeutics approaches. For instance, a sharp decrease of the majority of mature miRNAs is observed in tumor tissues. Expressions of proteins necessary for miRNA biogenesis are affected in many human cancers and studies link miRNAs' dysregulations to the progression of invasive tumors and metastatic stages (Png et al., 2011). In recent years, many studies have highlighted the benefits of using miRNAs, or proteins of these pathways, as biomarkers of a variety of tumors and more recently male infertility (Adams and Eischen, 2014; Khazaie and Nasr Esfahani, 2014; Oom et al., 2014; Ozen et al., 2008).

RNAi is the major mechanism for defense against RNA viruses in plants and invertebrates (Ding and Voinnet, 2007; Galiana-Arnoux et al., 2006; van Rij et al., 2006; Wang et al., 2006). In *Drosophila*, viral infection triggers the production of 21 nt siRNAs duplexes (also called vsiRNA for viral siRNA). They are processed from dsRNA viral intermediates by Dcr-2, loaded into Ago-2 protein complexes, unwounded and one of the strand guides the cleavage of target RNAs by sequence complementarity (Okamura et al., 2004) (Miyoshi et al., 2005; Tomari et al., 2007). A dramatic increase in sensitivity of *dcr-2* and *ago-2* mutants to viral infections reveals the importance of RNAi in *Drosophila* antiviral defense (Galiana-Arnoux et al., 2006; van Rij et al., 2006; Wang et al., 2006). While mammals have evolved an elaborate protein-

based immune system, it has been shown that RNAi pathways can also function in the antiviral defense in mouse embryonic stem cells and suckling mice, suggesting its implication in early antiviral defenses in mammals (Li et al., 2013b; Maillard et al., 2013).

Misregulation of the piRNA pathway has been reported in somatic cancers and some Piwi family protein members are abnormally expressed in cancer cells (Suzuki et al., 2012). For example, overexpression of the human Argonaute piRNA protein Hiwi is correlated with the formation of seminoma, a testicular tumor (Qiao et al., 2002). It has been shown in flies that an abnormal expression of Piwi in brain tumors is a causal effect for tumor growth (Janic et al., 2010). Furthermore, human orthologs of *Drosophila* Piwi proteins are abnormally expressed in tissues infected with Human Papillomavirus, a virus that can lead to cervical cancer (He et al., 2010). Similarly, these proteins are ectopically expressed in breast tumors (Liu et al., 2010), where they suppress cancer cells apoptosis, thus leading to cell survival and possible dissemination (Lee et al., 2010). Together, these observations suggest that ectopic expression of germ stem cell maintenance proteins, such as PIWI sub-clade of Argonautes, in somatic tissues could confer a tumorigenic potential. For a recent review on the implication of the piRNA pathway in cancer see (Liu, 2016).

PIWI proteins functions have been recently linked to another human pathology, azoospermia occurring during abnormal spermatogenesis. The study was performed in mice and implicated Miwi in normal spermatogenesis. Miwi is a mouse PIWI protein expressed in the meiosis pachytene stage during the late stages of spermatogenesis. It has been shown that degradation of Miwi at the end of spermatogenesis is important for the normally occurring replacement of histones with protamines during the condensation of the genetic information in the sperm heads, causing azoospermia. The authors found that this role of Miwi is independent of its associated piRNAs, as miwi mutant unable to load piRNAs still interfered with normal spermatogenesis (Gou et al., 2017). Furthermore, some human patients with azoospermia bearing mutations in the Hiwi gene, were described in the study. As the described functions of Miwi and potentially Hiwi involved protein domains are not conserved in *Drosophila*, such mechanism is probably not conserved in flies. However, this is one of the first examples of a lack of PIWI protein function implicated in human pathology.

7. Communication of the sncRNA pathways in flies

Several lines of evidence indicate that in *Drosophila* the three sncRNA pathways overlap, at least to a certain extent. A very strong example in flies was the aforementioned discovery that a subset of *Drosophila* miRNAs are loaded on the siRNA pathway effector Ago-2 (Forstemann et al., 2007). The siRNA

pathway is a main pathway for TE silencing in the somatic tissues of *Drosophila* (Czech et al., 2008; Ghildiyal et al., 2008). However, the siRNA pathway is also active in the germline (Czech et al., 2008) (Shpiz et al., 2014; Tchurikov and Kretova, 2011) and has been reported to be able to regulate affect some TEs (Czech et al., 2008). A cooperation between the siRNA and the piRNA pathway has been described by our team, in a study on the repression of TE in the head of *Drosophila* (van den Beek et al., 2018). Another study reported in Piwi-depleted follicle cells an up-regulation of 21 nt small RNAs complementary to the TE ZAM (Rozhkov et al., 2013), suggesting similar interconnection between the si and the piRNA pathway, as the one described by our group in *Drosophila* heads. Finally, a study from the Chambeyron's group also reports on such interconnection between the siRNA and the piRNA pathway in the regulation of some TEs families in gonadal somatic cells (Barckmann et al., 2018). Another example is the function of two microRNAs (miR 34 and 14), reported to be capable silence in flies reporters specific for the somatic piRNA pathway (Mugat et al., 2015).

Chapter II. tRFs provide a link between Epitranscriptomics and the regulation of sncRNAs

1. Transfer RNAs (tRNAs) biogenesis

tRNAs are ubiquitously expressed abundant small RNAs of 70-90 nt playing a central role in the cellular metabolism. Their role as adaptors during the translation of the genetic information is well established and will not be discussed in this manuscript. More recently, studies on the modifications of tRNAs and on their fragmentation to tRNA-derived fragments (tRFs), have provided new insights on the complexity of functions accomplished by tRNAs and its derivatives in the cell.

The biogenesis of tRNAs has been extensively studied in yeast and is also well characterized in mammals. As their biogenesis is very complex it will not be detailed here, but is reviewed in depth in (Chatterjee et al., 2018; Huang and Hopper, 2016). I will summarize key notions developed in those recent reviews. The studies of tRNA biogenesis in flies is much scarcer.

tRNAs have a characteristic cloverleaf secondary structure and are constituted of ssRNA *loops* and *stems*, dsRNA parts (Figure 5). They are constituted of an anticodon stem loop (AC) containing the anticodon at positions 34 (Wobble position) to 36 (Figure 5, red line); a D-loop containing a dihydrouridine (D); a TΨ-loop containing a pseudouridine (Ψ), a variable loop (sometimes); and an acceptor stem which is the amino acids attachment site at the 3'-end CCA trinucleotide.

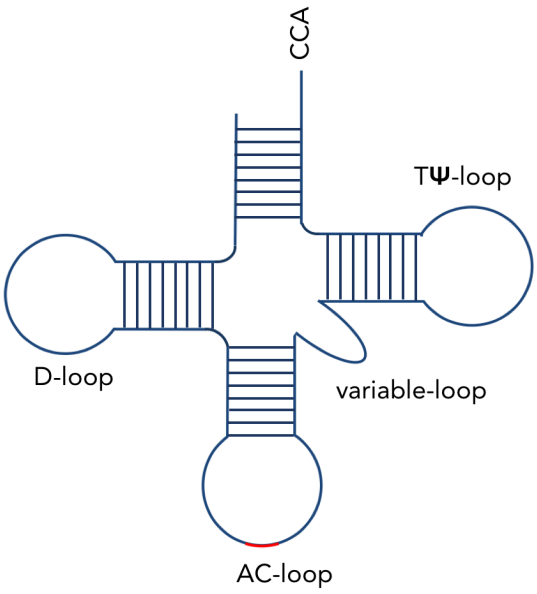

Figure 5. Secondary structure of tRNAs. Cloverleaf representation of a tRNA. Represented are the D-loop, the TΨ-loop, the AC-loop (including the anticodon sequence (red)) and the variable loop. 3'-end CCA trinucleotide is added to tRNAs post-transcriptionally in eukaryotes.

tRNA biogenesis involves multiple steps of maturation starting with the transcription by RNA polymerase III of their precursor tRNA (pre-tRNA) and continuing to the formation of a mature tRNA, charged with an amino acid, with a characteristic L-shaped three-dimensional structure. The majority of the steps in tRNA biogenesis occur in the nucleus, but some are taking place in the cytoplasm. This implies a spatio-temporal regulation of tRNA biogenesis in the cell (Figure 6).

First, in the nucleus, a process termed end-processing occurs (Figure 6). The majority of pre-tRNAs contain 5'-leader and 3' trailer sequences. Maturation of tRNAs usually starts with removal of the 5'-leader sequences by the endonuclease RNase P. The maturation of the 3'-end is more complex, as it may require both exo- and endonucleases. RNase Z is a well described endonuclease involved in the 3'-end processing of mitochondrial and nuclear encoded tRNAs. The tRNA binding La protein (Lhp1 in yeast), a protein involved in miRNA processing (Liang et al., 2013), is also involved in pre-tRNA processing. tRNAs contain a 3'-terminal CCA sequence which is important for tRNA aminoacylation. In yeast and vertebrates, CCA is not encoded in the genome and its addition to the 3'-end of tRNAs is enzymatic. In *E. coli* the CCA sequence is genome-encoded.

Some pre-tRNAs also contain introns, that are spliced, as they are not present in the mature tRNAs. Non-intron containing pre-tRNA in yeast and vertebrates, as well as intron-containing tRNAs in vertebrates are processed in the nucleus. These tRNAs maturation steps of end-processing, CCA-addition and splicing (only in vertebrates) occur in the nucleus. Then, the processed pre-tRNA is translocated in the cytoplasm, for charging with their cognate amino-acid by an amino-acyl tRNA synthetase. The translocation in the cytoplasm is termed primary tRNA nuclear export. It is interesting to mention that Nxf1 and Nxt1, the export factors involved in flies in primary piRNA biogenesis (Dennis et al., 2016) are characterized in other organisms as export factors of this primary pre-tRNA nuclear export. In flies, their implication with tRNA biogenesis is not described yet.

In yeast, intron containing tRNAs have a different fate. Their end processing and CCA addition also occur in the nucleus but their splicing occurs in the cytoplasm, on the mitochondrial surface. Once spliced there, they are also charged with the corresponding amino acids (Figure 6).

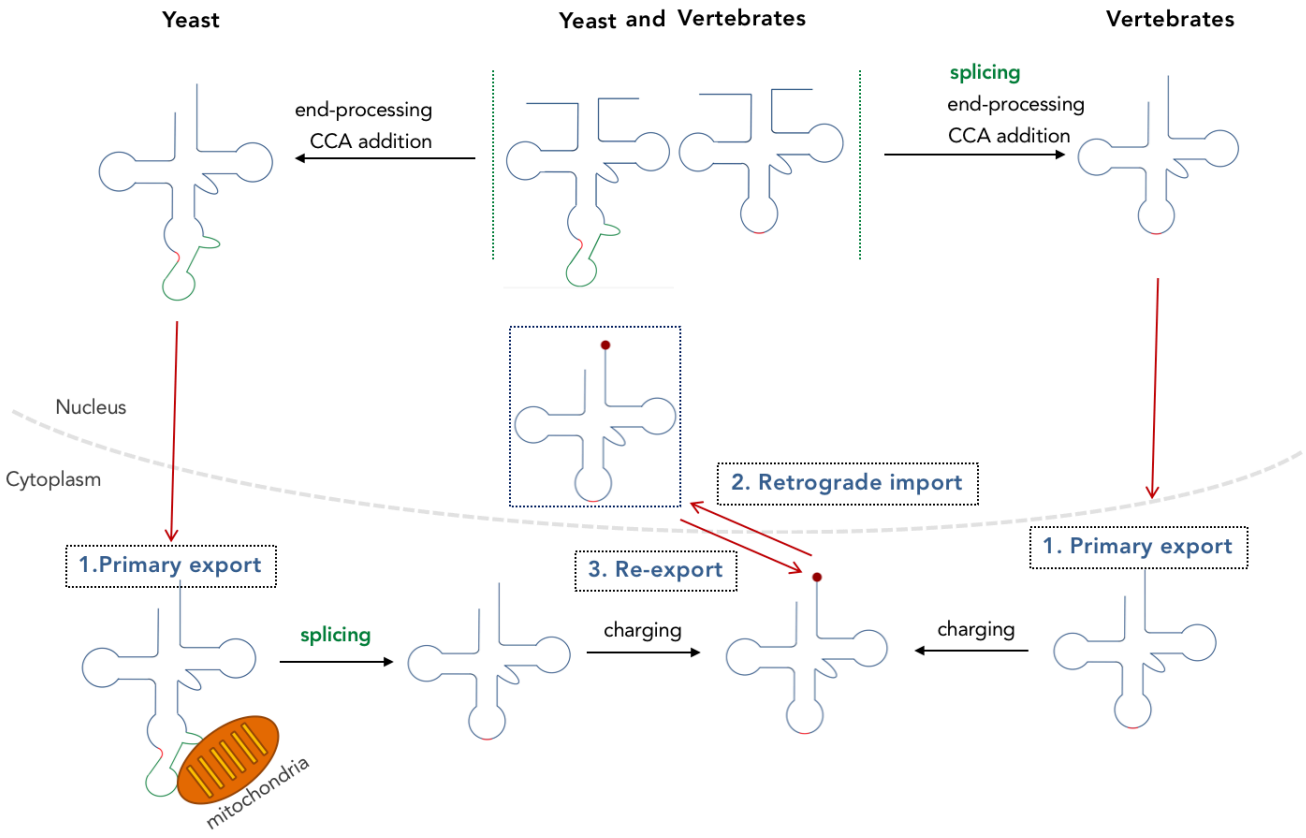

Figure 6. Overview of tRNA biogenesis in yeast and vertebrates. tRNA sequences retained in the mature tRNA are represented in blue, the anticodon is in red, intronic sequences that are spliced out of the mature tRNA are represented in green. The 3′ CCA sequence is represented as an extension on the 3′end of the tRNAs, red circles mark the amino acid in charged tRNAs; 5′ and 3′ processing of intron and non-intron containing pre-tRNA transcripts (upper part, surrounded by green dashed lines), as well as CCA addition to the 3′-end occurs in the nucleus. In vertebrates splicing also occurs in the nucleus. Yeast intron-containing tRNAs are spliced in the cytoplasm on the mitochondrial surface. Primary nuclear tRNA export assures the export of pre-tRNAs in the cytoplasm where the charging of an amino acid takes place. In yeast some posttranscriptional modifications require retrograde import of amino acid-charged tRNAs from the cytoplasm to the nucleus. Nuclear re-import occurs as well under some environmental circumstances and during tRNA quality control. When the post transcriptional maturation is accomplished, or the stimulus causing retrograde import is removed, a nuclear re-export shuttle returns the mature tRNAs to the cytoplasm. Figure inspired by (Chatterjee et al., 2018).

In flies Zuc processing of piRNA precursors is also located on the mitochondrial surface (Saito et al., 2010). This indicates, as discussed in the review of Chatterjee and colleagues, that post-transcriptional processing of some RNAs (at least piRNAs and tRNAs) can occur on the mitochondrial surface. Recently, Honda *et al.* reported specific tRNAs-derived piRNAs in *B. mori* BmN4 cells (Honda et al., 2017), whose biogenesis also occurs at the mitochondrial surface.

There are multiple tRNA nuclear export and import pathways. Apart from the primary nuclear export of uncharged pre-tRNAs, cytoplasmic tRNAs can be imported back in the nucleus *via* a tRNA retrograde nuclear import. Moreover, cytoplasmic tRNAs that have been re-imported in the nucleus can under certain circumstances return to the cytoplasm *via* tRNA nuclear re-export mechanisms. These mechanisms of

bidirectional movement of tRNAs between the nucleus and the cytoplasm is conserved between budding yeast and vertebrates and participates in multiple functions, including tRNA modification, tRNA quality control, cell cycle control and translation (reviewed in (Chatterjee et al., 2018)). The retrograde tRNA import pathway is important for regulation of retroviral replication, as it is used by retroviruses, *i.e.* HIV for their own replication cycle. It is known that tRNAs are "hijacked" by retroviral reverse transcriptases and serve as primers for reverse transcription initiation (Herschhorn and Hizi, 2010; Marquet et al., 1995) and recently reviewed in (Schorn and Martienssen, 2018).

Moreover, the traffic of tRNA between cellular compartments is linked to changes in the environmental conditions such as stress. In their review Chatterjee and colleagues suggest that tRNA subcellular trafficking could serve as a novel mechanism to regulate gene expression in response to changes in the environment (Chatterjee et al., 2018).

2. Epitranscriptomics: hidden layer of information in RNA modifications

The concept of *Epitranscriptomics* implies that, similarly to modifications on DNA and on histones, another layer of signalization information in the cell is "hidden" in RNA modifications. Ever since the discovery that some RNA modifications, *i.e.* m^6A, are reversible, and "read" by effector proteins with a consequence on variety of biological processes, a parallel with DNA and histone modifications function in conveying information has been made (briefly discussed in (Angelova et al., 2018)).

More than 140 RNA modifications contribute to an increase in the diversity of nucleotides (Boccaletto et al., 2018; Machnicka et al., 2013). Some of the most common RNA nucleotide modifications are represented in Figure 7. The presence of modified nucleotides alters the thermal stability of RNAs: RNA folding in secondary and tertiary structures is stabilized by post-transcriptional RNA modifications and hypo-modified RNAs can be degraded (reviewed in (Helm, 2006)). Post-transcriptional modifications have been implicated in gene expression regulation (reviewed in (Motorin and Helm, 2011)) and are often found on tRNAs, rRNAs and messenger RNAs (mRNAs).

RNA modification enzymes can have different substrates. For instance, in *E. coli*, FtsJ/RrmJ catalyzes 2'-O-methylation (Nm) addition on 23S rRNA (Caldas et al., 2000) as well as on tRNA (Bugl et al., 2000). Its yeast ortholog Trm7 catalyzes Nm at positions 32 and 34 on specific tRNAs ($tRNA^{Phe}$, $tRNA^{Trp}$ and $tRNA^{Leu\text{-}TAA}$)(Pintard et al., 2002).

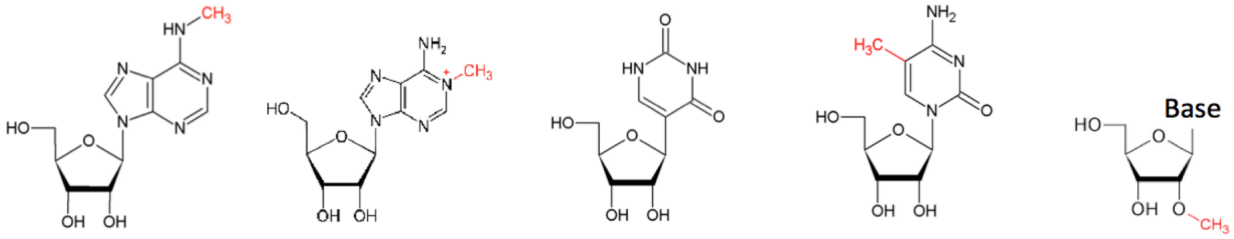

Figure 7. Some common RNA methylated nucleosides. From left to right: N6-methyladenosine (m^6A), N1-methyladenosine (m^1A), pseudouridine (**Ψ**), 5-methylcytidine (m^5C) and 2'-O-methylated bases (Nm). Modification sites are marked in red, (Image from (Angelova et al., 2018)

A. tRNAs modifications

Pioneer studies in the field of epitranscriptomics were made on tRNA and rRNA (Angelova et al., 2018). Numerous investigations on RNA modifications identified tRNAs as the most heavily modified RNAs in the cell (Boccaletto et al., 2018). tRNAs modifications occur post-transcriptionally and biosynthesis of hypermodified nucleosides usually involve several modification enzymes that act sequentially at different steps.

tRNAs modifications can be categorized in 3 groups (Machnicka et al., 2014). (i) Conserved position and identity of the modification in different tRNAs, *i.e.* dihydrouridine (D) or 2'-O-methylguanosine (Gm) in the D-loop; 5-methyluridine (m^5U) and pseudouridine (**Ψ**) in T**Ψ**-loop. These modifications are usually deposited by conserved enzymes. (ii) Conserved position of the modification in different tRNAs, *i.e.* hyper modified uridines at position 34 (Wobble) or purines at position 37. These modifications can be introduced by non-conserved enzymes. (iii) Simple nucleotides modifications often located at one position within some tRNAs, *i.e.* methylation of the base (mN; N is A, U, G or C) or the ribose (Nm; N is A, U, G or C). These modifications can be introduced by a diversity of enzymes. tRNAs modifications occurring frequency and conservation have been indexed in tRNAmodviz web server. This database contains information on tRNAs corresponding to 584 tRNAs and 5293 modifications in total from bacteria, fungi, metazoan and archaea (Machnicka et al., 2014). For instance, **Ψ** at position 55 in the T**Ψ**-loop is reported to be the most conserved modification followed by m^5U at position 54 and D at position 20.

Numerous tRNAs modifications are located in the anticodon (AC) loop (especially at the Wobble position 34) and on adjacent position 37. Positions 34 and 37 are conserved modified positions. There are

29 possible modifications found at the Wobble position 34 and 13 at position 37. They represent 70% of the modifications found on tRNAs. Previous studies demonstrated that modifications in the AC loop and the adjacent position 37 prevent frameshift during translation (Agris, 2004; Bjork et al., 1999; Urbonavicius et al., 2016) and are thus necessary for the decoding of the genetic information.

The most abundant modifications in tRNAs are: (i) an isomer of uridine, pseudouridine **Ψ** and its derivatives (*e.g.* 1-methylpseudouridine) and (ii) a reduced form of uridine, dihydrouridine D and 2'-O-methylation (Nm) on the ribose of A, U, G and C. There are few positions never found to be modified, particularly in the variable loop.

B. Implication of aberrant tRNA modification in disease

Defects in the methylation of tRNAs are associated with neurodegenerative diseases (for a review see the attached publication at the end of the Introduction section (Angelova et al., 2018)) as well as in cancer progression in human, mice and flies, similarly to the misregulation of the silencing sncRNAs: reviewed in (Anderson and Ivanov, 2014; Frye and Watt, 2006; Iyengar et al., 2014; Torres et al., 2014).

C. Role of RNA modifications for tRNA stability and tRFs accumulation

Aberrant function of the two m^5C tRNA methylases Dnmt2 and NSun2, has been linked to the stability of their tRNA substrates. In *Drosophila* mutants for *dnmt2,* as well as in NSun2-deficient mice and human patient-derived fibroblasts cultures, an increased cleavage of tRNAs into tRNA-derived fragments (tRFs) has been observed (Blanco et al., 2014; Schaefer et al., 2010). This phenomenon is enhanced under stressful conditions. The cleavage of tRNAs to tRFs in the absence of NSUN2 is catalyzed by the endonuclease angiogenin. Angiogenin was also reported to cleave the AC loop of tRNAs, giving rise to SHOT-RNA (another name of the tRNA-derived small RNA) (Honda et al., 2015).

In flies Dnmt2 adds m^5C to position 38 of $tRNA^{Asp}$, $tRNA^{Val}$, and $tRNA^{Gly}$. These methylations protect tRNAs from cleavage to tRFs (Schaefer et al., 2010). For instance, under stress conditions non-methylated tRNAs are susceptible to hydrolysis by endonucleases and generate tRFs (Schaefer et al., 2010). Studies, including our own in collaboration with Matthias Schaefer, have shown that the sncRNA pathways can be

disturbed by lack of some tRNAs modifications that correlate with increased tRFs (Durdevic et al., 2013b; Genennchер et al., 2018).

Multiple evidences have accumulated that tRNA fragments are not simple degradation products but rather functional molecules generated in a regulated manner. This suggests that those novel small RNAs represent individual biological entities. For instance, not all but only specific stress signals trigger tRNA halves accumulation (Thompson and Parker, 2009). This demonstrates that tRNA fragmentation is not a general answer for the cell to confront stressful situations but rather a selective response to certain kinds of stress. Probably the strongest argument that tRFs are not degradation products comes from the observation that the levels of mature tRNAs do not significantly change during tRNA halves accumulation (Lee and Collins, 2005; Thompson and Parker, 2009; Yamasaki et al., 2009). Moreover, tRNA fragments are often cleaved at specific positions (reviewed in (Sobala and Hutvagner, 2011)) and the fragmentation seems to be controlled by methylation such as m^5C as described for Dnmt2 (Schaefer et al., 2010) and NSun2 (Blanco et al., 2014), and also Nm as described by us, suggesting that cleavage can occur after tRNA maturation.

3. tRNA-derived fragments (tRFs)

A tRNA fragment (tRF) derived from $tRNA^{Leu}$ was first detected in the 1970s in *E. coli* soon after its infection with a T4 bacteriophage (Yudelevich, 1971). It took two more decades for the first functional study of a tRF isolated from human urinary bladder carcinoma to take place (Zhao et al., 1999). Currently, multiple studies reported the existence of tRFs (Cole et al., 2009; Haussecker et al., 2010; Kawaji et al., 2008; Lee et al., 2009; Li et al., 2008; Li et al., 2012; Reese et al., 2010). These pioneer studies on tRNA fragments biology have named them with different terms, leading to a rather nonhomogeneous nomenclature of tRFs. As of today, one can encounter the names tRNA-derived fragments (Lee et al., 2009), tRNA-derived small RNAs (tsRNAs), (Haussecker et al., 2010), tRNA halves (Fu et al., 2009; Lee and Collins, 2005), tRNA-derived halves (Garcia-Silva et al., 2010), stress-induced small RNAs (tiRNAs) (Emara et al., 2010; Yamasaki et al., 2009), or event urinary bladder carcinoma RNAs (ubcRNAs) (Zhao et al., 1999).

In one of the proposed nomenclatures, the discrimination is based on the size of the fragments and their position of origin on the mature tRNA (reviewed in (Sobala and Hutvagner, 2011)). Two classes of fragments appear: (i) tRNA halves and (ii) tRFs, characterized by smaller sizes. tRNA halves are produced

by cleavage near the AC loop and make 30–35 nt. Their production has been linked to the response to processes like oxidative stress, hypoxia, or apoptosis induction (Fu et al., 2009; Lee and Collins, 2005; Thompson et al., 2008). In mammals, tRNA halves are produced by the endonuclease Angiogenin (Fu et al., 2009) (Yamasaki et al., 2009) (Figure 8).

tRFs can be derived from the 5'- (5' tRF), the 3'-end (3'CCA tRFs) of mature tRNAs, or also from 3'-pre-tRNA trailers (3'U tRFs) and have a size around 20 nt. 3'U tRFs are formed during pre-tRNAs processing by the nuclease RNase Z.

The biogenesis mechanisms of 5'- and 3'CAA tRFs are not yet fully elucidated, but it is important to note that, as tRNAs contain double stranded parts similarly to miRNA and siRNA precursors, some tRFs were shown to be directly produced by the Dicer nuclease activity (Cole et al., 2009).

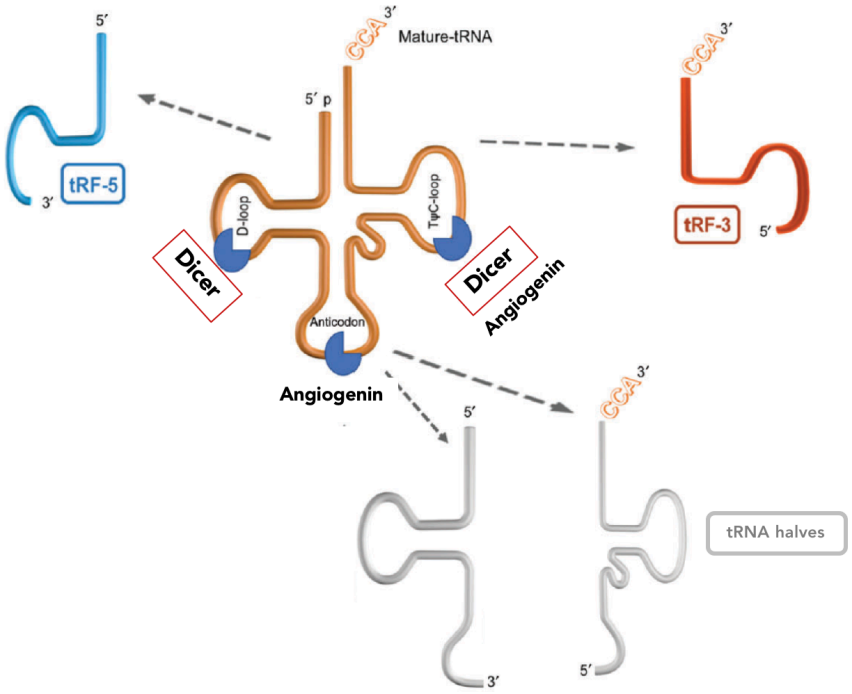

Figure 8. Biogenesis of tRFs from mature tRNAs. Represented is a cloverleaf structure of mature tRNAs with the different positions found to be cleaved by angiogenin or Dicer. Image adapted from (Soares and Santos, 2017).

Several studies have linked an up-regulation of tRNA halves during starvation or the oxidative stress response (Thompson et al., 2008; Thompson and Parker, 2009; Yamasaki et al., 2009). tRNA halves mediate translational inhibition and induction of expression of proteins involved in survival and repair responses during stress (Buchan and Parker, 2009) by promoting the assembly of stress granules (Emara et al., 2010). tRNAs can also inhibit apoptosis *via* titration of cytochrome-C, inhibiting its interaction with

Apaf-1 and thus blocking the subsequent caspase-9 activation. Moreover, RNase treatment against tRNAs was shown to enhance caspase-9 activation and apoptosis (Mei et al., 2010).

It has been shown that the abundance of tRFs does not correlate with tRNA genes copy numbers and distribution of codon usage (Kawaji et al., 2008). In their study the group compared the frequencies of tRNA-mapping small RNAs from small RNA sequencing experiment per tRNA isoacceptor family (groups of tRNA genes associated with the same amino acid) with the tRNA gene copy number on one hand and with the amino acid frequencies on the other. As they found no correlation of the abundance of tRNA-derived small RNA sequences with any of those two categories the group concluded that tRNA cleavage is oriented towards specific types of tRNAs and that the abundance of tRNA-derived small RNAs is not a direct consequence of tRNA expression.

However, in contrast to this study it has been reported that upon overexpression of particular tRNAs, distinct tRFs-3 levels increase and have a silencing activity on luciferase reporters (Kumar et al., 2015). In a later study of the same group, an overexpression of tRNAs in HEK293T leads to an increase in the levels of endogenous 18 nt tRF-3 (Kuscu et al., 2018). Furthermore, these tRFs were shown to post-transcriptionally silence genes by base-pairing to the 3' UTR of their target mRNAs. The repression of the target genes was shown to be Dicer-independent, Argonaute-dependent and accomplished by association with GW182 proteins, known translation repressors in miRNA silencing, capable to promote the target mRNAs degradation.

Interestingly, in their review on tRNA biogenesis, Chatterjee and colleagues discussed that many of the cellular stresses which induce retrograde tRNA nuclear import also induce tRNA fragmentation and speculate on the possibility that the imported tRNA fragments could play an important role in processes such as tRNA quality control and translation efficiency (Chatterjee et al., 2018).

On overview of the variety of functions into which tRFs have been implicated is reviewed in (Cristodero and Polacek, 2017) (Figure 9).

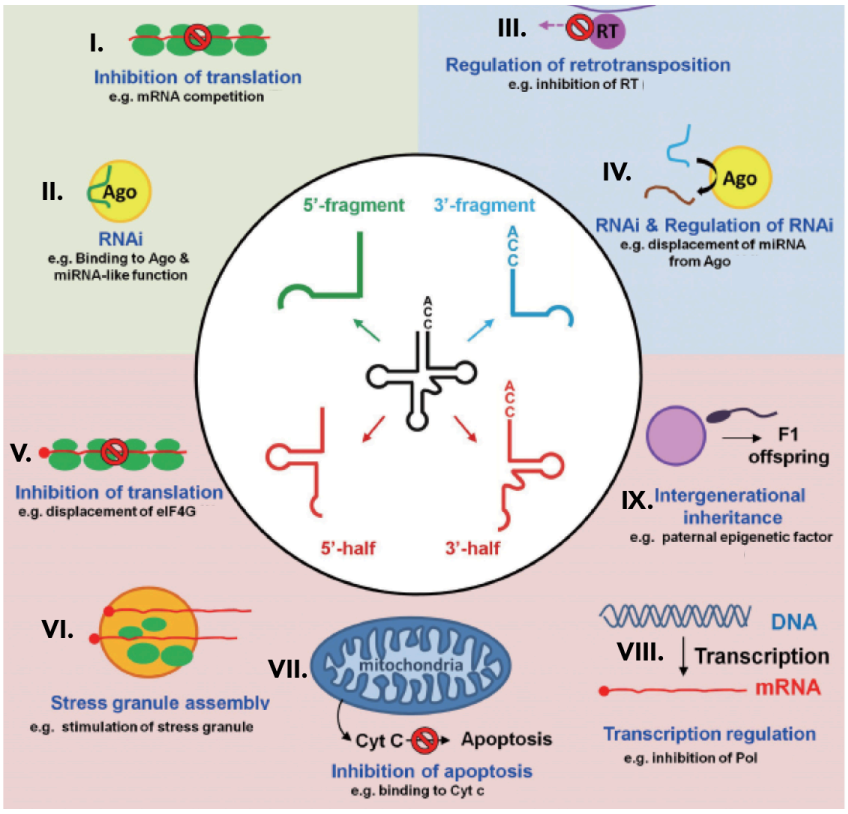

Figure 9. Multiple functions of tRNA-derived fragments. Overview of the different biological processes in which tRFs have been implicated (adapted from (Cristodero and Polacek, 2017)).

A. tRFs as a novel class of sncRNA

Similarly to the "canonical" sncRNA pathways, it has been shown that tRNA-derived fragments can regulate gene expression. A recent study on this topic was accomplished by Yeung and colleagues (Yeung et al., 2009). They characterized the small RNA population in cells infected with the human immunodeficiency virus (HIV). The authors detected an abundant 18 nt small RNA derived from a hybrid RNA duplex formed between the host human $tRNA^{Lys}$ and the virus PBS (primer-binding site). Furthermore, this $tRNA^{Lys}$-derived 3'tRF was shown to associate with Dicer and Ago-2 in order to silence the genetic reporter used in the study. This suggested that the function of this 3'tRF in HIV antiviral response can be accomplished *via* an RNAi-similar mechanism.

In mammals the tRNAs biogenesis proteins RNA Pol III and RNase Z are used by herpes viruses to generate functional viral miRNAs. RNA Pol III generates short transcripts composed of a 5' tRNA-like molecule that is flanked by a 3'-pre-miRNA hairpin. After cleavage of the pre-miRNA hairpins by the host RNase Z, they are processed by Dicer to mature viral miRNAs (Bogerd et al., 2010; Diebel et al., 2010; Reese et al., 2010). In mice, RNA Pol III has also been found to produce small RNAs from similar structures,

indicating that generation of small RNAs from Pol III transcripts is not a unique property of viral genomes (Reese et al., 2010). 3'U tRF (tRF-1001, cand45) detected in human cancerous cells is generated by mechanism, similar to the above described generation of murine herpes viral miRNAs (Bogerd et al., 2010; Diebel et al., 2010; Reese et al., 2010), although unlike for the viral miRNAs, no silencing activity was detected for this tRF on luciferase reporters, suggesting that it functions by different mechanisms (Haussecker et al., 2010; Lee et al., 2009).

The abnormal production of tRFs has also been shown to impair the siRNA pathway in *Drosophila*. Indeed, a mutation of the tRNA methylase Dnmt2 leads to an important increase of tRFs which inhibit Dicer-2 RNase activity, causing reduced siRNA production and misregulation of Dicer-2-dependent silencing (Durdevic et al., 2013b). As mentioned, Dicer generates some tRFs (Cole et al., 2009), that can thus potentially compete with other Dicer substrates. The interactions of abnormally abundant tRFs with sncRNAs' effector machineries, i.e. Dicers and Argonautes (Cole et al., 2009; Keam et al., 2014; Kuscu et al., 2018; Shigematsu and Kirino, 2015), thus affects sncRNA-mediated gene regulation.

Interestingly, recent studies showed that tRFs interact with human Piwi proteins in cancer cells (Keam et al., 2014). Moreover, a correlation between tRNA biogenesis defects and defect on piRNA pathway has been recently reported in flies (Genenncher et al., 2018; Molla-Herman et al., 2015). Furthermore, in Tetrahymena 3'CCA tRFs were found in association with Twi12, a PIWI protein that is essential for this organism (Couvillion et al., 2010).

tRFs were also found to act in relation to transposons. For instance, it has been shown that an under aminoacylated isoform of the human $tRNA^{Asp}$ isodecoder adopts an alternative hairpin secondary structure and binds to the 3'UTR of its aspartyl-tRNA synthetase mRNA increasing its translation efficiency (Rudinger-Thirion et al., 2011). Interestingly, $tRNA^{Asp}$ binds to an embedded *Alu* element in the 3'UTR *via* extensive base pairing from positions 21 through 55, leaving a part of the tRNA resembling a 5'tRF accessible for nuclease cleavage. These findings suggest a potential alternative mechanism for tRFs generation. A recent study in human cancer and mouse embryonic cells described Dicer-independent, 16–18 nt long 3'CCA tRFs that associated with Ago-2 and mediated *in vitro* silencing of target RNAs (Li et al., 2012). Similarly to the study of (Rudinger-Thirion et al., 2011) the tRFs showed high complementary to endogenous retroviral PBS in the human genome, suggesting an RNAi based function of tRFs in the regulation of retrotransposons.

A very recent example is the study of Schorn and colleagues (Schorn et al., 2017; Schorn and Martienssen, 2018). The study focused on a specific tRNA, hijacked by retroviruses for priming retroviral reverse transcription in mouse embryonic fibroblasts (MEFs). Schorn *et al.* showed that the same tRNA is being fragmented by the host in response to retrotransposon activation, giving rise to two tRFs of 18 and 22nt length, each one participating in distinct pathways in the host defense against the retrotransposons. The 18nt tRF served to block reverse transcription (RT) initiation, as they were shown to compete with the full length tRNA priming binding site, thus accomplishing a mechanism similar to the piRNA-mediated TGS, but on the level of RT (that could be called "RTGS"). If this first level of defense fails to repress the TEs, 22nt Ago-loaded tRFs assure a second layer of defense by post-transcriptionally repressing the retroviral propagation *via* a PTGS mechanism, again as shown for the sncRNA pathways. This not only solidifies the concept of information "layers" within the same molecules encoded by unique genomic sequence, but also characterize tRFs as molecules similar to piRNA and endo-siRNA in terms of target (TEs) as well as in terms of mechanism of action.

Another example of tRFs targeting TEs was described in the pollen of *Arabidopsis thaliana* where tRFs were shown to be processed by Dicer-like 1 and incorporated into Argonaute1, similarly to microRNAs (Martinez et al., 2017). The similarities between tRFs and sncRNAs are summarized in Figure 10.

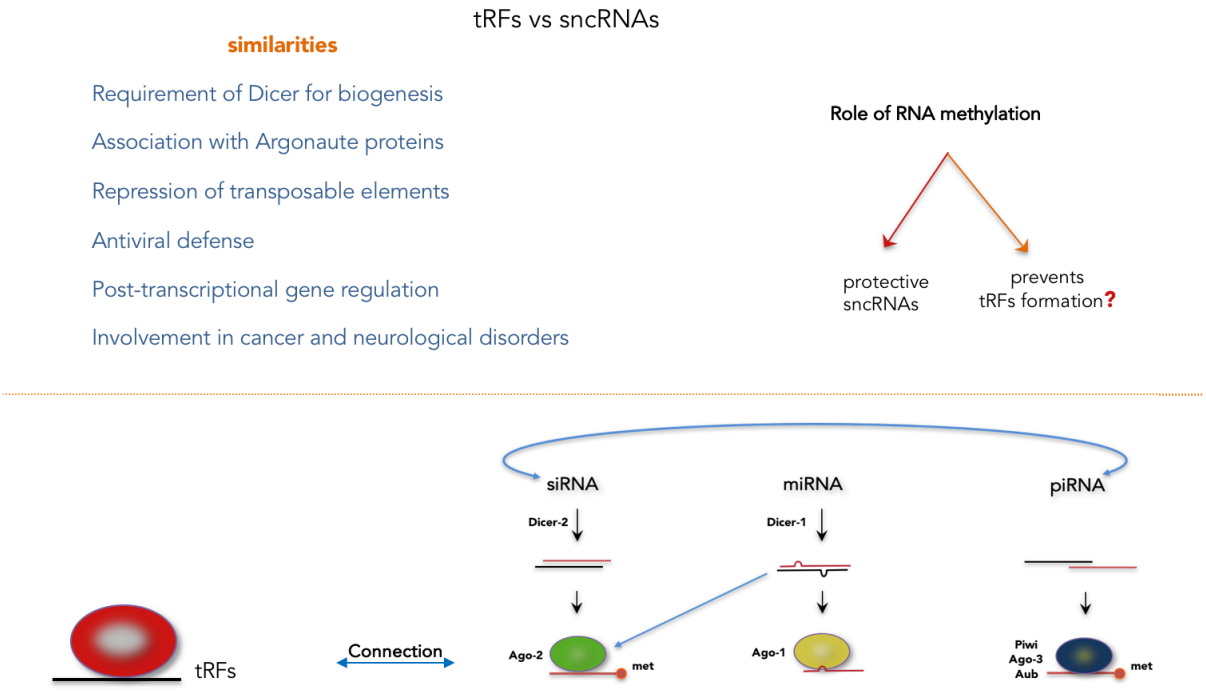

Figure 10. tRFs as novel class of small non-coding RNAs. Summary of the arguments that tRNA-derived fragments can be considered as novel class of small non-coding RNAs.

B. Implication of tRFs in disease

Already in the late 1970s, elevated levels of tRNA breakdown products were found in the urine of humans and mice with different malignancies (Speer et al., 1979). More generally, differentially expressed tRFs have now been identified in a number of cancers (for a review see (Green et al., 2016) and (Fu et al., 2015)), *e.g.* lung cancer and chronic lymphocytic leukemia (Pekarsky et al., 2016).

Differentially expressed tRFs were identified in patients with myelodysplastic syndrome (MDS), as well as in responders and non-responders to therapy (Guo et al., 2015). The expression of some of those tRFs could explain 67 % of the variation in treatment response of patients with MDS, an observation that highlights the potential utility of tRFs in clinical settings to select patients likely to respond to treatment. tRFs could thus become valuable diagnostic and prognostic markers in MDS, but also in prostate cancer, as reported by Olvedy and colleagues (Olvedy et al., 2016).

An example of the capacity of tRFs to titrate RNA binding proteins was recently reported: specific tRFs suppress breast cancer progression *via* the displacement of the mRNA-stabilizing protein YBX1 from pro-oncogenic transcripts (Goodarzi et al., 2015), destabilizing pro-oncogenic mRNAs and suppressing metastatic progression. When these tRFs were introduced into breast cancer cells, they inhibited breast cancer growth, cell invasion, and metastasis. These findings reveal a tumor-suppressive role for tRFs, as well as their potential use as cancer biomarkers. It is hence possible that this RNA displacement-based mechanism may be a general phenomenon, applying to other RNA binding proteins, like Argonautes and Dicers and playing a major role in tumorigenesis.

Angiogenin, the endonuclease responsible for the production of tRNA halves, is a potential oncogene and a vascularization agent in normal, as well as in malignant, cells. RNase Z, the nuclease responsible for the processing of pre-tRNAs and the generation of 3'U tRFs is also known, as the gene ELAC2, to susceptibility for prostate cancer (Lee et al., 2009). tRFs can influence cell proliferation, as in prostate cancer cell lines expression of 3'U tRFs was positively correlated with the cellular proliferation rate and knockdown of these 3'U tRF by siRNAs resulted in impairment of the cell proliferation (Lee et al., 2009). In prostate cancer cell lines (Lee et al., 2009) and in human hepatitis delta transformed embryonic kidney cells (Haussecker et al., 2010) a 3'U tRF (tRF-1001, cand45). tRF-1001 found to play defined biological roles in proliferation (Lee et al., 2009). The Angiogenin-generated tRNA-derived small RNA (SHOT-RNA) described in the study of Honda and colleagues were also found to be highly expressed in breast and prostate cancer (Honda et al., 2015).

Stress granule assembly mediated by tRNA fragments may play a role in tumor cells for hypoxia and starvation-induced stress signaling (Holcik and Sonenberg, 2005). tRFs have been shown to be upregulated in prostate cancer (Martens-Uzunova et al., 2012). It is also worthy to mention that as described above tRNA halves increase during nutritional starvation and oxidative stress, both processes characteristic for tumor environments (Thompson et al., 2008; Thompson and Parker, 2009).

The human long non-coding RNA MALAT1 which can interact and be processed by factors of the tRNA biogenesis machinery is known to be deregulated in many human cancers (Wilusz et al., 2008).

These are only some of numerous examples from the recent years, linking tRFs to pathology in human. Multiple reviews were published on this thematic. More generally, apart cancer, tRFs have been also been implicated in neurological disorders and in antiviral defense in different species. This has been recently reviewed by (Soares and Santos, 2017).

As discussed sncRNA pathways deregulations have been strongly associated with the emergence of cancer and neurodegenerative diseases. On the other hand, tRFs arise as novel class of sncRNAs, shown to be able to compete with sncRNAs for common factors, or targeting similar targets by similar mechanisms. It is tempting to hypothesize that neurodegenerative pathologies and cancer progression, as observed in mutants for tRNA biogenesis genes, are at least partially due to their ability to interfere with sncRNA pathways.

frontiers in Bioengineering and Biotechnology

REVIEW
published: 13 April 2018
doi: 10.3389/fbioe.2018.00046

The Emerging Field of Epitranscriptomics in Neurodevelopmental and Neuronal Disorders

Margarita T. Angelova[1†], Dilyana G. Dimitrova[1†], Nadja Dinges[2†], Tina Lence[2†], Lina Worpenberg[2†], Clément Carré[1] and Jean-Yves Roignant[2*]*

[1] Drosophila Genetics and Epigenetics, Sorbonne Université, Centre National de la Recherche Scientifique, Biologie du Développement—Institut de Biologie Paris Seine, Paris, France, [2] Laboratory of RNA Epigenetics, Institute of Molecular Biology, Mainz, Germany

OPEN ACCESS

Edited by:
Giovanni Nigita,
The Ohio State University,
United States

Reviewed by:
Jernej Ule,
University College London,
United Kingdom
Martin Kos,
Universität Heidelberg, Germany

****Correspondence:***
Clément Carré
clement.carre@upmc.fr
Jean-Yves Roignant
j.roignant@imb-mainz.de

[†]These authors have contributed equally to this work.

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *07 March 2018*
Accepted: *29 March 2018*
Published: *13 April 2018*

Citation:
Angelova MT, Dimitrova DG, Dinges N, Lence T, Worpenberg L, Carré C and Roignant J-Y (2018) The Emerging Field of Epitranscriptomics in Neurodevelopmental and Neuronal Disorders. Front. Bioeng. Biotechnol. 6:46. doi: 10.3389/fbioe.2018.00046

Analogous to DNA methylation and histone modifications, RNA modifications represent a novel layer of regulation of gene expression. The dynamic nature and increasing number of RNA modifications offer new possibilities to rapidly alter gene expression upon specific environmental changes. Recent lines of evidence indicate that modified RNA molecules and associated complexes regulating and "reading" RNA modifications play key roles in the nervous system of several organisms, controlling both, its development and function. Mutations in several human genes that modify transfer RNA (tRNA) have been linked to neurological disorders, in particular to intellectual disability. Loss of RNA modifications alters the stability of tRNA, resulting in reduced translation efficiency and generation of tRNA fragments, which can interfere with neuronal functions. Modifications present on messenger RNAs (mRNAs) also play important roles during brain development. They contribute to neuronal growth and regeneration as well as to the local regulation of synaptic functions. Hence, potential combinatorial effects of RNA modifications on different classes of RNA may represent a novel code to dynamically fine tune gene expression during brain function. Here we discuss the recent findings demonstrating the impact of modified RNAs on neuronal processes and disorders.

Keywords: RNA modification, m^5C, Nm, pseudouridine, m^6A, neurons, disease

INTRODUCTION

An estimated 1–2% of all genes in a given organism contribute to nucleic acid modification systems, suggesting biological importance of modified nucleotides (Grosjean, 2009). A classic example is the methylation of cytosine on DNA, which acts as a critical epigenetic regulator of gene expression (Bird, 2002). Additionally, current advances in RNA modification research report over 140 distinct post-transcriptional RNA modifications (Cantara et al., 2011; Machnicka et al., 2013). Initial knowledge has been derived from studies on abundant non-coding RNAs (ncRNAs), such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), in prokaryotes and simple eukaryotes. These pioneer investigations described a diverse, chemically complex, and strongly conserved nature of RNA nucleotide modifications (Cantara et al., 2011; Machnicka et al., 2013). The most heavily modified RNAs in any cell type and organism are tRNAs. Up to 20% of nucleotides

in mammalian cytoplasmic tRNAs carry modifications (Motorin and Helm, 2011; Pan, 2018). Modified nucleotides outside the anticodon loop of tRNAs occur non-randomly at conserved positions across diverse species and affect in general its stability (Helm, 2006; Motorin and Helm, 2010). In addition, modifications in the anticodon loop can contribute to optimize mRNA decoding by directly affecting codon-anticodon interactions (Agris, 2008).

Aberrant tRNA and rRNA modifications have been linked to various human disease syndromes and the phenotypes are often observed in specific tissues such as the gonads and the nervous system (Torres et al., 2014). Notably, increasing number of predicted human transfer RNA (tRNA) modification genes have been associated with neurological disorders, in particular with intellectual disability (ID) (for recent review see Bednárová et al., 2017). ID, or previously known as Mental Retardation (MR), is characterized by non-progressive cognitive impairment and affects 1–3% of the general population (Daily et al., 2000). It is presently unclear whether all observed phenotypes are caused by aberrant tRNA modifications, by effects on unidentified other RNA substrates (see below) and/or by a modification-independent function of the involved enzymes (Guo and Schimmel, 2013; Genenncher et al., 2018). Likewise, it is unknown why some tissues, in particular the brain, are more sensitive to the loss of these modifications.

Importantly, besides the heavily modified tRNAs and rRNAs, mRNAs, small and long non-coding RNAs were also found to harbor post-transcriptional modifications. Recent technological advances that allowed mapping of selected RNA modifications on a transcriptome-wide scale revealed widespread distribution of N6-methyladenosine (m^6A), pseudouridine (Ψ) and ribose 2′-O-methylation (Nm) on mRNA (Dominissini et al., 2012; Meyer et al., 2012; Carlile et al., 2014; Schwartz et al., 2014a; Dai et al., 2017). The prevalence of some others, including N1-methyladenine (m^1A) and 5-methylcytidine (m^5C) is still debated (Dominissini et al., 2016; Li et al., 2016, 2017c; Dominissini and Rechavi, 2017; Legrand et al., 2017; Safra et al., 2017). m^6A, the most abundant mRNA modification, was shown to affect almost every step of mRNA biogenesis, including splicing, export, translation, and mRNA decay (Lence et al., 2017; Roignant and Soller, 2017). It is thus not surprising that misregulation of m^6A results in several physiological defects, including brain development abnormalities, obesity, cancer, and other diseases (Batista, 2017; Dai et al., 2018). In addition, the discovery of m^6A RNA demethylases (Jia et al., 2011; Zheng et al., 2013; Jacob-Hirsch et al., 2018) and the identification of m^6A-binding proteins (Dominissini et al., 2012) indicated that similarly to DNA modification, RNA methylation can be reversible and convey information *via* recognition of effector proteins.

Altogether these recent studies revealed an entire new layer of regulation of gene expression, which has been central to the development of a novel concept called "RNA epigenetics or epitranscriptomics" (He, 2010; Meyer et al., 2012). However, the exact biological function of the majority of modified RNA nucleotides remains to be discovered. In this review, we will focus on several RNA modifications and will discuss their involvement in the development of the brain and neurological disorders.

5-METHYLCYTOSINE (m^5C)

Cytosine can be methylated at the 5^{th} position of the pyrimidine ring to form 5-methylcytosine (m^5C) (**Figure 1**). Various eukaryotic cytosine-5-RNA methyltransferases catalyze the formation of m^5C at specific positions (Motorin and Grosjean, 1999; Brzezicha et al., 2006; Sharma et al., 2013; Metodiev et al., 2014; Haag et al., 2015; Schosserer et al., 2015). The analysis of genetic mutations in two particular RNA cytosine-5 methyltransferase family members (Dnmt2/Trdmt and NCL1/TRM4/NSun2) has provided important insights into the biological effects of aberrant m^5C deposition.

Dnmt2

Dnmt2 is a member of the most widely conserved eukaryotic cytosine-5-DNA methyltransferase protein family (Goll and Bestor, 2005). Despite this classification, only few studies reported Dnmt2-mediated DNA methylation (Hermann et al., 2003; Kunert et al., 2003; Phalke et al., 2009) and it is today acknowledged that Dnmt2 functions mainly as a tRNA methylase (Okano et al., 1998; Schaefer and Lyko, 2010a,b; Raddatz et al., 2013). Dnmt2 methylation activity on position C38 of three tRNAs, which include $tRNA^{Asp}$, $tRNA^{Val}$, and $tRNA^{Gly}$, has been described in yeast, *Drosophila*, mouse, and human cells (Goll et al., 2006; Jurkowski et al., 2008; Schaefer et al., 2010). Knockdown of Dnmt2 in zebrafish embryos leads to differentiation defects in some organs, and notably, to abnormal neurogenesis in the hypothalamus and diencephalon (Rai et al., 2007). In *Dnmt2* mutant flies, reduced viability under stress conditions was observed (Schaefer et al., 2010). This is in accordance with previous studies that suggest an increased tolerance for stress in *Drosophila* and *Entamoeba* upon Dnmt2 overexpression (Lin et al., 2005; Fisher et al., 2006). Nevertheless, the majority of studies suggests that *Dnmt2* mutation does not trigger strong detrimental phenotypes in yeast, *Drosophila* and mice (Wilkinson et al., 1995; Kunert et al., 2003; Goll et al., 2006; Schaefer et al., 2010), which raises the question why zebrafish relies on Dnmt2 for proper development, whereas mice and flies do not. One possible explanation is that these organisms have redundant mechanisms that compensate for the loss of Dnmt2, which may be absent or less robust in zebrafish. Consistent with this possibility, it was shown that *Dnmt2* mutant mice exhibit lethal phenotypes in the absence of a second m^5C methyltransferase, NSun2 (Tuorto et al., 2012). In human, polymorphisms in *DNMT2* have been associated with spina bifida, a congenital malformation of the central nervous system (Franke et al., 2009).

NSun2

Unlike Dnmt2, it has been established that mammalian NSun2 does not only modify tRNAs (Blanco et al., 2011; Tuorto et al., 2012) but also other small ncRNAs such as 7SK, vault, and Y-RNAs (Hussain et al., 2013; Khoddami and Cairns, 2013). *Dnmt2* and *Nsun2* double knockout mice showed a lethal

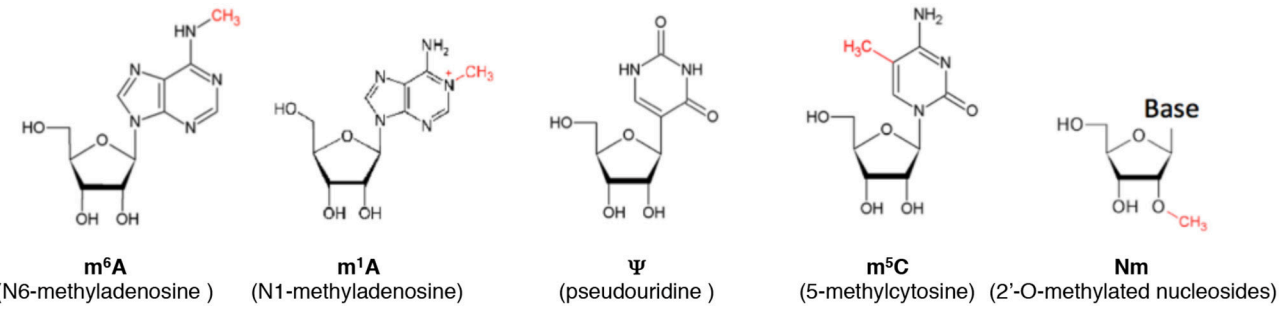

FIGURE 1 | Some of the most common modifications in RNA.

phenotype. However, deletion of *NSun2* alone (Blanco et al., 2011; Tuorto et al., 2012; Hussain et al., 2013) or in combination with *Dnmt2* (Rai et al., 2007; Tuorto et al., 2012) in specific tissues impairs cellular differentiation pathways in mammalian skin, testes, and brain. The function in the brain appears conserved as *Nsun2* mutations are associated with ID and Dubowitz-like syndrome in humans (Abbasi-Moheb et al., 2012; Khan et al., 2012; Martinez et al., 2012), as well as with microcephaly in human and mice (Blanco et al., 2014). In addition, a recent study in human and mice neuron precursor cells showed that m^5C deposited by Nsun2 regulates neural stem cell (NSC) differentiation and motility (Flores et al., 2017). This study thus provides some links between the failure of RNA m^5C deposition and the associated brain development diseases.

It is intriguing that patient fibroblasts and *Nsun2*-deficient mice (Blanco et al., 2014), as well as *Dnmt2* mutant flies (Durdevic et al., 2013), exhibit increase cleavage of tRNA, and elevated production of tRNA fragments (tRFs). This accumulation of tRFs reduces protein translation rates and increases oxidative stress as well as neuronal apoptosis. Interestingly, reducing tRNA cleavage in *Nsun2*-deficient brains is sufficient to rescue sensitivity to oxidative stress, implying that tRFs play a role in Nsun2-mediated defects.

(2′-O)-METHYLATION (Nm)

2′-O-methylation (Nm) is a common nucleoside modification of RNA, where a methyl group is added to the 2′ hydroxyl of the ribose moiety (**Figure 1**). Nm increases hydrophobicity, protects RNAs from nuclease attacks and stabilizes helical structures (Kurth and Mochizuki, 2009; Byszewska et al., 2014; Kumar et al., 2014; Yildirim et al., 2014). Nm is predominantly found internally in ribosomal RNAs and small nuclear RNAs as well as in tRNAs and in a number of sites on mRNA (Darzacq et al., 2002; Rebane et al., 2002; Kurth and Mochizuki, 2009; Zhao et al., 2012; Somme et al., 2014; Dai et al., 2017). This modification is also present at the 3′-end of miRNAs and siRNAs in plants (Li et al., 2005; Yu et al., 2005), as well as in siRNAs and piRNAs in animals (Horwich et al., 2007; Saito et al., 2007). Nm methyltransferases acting on tRNAs are highly conserved from bacteria and archaea to humans (Somme et al., 2014) and usually target positions in the anticodon loop. For instance, TRM7 in *S. cerevisiae* modifies positions 32 and 34 of selected tRNA, amongst which is $tRNA^{Phe}$ (Pintard et al., 2002; Guy et al., 2012). Strikingly, FTSJ1, the TRM7 ortholog in human, methylates the exact same positions of the exact same tRNAs (Guy and Phizicky, 2015; Guy et al., 2015). Consistent with a conserved function, expression of human FTSJ1 can suppress the severe growth defect of *S. cerevisiae Δtrm7* mutants (Pintard et al., 2002; Guy and Phizicky, 2015). Reduction of the modification level in $tRNA^{Phe}$ was reported in carcinoma and neuroblastoma in mice (Pergolizzi and Grunberger, 1980; Kuchino et al., 1982) and is associated with ID in human (see below).

FTSJ1

One of the best characterized associations between ID in human and mutations in a gene encoding for Nm is the one between non-syndromic X-linked ID (NSXLID) and mutations in the *FTSJ1* gene (OMIM:300499) (Guy et al., 2015). One third of the X-linked ID (XLID) conditions are syndromic (S-XLID) and the other two thirds are non-syndromic (NS-XLID) (Lubs et al., 2012). NSXLID is associated with no obvious and consistent phenotype other than mental retardation (IQ < 70), indeed NSXLID disorders are clinically diverse and genetically heterogeneous. *FTSJ1* loss of function causes NSXLID retardation in males (Froyen et al., 2007; Takano et al., 2008). Heterozygous loss of function mutations in females do not cause the disease, which is probably due to inactivation of the affected X chromosome**.** Several alleles of *FTSJ1* from six independent families correlate with NSXLID. All of these alleles lead to a reduction in mRNA levels and/or protein function (Willems et al., 1993; Hamel et al., 1999; Freude et al., 2004; Ramser et al., 2004; Froyen et al., 2007; Takano et al., 2008; Guy et al., 2015; **Table 1**). Consistently with the 2′-O-methyltransferase activity of FTSJ1 on tRNAs, Guy and Phizicky reported that two genetically independent lymphoblastoid cell lines (LCLs) of NSXLID patients with *FTSJ1* loss of function mutations nearly completely lack Cm_{32} and Gm_{34} on $tRNA^{Phe}$ (Guy et al., 2015). Additionally, $tRNA^{Phe}$ from a patient carrying an *FTSJ1-p.A26P* missense allele specifically lacks Gm_{34}, but has normal levels of Cm_{32}. $tRNA^{Phe}$ from the corresponding *Saccharomyces cerevisiae TRM7-A26P* mutant also specifically lacks Gm_{34}. Altogether, these findings strongly suggest that the absence of Gm_{34}, but not Cm_{32} modification on $tRNA^{Phe}$ causes NSXLID in patients

TABLE 1 | Proteins required for writing, reading, or removal of different RNA modifications and their mutations associated with altered brain functions.

Modification	Gene	Organism	Defect	Effect/Disease	References
m^5C	Dnmt2/Trdmt	*Dr*	KD	Abnormal neurogenesis in the hypothalamus and diencephalon. Defects in retina and liver.	Rai et al., 2007
		Dm	Lof	Decreased tolerance to stress (reduced viability under stress conditions).	Schaefer et al., 2010
	NSun2/NCL1/TRM4	*Hs*	Lof	Autosomal-recessive ID, facial dysmorphism, microcephaly and Dubowitz-like syndrome.	Abbasi-Moheb et al., 2012; Khan et al., 2012; Martinez et al., 2012
		Mm	Lof	Impaired cortical, hippocampal and striatal expansion during development Microcephaly. Decrease in neural stem cell (NSC) differentiation and motility.	Blanco et al., 2014; Flores et al., 2017
		Dm	Lof	Severe short-term memory deficits.	Abbasi-Moheb et al., 2012
	Dnmt2, NSUN2 double mutant	*Ms*	KO	Reduced proliferation rates, underdeveloped pheno- type in several tissues, including thickness and organization of the cerebral cortex.	Tuorto et al., 2012
Nm	FTSJ1	*Hs*	Lof	Nonsyndromic X-linked ID (NSXLID) in males.	Willems et al., 1993; Hamel et al., 1999; Freude et al., 2004; Guy et al., 2015
			SNPs	Impact on general cognitive ability, verbal comprehension, and perceptual organization in males.	Gong et al., 2008
			Gof	ID	Giorda et al., 2009; Honda et al., 2010
	TRMT44	*Hs*	SNPs	Partial epilepsy with pericentral spikes (PEPS).	Leschziner et al., 2011
	C/D box snoRNAs SNORD115 (HBII-52); SNORD116 (HBII-85) and others in the 15q11-q13 region	*Hs*	Lof	Prader-Willi syndrome (PWS).	Cavaillé et al., 2000; Kishore and Stamm, 2006; Peters, 2008; Sahoo et al., 2008; Sridhar et al., 2008; Doe et al., 2009; Duker et al., 2010
	15q11-q13 region	*Hs*	Gof	Autism	Bolton et al., 2004; Cook and Scherer, 2008
	Hen1	*Dm*	Lof	Accelerated neurodegeneration-related phenotypes (brain vacuolization, memory defaults and shorter life span).	Abe et al., 2014
Ψ	Unknown	*Hs*	Gof	Alzheimer's disease	Lee et al., 2007
		Hs	Lof	Myotonic dystrophy type 2 (DM2).	Delorimier et al., 2017
	Pus1 (TruA family member)	*Hs*	Lof	Mild-cognitive impairment, mitochondrial myopathy and sideroblastic anemia.	Cao et al., 2016
	Pus3 (TruA family member)	*Hs*	Lof	ID	Shaheen et al., 2016
	DKC1 (dyskerin)	*Hs*	Lof	X-linked recessive dyskeratosis congenita (DKC)	Heiss et al., 1998
m^6A	METTL3 (m^6A writer)	*Hs*	Lof	Impaired neuronal differentiation and formation of mature neurons from embryoid bodies.	Batista et al., 2014; Geula et al., 2015
		Dm	SNPs	Severe locomotion defects due to altered neuronal functions.	Haussmann et al., 2016; Lence et al., 2016; Kan et al., 2017
	Mettl14 (m^6A writer)	*Mm*	cKO	Delayed specification of different neuronal subtypes during brain development. Altered axon regeneration.	Yoon et al., 2017; Wang et al., 2018; Weng et al., 2018
	Wtap (m^6A writer)	*Dr*	KD	Smaller brain ventricles and curved notochord.	Ping et al., 2014
	ZC3H13	*Hs*	SNP	Schizophrenia	Oldmeadow et al., 2014
	Spenito (m^6A writer)	*Dm*	Lof	Control axon outgrowth, branching and synaptic bouton formation.	Gu et al., 2017
	Ythdc1 (m^6A reader)	*Dm*	KD	Enhancement of SCA1-induced neurodegeneration	Fernandez-Funez et al., 2000
	ALKBH5 (m^6A eraser)	*Hs*	SNP	Major depressive disorder (MDD).	Du et al., 2015

(Continued)

TABLE 1 | Continued

Modification	Gene	Organism	Defect	Effect/Disease	References
	FTO (m^6A eraser)	*Hs*	SNP	Decreased brain volume, increased risk for attention-deficit/hyperactivity disorder (ADHD) and Alzheimer's disease.	Keller et al., 2011; Reitz et al., 2012; Choudhry et al., 2013; Melka et al., 2013; Li et al., 2017a
		Mm	KO	Altered behavior (e.g., locomotion defects), abnormal electrophysiological response to cocaine (impaired dopamine type 2 and 3 receptor response) and enhanced consolidation of cued fear memory.	Hess et al., 2013; Widagdo et al., 2016

ID, intellectual disability; KD, knockdown; KO, knock out; cKO, conditional knock out; Lof, reduced function or loss of function; Gof, gain of function; SNP, single nucleotide polymorphism; Hs, Homo sapiens; Mm, Mus musculus; Dr, Danio rerio; Dm, Drosophila melanogaster N6-methyladenosine (m^6A); pseudouridine (Ψ), 5-methylcytosine (m^5C); and 2′-O-methylation (Nm).

carrying distinct *FTSJ1* alleles. Nevertheless, the molecular consequences arising from the loss of this 2′-O-methylation are not yet determined. Furthermore, it is noteworthy to mention two additional studies involving families from the Chinese Han population (Dai et al., 2008; Gong et al., 2008), where three single nucleotide polymorphisms (SNPs) in the *FTSJ1* gene were analyzed. Authors found a positive association with occurrence of NSXLID (Dai et al., 2008) as well as with general cognitive ability, verbal comprehension, and perceptual organization in male individuals (Gong et al., 2008). Although it seems tempting to link the variance of *FTSJ1* gene to general human cognitive ability, more profound studies are needed to support this idea.

TRMT44

TRMT44 is a putative 2′-O-methyluridine methyltransferase predicted to methylate residue 44 in $tRNA^{Ser}$ (Leschziner et al., 2011). Mutations in this gene were identified as a causative mutation in partial epilepsy with pericentral spikes (PEPS), a novel mendelian idiopathic epilepsy (Leschziner et al., 2011). However, the underlying mechanisms are currently unknown.

Small Nucleolar RNAs (snoRNAs)

snoRNAs are a class of regulatory RNAs responsible for post-transcriptional modification of ribosomal RNAs (rRNAs). Two families of snoRNAs have been described, based on their structure and function: C/D box snoRNAs are responsible for 2′-O-methylation (Cavaillé et al., 1996), whereas H/ACA box snoRNAs mediate pseudouridylation (Ganot et al., 1997 and see the following chapter in this review). In zebrafish, loss of three snoRNAs results in impaired rRNA modifications, causing severe developmental defects including growth delay and deformations in the head region (Higa-Nakamine et al., 2012). In human, C/D box snoRNAs have been implicated in Prader-Willi syndrome (PWS), a complex neurological disease characterized with mental retardation, low height, obesity, and muscle hypotonia (Sridhar et al., 2008; Doe et al., 2009). In several independent studies, PWS was shown to be caused by the loss of imprinted snoRNAs in locus 15q11-q13. Large deletions of this region underlie about 70% of cases of PWS (Peters, 2008), whereas duplication of the same region is associated with autism (Belmonte et al., 2004; Bolton et al., 2004; Cook and Scherer, 2008). Locus 15q11–q13 contains numerous copies of two C/D box snoRNAs—SNORD115 (HBII-52), and SNORD116 (HBII-85) (Cavaillé et al., 2000). SNORD115 is believed to play key roles in the fine-tuning of serotonin receptor (5-HT2C) by influencing its pre-mRNA splicing (Vitali et al., 2005; Kishore and Stamm, 2006; Falaleeva et al., 2017), whereas SNORD116 loss is thought to contribute to the etiology of the PWS (Cavaillé et al., 2000; Sahoo et al., 2008; Duker et al., 2010).

Hen1/Pimet

Hen1/Pimet is a conserved enzyme, which adds 2′-O-methyl group to 3′-terminal nucleotides of miRNAs and siRNAs in plants, and of siRNAs and piRNAs in animals. Addition of this modification protects these small non-coding RNAs (sncRNAs) from 3′ → 5′ exonuclease degradation (Li et al., 2005; Horwich et al., 2007; Saito et al., 2007; Terrazas and Kool, 2009; Ross et al., 2014). In the absence of Hen1/Pimet, piRNA, and siRNA are destabilized and sncRNA silencing activities are compromised. Surprisingly, *Hen1* mutant flies display neither increased lethality nor sterility under normal laboratory conditions but show however accelerated neurodegeneration (brain vacuolization), memory default, and shorter lifespan (Abe et al., 2014). This suggests a protective effect of Nm and small RNA pathways against age-associated neurodegenerative events. Accordingly, *Drosophila* lacking the siRNA effector, Argonaute 2 (Ago2), are viable but exhibit memory impairment and shortened lifespan (Li et al., 2013).

PSEUDOURIDINE (ψ)

Pseudouridine (also known as 5-ribosyluracil or ψ) is the first discovered (Cohn and Volkin, 1951) and most abundant RNA modification, present in a broad range of non-coding RNA, and was also recently detected in coding mRNA (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014a; Li et al., 2015). The isomerization of uridine into ψ improves the base stacking in RNAs by the formation of additional hydrogen bonds, which influences RNA secondary structure and increases the stability of RNA duplexes (Arnez and Steitz, 1994; Davis, 1995). Pseudouridylation was shown to have a strong impact on different aspects of cellular processes, including translation efficiency, splicing, telomere maintenance, and the

regulation of gene expression (Mochizuki et al., 2004; Carlile et al., 2014; Schwartz et al., 2014a). This base modification is catalyzed by pseudouridine synthases (Pus) that act on their substrates by two distinct mechanisms. One of those mechanisms is the guide RNA-dependent pseudouridylation, in which H/ACA box snoRNAs target RNAs for pseudouridylation *via* specific sequence interactions between the snoRNAs and the target RNA. A specific enzyme present in the snoRNP (snoribonucleoprotein) particule catalyzes the uridine modification (dyskerin in human, Cbf5 in yeast; Duan et al., 2009; Liang et al., 2009). Alternatively, RNA-independent pseudouridylation requires stand-alone pseudouridine synthases (Pus) that directly catalyze ψ formation at particular target RNA (Yu et al., 2011; Carlile et al., 2014; Rintala-Dempsey and Kothe, 2017). Each enzyme has a unique specificity for its target RNA and modifies uridine in a certain consensus sequence. Pus enzymes are present in all kingdoms of life, evolutionary conserved and are categorized into six families, based on their consensus sequences: TruA, TruB, TruD, RluA, and RsuA. The sixth family member, Pus10, is exclusive to eukaryotes and archaea (Koonin, 1996; Kaya and Ofengand, 2003; Fitzek et al., 2018).

Several pieces of evidence hint toward an implication of ψ in regulating neuronal functions. For instance, patients with mild-to-moderate severity of Alzheimer's disease show significantly elevated levels of urinal ψ (Lee et al., 2007) but it is currently unknown whether there is a link between this increase and the Alzheimer's disease etiology. Furthermore, it has been suggested that pseudouridylation can serve as a direct indicator of oxidative stress, which in turn has been linked to an increasing risk of neurodegeneration (Roth et al., 1999; Uttara et al., 2009). Accordingly, in cells exposed to acute oxidative stress by H_2O_2 treatment, Li and colleagues detected an elevation by ∼40–50% in mRNA ψ levels, demonstrating that mRNA pseudouridylation acts as a direct response to cellular stress (Li et al., 2015).

A recent report demonstrated a direct implication of ψ in neuronal disorders from patients with myotonic dystrophy type 2 (DM2) (Delorimier et al., 2017). DM2 is a neuromuscular disease characterized by severe gray matter changes, including neuronal loss and global neuronal impairment (Minnerop et al., 2011; Meola and Cardani, 2015). DM2 patients have an increased binding of Muscleblind-like 1 protein (MBNL1) to CCUG repeats in an intron of the *CNBP* gene (Cho and Tapscott, 2007). Interestingly, it was recently reported that pseudouridylation within CCUG repeats reduces RNA flexibility and thus modestly inhibits MBNL1 binding (Delorimier et al., 2017). Similarly, ψ modification of a minimally structured model RNA resulted in an even more drastic reduction of MBNL1 binding to CCUG repeats. This study shows that ψ can reduce the disease-causing binding of MBNL1 at extended CCUG repeats and offers a basis for future research in treating neurodegenerative diseases.

Pus1

Pus1 is a member of the TruA family that typically pseudouridylates tRNA but was also recently found to act on rRNA, snRNA, and mRNA (Schwartz et al., 2014a; Carlile et al., 2015). Mutations of *Pus1* in human lead to mitochondrial myopathy and sideroblastic anemia (Bykhovskaya et al., 2004; Fernandez-Vizarra et al., 2007; Bergmann et al., 2010). Recently, a mild cognitive impairment was also characterized in a long-surviving patient with two novel *Pus1* mutations (Cao et al., 2016). A different study demonstrated a Pus1-dependent pseudouridylation of the steroid RNA activator (SRA). Pseudouridylated SRA acts as a co-activator of the nuclear estrogen receptor α (ERα) (Zhao et al., 2004; Leygue, 2007). Given that ERα was shown to regulate neuronal survival (Gamerdinger et al., 2006; Foster, 2012), it is conceivable that one of the functions of Pus1 in brain activity is mediated *via* the control of the ER pathway.

Pus3

Pus3 is another member of the TruA family, which has a strong sequence homology to Pus1 but acts on distinct target RNA. *In situ* hybridization showed accumulation of *Pus3* mRNA in the nervous system of mice embryos, suggesting a role of Pus3 in neural development (Diez-Roux et al., 2011). Accordingly, a truncated form of Pus3 accompanied by reduced levels of ψ U39 in tRNA was detected in patients with ID (Shaheen et al., 2016). Taken together, the discovery of impaired cognition caused by mutations of TruA enzymes emphasizes their importance in neuronal development and maintenance regulation.

Dyskerin and RluA-1

The pseudouridine synthase dyskerin is essential for the H/ACA-box mediated pseudouridylation in human (Heiss et al., 1998; Lafontaine et al., 1998). Mutations of the dyskerin-encoding gene DKC1 causes X-linked recessive dyskeratosis congenita (DKC), a rare progressive congenital disorder that mostly affects highly regenerative tissues, such as the skin and bone marrow (Heiss et al., 1998; Mochizuki et al., 2004). Cells of the affected patients have decreased telomerase activity and thus reduced telomere length, which may be responsible for the disease (Mitchell et al., 1999). Interestingly, expression analysis of *Dyskerin 1* showed a high level in embryonic neural tissue, as well as in specific subsets of neurons in the cerebellum and olfactory bulb of adult brains (Heiss et al., 2000). While the function of dyskerin in hematopoiesis has been studied intensively, we are yet lacking a detailed understanding about its potential nervous system function in adult brains. In *Drosophila melanogaster*, RluA enzymes modify uridines in rRNA and tRNA. *In situ* hybridization in embryos revealed a specific RluA-1 mRNA localization to dendrites of a subset of peripheral neurons, which also raises the question about the molecular function of RluA-1 and its target RNA in the peripheral nervous system during embryonic development (Wang et al., 2011).

N6-METHYLADENOSINE (m^6A)

m^6A is an abundant mRNA modification that regulates nearly all aspects of mRNA processing including splicing, export, translation, stability, and decay (Meyer and Jaffrey, 2017; Roignant and Soller, 2017; Roundtree et al., 2017). This modification is catalyzed by a stable protein complex composed of two methyltransferases, Methyltransferase like-3 (Mettl3) and

Methyltransferase like-14 (Mettl14) (Sledz and Jinek, 2016; Wang et al., 2016a,b; Schöller et al., 2018). Additional proteins required for m^{6}A deposition are Wilms' tumor 1-associating protein (Wtap) (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014), Vir like m^{6}A methyltransferase associated (Virma) (Schwartz et al., 2014b; Yue et al., 2018), Zinc finger CCCH domain-containing protein 13 (Zc3h13) (Guo et al., 2018; Knuckles et al., 2018; Wen et al., 2018), RNA binding protein 15 (Rbm15) and its paralog Rbm15B (Patil et al., 2016). Mettl3 has the catalytic activity and can accommodate the SAM substrate, while Mettl14 serves to stabilize the binding to RNA (Sledz and Jinek, 2016; Wang et al., 2016a,b; Schöller et al., 2018). In vertebrates, m^{6}A modification is dynamically regulated and can be reversed by two demethylases belonging to the family of α-ketoglutarate dependent dioxygenases, Fat mass and obesity associated protein (FTO) and ALKBH5 (Jia et al., 2011; Zheng et al., 2013). Recent advances in techniques to map m^{6}A modification in a transcriptome wide manner enabled identification of thousands of modified mRNAs and lncRNAs (Dominissini et al., 2012; Meyer et al., 2012). While m^{6}A has been involved in many physiological processes, increasing evidence suggests an importance of m^{6}A modification in brain development and in the function of the nervous system.

m^{6}A Writer Complex

m^{6}A levels are particularly high in the nervous system, as shown in the developing mouse brain (Meyer et al., 2012), and in heads of adult flies (Lence et al., 2016). Furthermore, a recent study detected higher m^{6}A content in the mouse cerebellum and in neurons compared to glia (Chang et al., 2017). Using *in situ* hybridization in zebrafish embryos, Ping et al. showed that Wtap is ubiquitously expressed at 36 h post-fertilization with enrichment in the brain region (Ping et al., 2014). Consistently, Wtap depletion using morpholino treatment resulted in severe developmental defects, including appearance of smaller brain ventricles and curved notochord at 24 h post-fertilization. Importance of m^{6}A during neuronal development was further demonstrated by depletion of METTL3 in human embryonic stem cells (hESC), which strongly impaired neuronal differentiation (Batista et al., 2014), as well as the formation of mature neurons from embryoid bodies (Geula et al., 2015). Notably, m^{6}A mRNA modification is essential for mouse survival as mice lacking Mettl3 die at E6.5 (Geula et al., 2015). However, two recent studies performed a conditional KO (cKO) of Mettl14 specifically in neurons and revealed an essential role of m^{6}A in embryonic cortical neurogenesis (Yoon et al., 2017; Wang et al., 2018). Mettl14 cKO animals showed a decreased NSC proliferation and premature differentiation of NSCs (Wang et al., 2018), as well as delayed specification of different neuronal subtypes during brain development (Yoon et al., 2017). Yoon at al. further demonstrated that m^{6}A modification is required for timely decay of transcripts involved in stem cell maintenance and cell cycle regulation in cortical neuronal progenitors. This allows accurate progression of the cell cycle and in turn induces the spatiotemporal formation of different neuronal subtypes. Interestingly, the authors also observed that many transcripts linked to mental disorders (autism, schizophrenia) are m^{6}A modified in human, but not in mouse cultures of neuronal progenitor cells (NPC), raising the possibility that m^{6}A regulates specifically these human diseases (Yoon et al., 2017). Consistent with this hypothesis, polymorphisms in *ZC3H13* have been associated with schizophrenia (Oldmeadow et al., 2014).

Beyond the role in neuronal development, m^{6}A modification also plays a critical role in the process of axon regeneration in mature mouse neurons (Weng et al., 2018). Weng et al. showed that upon peripheral nerve injury m^{6}A levels of many transcripts in dorsal root ganglion (DRG) were elevated, which led to increased translation during the time of axon regeneration, *via* the specific m^{6}A reader protein Ythdf1. Mettl14 and Ythdf1 conditional KO mice displayed strong reduction of sensory axon regeneration, resulting from reduced protein synthesis, revealing the critical role of m^{6}A modification in response to injury (Weng et al., 2018).

In *Drosophila*, loss of components of the methyltransferase complex results in severe locomotion defects due to altered neuronal functions (Haussmann et al., 2016; Lence et al., 2016; Kan et al., 2017). *Mettl3* mutants display alterations in walking speed and orientation, which can be rescued by ectopic expression of Mettl3 cDNA in neurons. Whether a particular subset of neurons is responsible for the observed alterations awaits further investigations. Interestingly, another member of the m^{6}A methyltransferase complex, Nito (RBM15 in human), was recently shown to control axon outgrowth, branching and to regulate synaptic bouton formation *via* the activity of the CCAP/bursicon neurons (Gu et al., 2017), providing first insights toward addressing this question.

m^{6}A Readers

As mentioned above, most characterized functions of m^{6}A rely on the direct binding of the so-called m^{6}A "reader" proteins to the modified site. The best-studied m^{6}A readers are the YTH-domain containing proteins, which can specifically bind m^{6}A *via* their YTH domain (Luo and Tong, 2014; Theler et al., 2014; Xu et al., 2014). RNA *in situ* hybridization of rat brain sections showed that one particular member of the YTH family, Ythdc1, is enriched in specific cells in the brain (Hartmann et al., 1999). Interestingly, in a yeast two-hybrid screen to identify Ythdc1-interacting proteins, libraries from P5 and E16 brains were screened and the rat homolog of Sam68 was found as the main interactor (Hartmann et al., 1999). Sam68 is known to regulate neuronal activity-dependent alternative splicing events (i.e., of neurexin-1) (Iijima et al., 2011). In line with this function, *in situ* hybridization assays showed that *Drosophila Ythdc1* localizes in the ventral neurectoderm and central nervous system of *Drosophila* embryos (Lence et al., 2016) and a reduced level of Ythdc1 was found to enhance SCA1-induced neurodegeneration (Fernandez-Funez et al., 2000).

Apart from the YTH-protein family of conventional m^{6}A reader proteins, a number of other proteins that bind RNA in m^{6}A-dependent fashion has been recently identified (Edupuganti et al., 2017). Among them, Fragile X mental retardation protein (FMRP, also known as POF; FMR1; POF1; FRAXA) was shown to preferentially bind an RNA probe containing m^{6}A sites (Edupuganti et al., 2017). FMRP plays critical roles in synaptic

plasticity and neuronal development. Its loss of function in human leads to the Fragile X syndrome, which is the most prevalent form of inherited ID and the foremost monogenic cause of autism (Bardoni et al., 2001; Lubs et al., 2012; Wang et al., 2012; Hagerman and Polussa, 2015). FMRP has a central role in neuronal development and synaptic plasticity through the regulation of alternative mRNA splicing, mRNA stability, mRNA dendritic transport and postsynaptic local protein synthesis of a subset of mRNAs (Antar et al., 2006; Didiot et al., 2008; Bechara et al., 2009; Ascano et al., 2012; Guo et al., 2015). Moreover, it represses mRNA translation during the transport of dendritic mRNAs to postsynaptic dendritic spines and activates mRNA translation of a subset of dendritic mRNAs at synapses (Bechara et al., 2009; Fähling et al., 2009). Consistent with a potential interplay between FMRP and m^6A, a recent study found that m^6A is present on many synaptic mRNAs that are known targets of FMRP protein (Chang et al., 2017). Future research will seek to further illuminate the potential role of FMRP within the m^6A pathway. It is interesting to note that Nm also appears to contribute to FMRP-mediated translation regulation at synapses. FMRP can form a complex with the non-coding RNA, *brain cytoplasmic RNA* (*BC1*), to repress translation of a subset of FMRP target mRNAs (Zalfa et al., 2003). This interaction is modulated by the Nm status of *BC1* RNA. In both nucleus and cytoplasm in the cell body, Nm is present on *BC1*, but it is virtually absent at synapses (Lacoux et al., 2012). The authors suggested that changes in the 2′-O-methylation status of *BC1* RNA contribute to the fine-tuned regulation of gene expression at synapses and consequently to neuronal plasticity by influencing FMRP local translational control. This example supports a likely combinatorial role of RNA modifications in the regulation of similar targets and/or processes during brain function.

m^6A Erasers

Two proteins in humans were reported to act as m^6A erasers: (FTO) and AlkB homolog 5 (ALKBH5) (Jia et al., 2011; Zheng et al., 2013). Both belong to the family of Fe^{2+}-α-ketoglutarate-dependent deoxygenases and catalyze the removal of the methyl group of m^6A by oxidation. Interestingly, even though ALKBH5 is only moderately expressed in the brain, it has been associated with mental disorders. Du et al. found that certain polymorphisms within the *ALKBH5* gene correlate with the major depressive disorder (MDD), suggesting an involvement of ALKBH5 in conferring risk of MDD (Du et al., 2015).

In comparison, FTO is highly expressed in the human brain, especially in the hypothalamus and the pituitary gland and displays dynamic expression during postnatal neurodevelopment. Polymorphic alleles of *FTO* in human were identified to increase the risk for hyperactive disorder (Choudhry et al., 2013), for Alzheimer's disease (Keller et al., 2011; Reitz et al., 2012) and to affect the brain volume (Melka et al., 2013; Li et al., 2017a). Molecular and functional studies have shown that *FTO* knockout mice display altered behavior, including locomotion defects, but also influences learning and memory. For instance, fear conditioned mice showed a significant increase in m^6A intensity on several neuronal targets, and knockdown of FTO further enhanced consolidation of cued fear memory (Widagdo et al., 2016). In line with this finding, *FTO* deficiency reduces the proliferation and neuronal differentiation of adult NSCs, which leads to impaired learning and memory (Li et al., 2017a). Electrophysiological tests in *FTO* knockout mice demonstrated an impaired dopamine type 2 and 3 receptor response, resulting in an abnormal response to cocaine (Hess et al., 2013). Strikingly, a recent study in mouse embryonic dorsal root ganglia found that FTO is enriched and specifically expressed in axons, influencing translation of axonal mRNAs (Yu et al., 2018). This demonstrates the dynamic role of m^6A modification in regulating local translation. However, it is important to stress that FTO was also demonstrated to additionally demethylate m^6Am, which is present next to the 7mG cap modification (Mauer et al., 2017). In comparison to m^6A, m^6Am also contains a methyl group on the ribose. As available antibodies recognize both modifications indistinguishably, it is currently difficult to assign their respective contribution in the context of brain activity.

CONCLUSION

The association of aberrant RNA modifications with various neurological disorders highlights the importance of these chemical moieties for proper brain development and cognition. However, today the role of RNA modifications in these processes is not completely understood. One of the current challenges lies in identifying the class and identity of RNAs that are targeted by RNA modification enzymes and are causative of the neurological defects. Common targets for many of these enzymes are tRNAs and rRNAs, thus it is likely that in many cases their dysfunction plays an important role in the etiology of the disease. Yet, this does not explain why the phenotypes observed upon mutations of these enzymes are often restricted to the brain. Some of these enzymes or snoRNAs are predominantly expressed in the nervous system, which indicates their importance in this tissue and suggests the existence of differentially modified ribosomes (also recently called specialized ribosomes) that may carry distinct functions (Briggs and Dinman, 2017; Sloan et al., 2017). However, other snoRNA or enzymes exhibit wider tissue distribution, suggesting that brain specific phenotypes may reflect a higher sensitivity to altered translation in this tissue compared to others organs. This would be consistent with other disease-causing mutations in ribosomal proteins or tRNA synthetase genes that also manifest their effect specifically in the nervous system (Antonellis et al., 2003; Antonellis and Green, 2008; Yao and Fox, 2013; Brooks et al., 2014). What could be the reason(s) behind this increased sensitivity? It has been observed that tRNA level is in general higher in the brain compared to other tissues, suggestive of a bigger translational demand (Dittmar et al., 2006). This neuronal specificity may arise from the local translation that occurs at synapses upon environmental changes. In this context, RNA modifications represent an attractive system to regulate this acute need in a dynamic and flexible manner. However, other examples point toward a translation-independent mechanism. For instance, in the case of *FTSJ1* mutations and the associated NSXLID, neither

the amount, nor the charging of the concerned tRNAs appear to be affected (Guy et al., 2012, 2015), seemingly ruling out a general translation defect. Perhaps in this case, the absence of modification can stimulate tRNA cleavage and generate tRFs that can also interfere with translation. This increase in tRFs would not necessarily be associated with a corresponding reduction of the uncleaved tRNA since tRNA levels are tightly regulated (Wilusz, 2015). Alternatively, the absence of modification on the tRNA may affect its interaction with the ribosome and thus influence translation efficiency or fidelity. Therefore, careful examination at different molecular levels is required to appreciate the effect of tRNA and rRNA modification enzyme mutations on translation and their consequence on the neurological phenotype.

That being said, the general picture is probably more complex. For instance, recent reports show that tRFs not only interfere with translation but can also affect transposon regulation and genome stability (Durdevic et al., 2013; Martinez et al., 2017; Schorn et al., 2017; Zhang et al., 2017). This activity could in principle also contribute to neurological disorders as growing evidence suggests associations between (re)expression of transposable elements and the occurrence of neuropathies (Perrat et al., 2013; Krug et al., 2017; Zahn, 2017; Jacob-Hirsch et al., 2018). In addition, beyond tRNA and tRF, the recent studies on m^{6}A clearly demonstrate the involvement of mRNA modification in different aspects of neuronal development and regulation. The large diversity of RNA processing events in the brain, including the high rates of alternative splicing and recursive splicing (Duff et al., 2015; Sibley et al., 2015), the inclusion of microexons (Irimia et al., 2014) and the biogenesis of circular RNAs (cirRNAs) can all in principle be affected by RNA modifications. For instance, m^{6}A modification was recently found on circRNAs, and enables their translation (Yang et al., 2017; Zhou et al., 2017). Given that misexpression of circRNAs has been associated with neurological disorders (Shao and Chen, 2016; van Rossum et al., 2016; Li et al., 2017b), some of m^{6}A brain functions may rely on circRNA-mediated regulation. Thus, because neurons face distinct challenges with regards to localisation of RNAs to distal processes and to localized translation it is likely that in the brain, more than in any other tissues, a combinatorial effect of RNA modifications on different classes of RNAs represents a critical informational

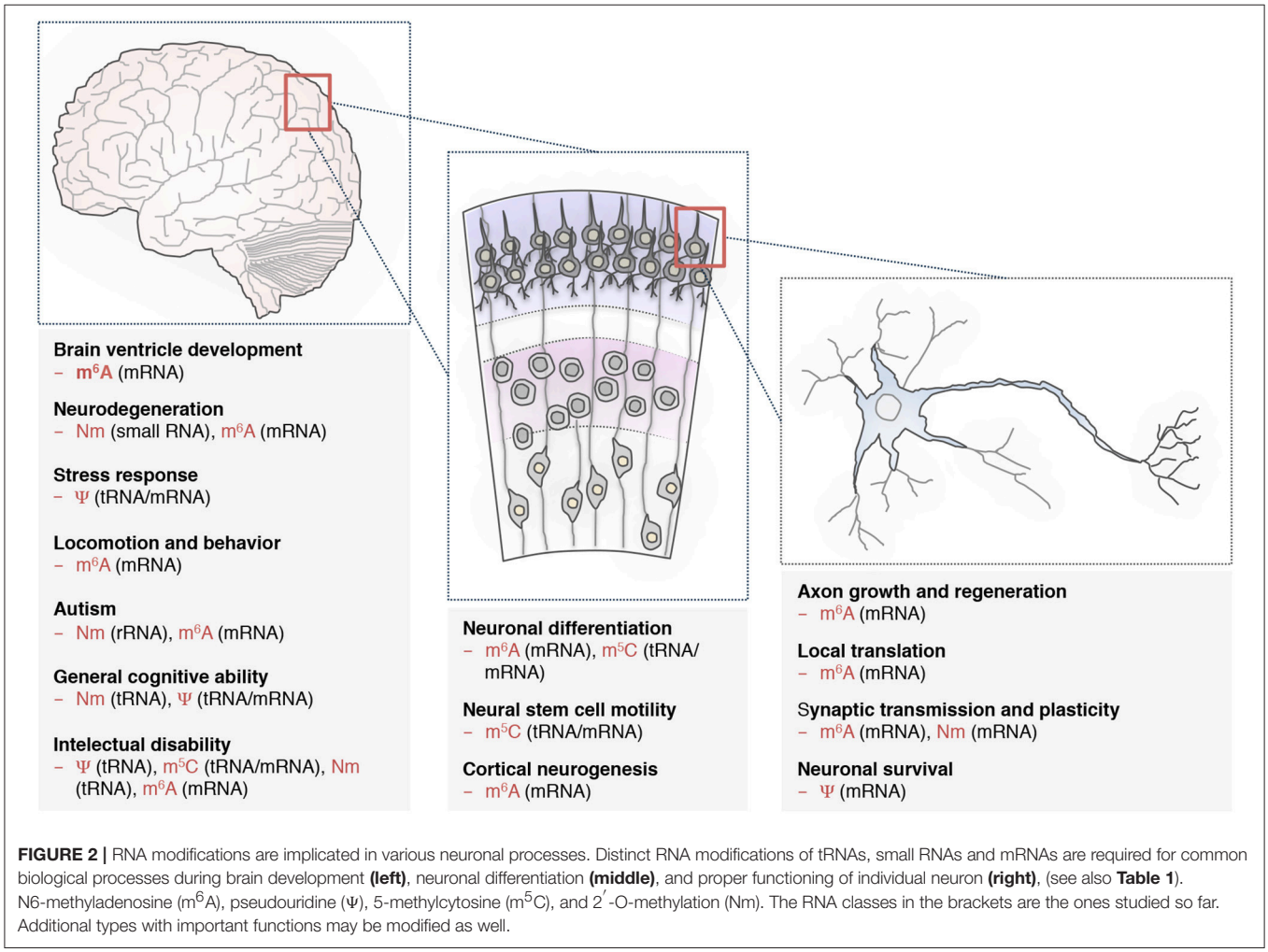

FIGURE 2 | RNA modifications are implicated in various neuronal processes. Distinct RNA modifications of tRNAs, small RNAs and mRNAs are required for common biological processes during brain development **(left)**, neuronal differentiation **(middle)**, and proper functioning of individual neuron **(right)**, (see also **Table 1**). N6-methyladenosine (m^{6}A), pseudouridine (Ψ), 5-methylcytosine (m^{5}C), and 2′-O-methylation (Nm). The RNA classes in the brackets are the ones studied so far. Additional types with important functions may be modified as well.

layer that dynamically fine-tunes gene regulation (**Figure 2**). Exciting discoveries are lying ahead for deciphering this intricate epitranscriptomics code.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The work presented was supported by DFG (RO4681/5-1), the DIP (RO4681/6-1) to J-YR, by the IBPS Actions Incitatives 2018 to CC, the French MESR to DD and an ARC Ph.D. fellowship to MA. CC acknowledges the Réseau André Picard for its financial support. J-YR and CC are members of the European Epitranscriptomic COST Action (CA16120).

REFERENCES

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., et al. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 90, 847–855. doi: 10.1016/j.ajhg.2012.03.021

Abe, M., Naqvi, A., Hendriks, G. J., Feltzin, V., Zhu, Y., Grigoriev, A., et al. (2014). Impact of age-associated increase in 2′-O-methylation of miRNAs on aging and neurodegeneration in Drosophila. *Genes Dev.* 28, 44–57. doi: 10.1101/gad.226654.113

Agris, P. F. (2008). Bringing order to translation, the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.* 9, 629–635. doi: 10.1038/embor.2008.104

Antar, L. N., Li, C., Zhang, H., Carroll, R. C., and Bassell, G. J. (2006). Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses. *Mol. Cell. Neurosci.* 32, 37–48. doi: 10.1016/j.mcn.2006.02.001

Antonellis, A., and Green, E. D. (2008). The role of aminoacyl-tRNA synthetases in genetic diseases. *Annu. Rev. Genomics Hum. Genet.* 9, 87–107. doi: 10.1146/annurev.genom.9.081307.164204

Antonellis, A., Ellsworth, R. E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S. Q., et al. (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am. J. Hum. Genet.* 72, 1293–1299. doi: 10.1086/375039

Arnez, J. G., and Steitz, T. A. (1994). Crystal-structure of unmodified Trna(Gln) complexed with glutaminyl-transfer-rna synthetase and atp suggests a possible role for pseudo-uridines in stabilization of Rna structure. *Biochemistry* 33, 7560–7567. doi: 10.1021/bi00190a008

Ascano, M., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382–386. doi: 10.1038/nature11737

Bardoni, B., Schenck, A., and Mandel, J. L. (2001). The fragile X mental retardation protein. *Brain Res. Bull.* 56, 375–382. doi: 10.1016/S0361-9230(01)00647-5

Batista, P. J. (2017). The RNA modification N(6)-methyladenosine and Its implications in human disease. *Genomics Proteomics Bioinformatics* 15, 154–163. doi: 10.1016/j.gpb.2017.03.002

Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* 15, 707–719. doi: 10.1016/j.stem.2014.09.019

Bechara, E. G., Didiot, M. C., Melko, M., Davidovic, L., Bensaid, M., Martin, P., et al. (2009). A novel function for fragile x mental retardation protein in translational activation. *PLoS Biol.* 7:e16. doi: 10.1371/journal.pbio.1000016

Bednárová, A., Hanna, M., Durham, I., VanCleave, T., England, A., Chaudhuri, A., et al. (2017). Lost in translation: defects in transfer rna modifications and neurological disorders. *Front. Mol. Neurosci.* 10:135. doi: 10.3389/fnmol.2017.00135

Belmonte, M. K., Cook, E. H., Anderson, G. M., Rubenstein, J. L. R., Greenough, W. T., Beckel-Mitchener, A., et al. (2004). Autism as a disorder of neural information processing: directions for research and targets for therapy. *Mol. Psychiatry* 9, 646–663. doi: 10.1038/sj.mp.4001499

Bergmann, A. K., Campagna, D. R., McLoughlin, E. M., Agarwal, S., Fleming, M. D., Bottomley, S. S., et al. (2010). Systematic molecular genetic analysis of congenital sideroblastic anemia: evidence for genetic heterogeneity and identification of novel mutations. *Pediatr. Blood Cancer* 54, 273–278. doi: 10.1002/pbc.22244

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102

Blanco, S., Dietmann, S., Flores, J. V., Hussain, S., Kutter, C., Humphreys, P., et al. (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.* 33, 2020–2039. doi: 10.15252/embj.201489282

Blanco, S., Kurowski, A., Nichols, J., Watt, F. M., Benitah, S. A., and Frye, M. (2011). The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate. *PLoS Genet.* 7:e1002403. doi: 10.1371/journal.pgen.1002403

Bolton, P. F., Veltman, M. W., Weisblatt, E., Holmes, J. R., Thomas, N. S., Youings, S. A., et al. (2004). Chromosome 15q11-13 abnormalities and other medical conditions in individuals with autism spectrum disorders. *Psychiatr. Genet* 14, 131–137. doi: 10.1097/00041444-200409000-00002

Briggs, J. W., and Dinman, J. D. (2017). Subtractional heterogeneity: a crucial step toward defining specialized ribosomes. *Mol. Cell* 67, 3–4. doi: 10.1016/j.molcel.2017.06.022

Brooks, S. S., Wall, A. L., Golzio, C., Reid, D. W., Kondyles, A., Willer, J. R., et al. (2014). A novel ribosomopathy caused by dysfunction of RPL10 disrupts neurodevelopment and causes X-linked microcephaly in humans. *Genetics* 198, 723–733. doi: 10.1534/genetics.114.168211

Brzezicha, B., Schmidt, M., Makalowska, I., Jarmolowski, A., Pienkowska, J., and Szweykowska-Kulinska, Z. (2006). Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). *Nucleic Acids Res.* 34, 6034–6043. doi: 10.1093/nar/gkl765

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A., and Fischel-Ghodsian, N. (2004). Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am. J. Hum. Genet.* 74, 1303–1308. doi: 10.1086/421530

Byszewska, M., Smietanski, M., Purta, E., and Bujnicki, J. M. (2014). RNA methyltransferases involved in 5' cap biosynthesis. *RNA Biol.* 11, 1597–1607. doi: 10.1080/15476286.2015.1004955

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., et al. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39, D195–D201. doi: 10.1093/nar/gkq1028

Cao, M., Dona, M., Valentino, M. L., Semplicini, C., Maresca, A., Cassina, M., et al. (2016). Clinical and molecular study in a long-surviving patient with MLASA syndrome due to novel PUS1 mutations. *Neurogenetics* 17, 143–143. doi: 10.1007/s10048-016-0475-3

Carlile, T. M., Rojas-Duran, M. F., and Gilbert, W. V. (2015). Transcriptome-wide identification of pseudouridine modifications using Pseudo-seq. *Curr. Protoc. Mol. Biol.* 112, 4. 25. 21-24. doi: 10.1002/0471142727.mb0425s112

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146. doi: 10.1038/nature13802

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., et al. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14311–14316. doi: 10.1073/pnas.250426397

Cavaillé, J., Nicoloso, M., and Bachellerie, J. P. (1996). Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature* 383, 732–735. doi: 10.1038/383732a0

Chang, M., Lv, H., Zhang, W., Ma, C., He, X., Zhao, S., et al. (2017). Region-specific RNA m(6)A methylation represents a new layer of control in the gene regulatory network in the mouse brain. *Open Biol.* 7:170166. doi: 10.1098/rsob.170166

Cho, D. H., and Tapscott, S. J. (2007). Myotonic dystrophy: emerging mechanisms for DM1 and DM2. *Biochim Biophys. Acta* 1772, 195–204. doi: 10.1016/j.bbadis.2006.05.013

Choudhry, Z., Sengupta, S. M., Grizenko, N., Thakur, G. A., Fortier, M. E., Schmitz, N., et al. (2013). Association between obesity-related gene FTO and ADHD. *Obesity* 21, E738–E744. doi: 10.1002/oby.20444

Cohn, W. E., and Volkin, E. (1951). Nucleoside-5′-phosphates from ribonucleic acid. *Nature* 167, 483–484. doi: 10.1038/167483a0

Cook, E. H., and Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature* 455, 919–923. doi: 10.1038/nature07458

Dai, D., Wang, H., Zhu, L., Jin, H., and Wang, X. (2018). N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis.* 9:124. doi: 10.1038/s41419-017-0129-x

Dai, L., Xing, L. , Gong, P., Zhang, K., Gao, X., Zheng, Z., et al. (2008). Positive association of the FTSJ1 gene polymorphisms with nonsyndromic X-linked mental retardation in young Chinese male subjects. *J. Hum. Genet.* 53, 592–597. doi: 10.1007/s10038-008-0287-x

Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., et al. (2017). Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* 14, 695–698. doi: 10.1038/nmeth.4294

Daily, D. K., Ardinger, H. H., and Holmes, G. E. (2000). Identification and evaluation of mental retardation. *Am. Fam. Physician* 61, 1059–1067. Available online at: https://www.aafp.org/afp/2000/0215/p1059.html

Darzacq, X., Jády, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.* 21, 2746–2756. doi: 10.1093/emboj/21.11.2746

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* 23, 5020–5026. doi: 10.1093/nar/23.24.5020

Delorimier, E., Hinman, M. N., Copperman, J., Datta, K., Guenza, M., and Berglund, J. A. (2017). Pseudouridine modification inhibits muscleblind-like 1 (MBNL1) binding to CCUG repeats and minimally structured RNA through reduced RNA flexibility. *J. Biol. Chem.* 292, 4350–4357. doi: 10.1074/jbc.M116.770768

Didiot, M. C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J. L., and Moine, H. (2008). The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* 36, 4902–4912. doi: 10.1093/nar/gkn472

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* 9:e1000582. doi: 10.1371/journal.pbio.1000582

Dittmar, K. A., Goodenbour, J. M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS Genet.* 2:e221. doi: 10.1371/journal.pgen.0020221

Doe, C. M., Relkovic, D., Garfield, A. S., Dalley, J. W., Theobald, D. E. H., Humby, T., et al. (2009). Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT(2C)R-mediated behaviour. *Hum. Mol. Genet.* 18, 2140–2148. doi: 10.1093/hmg/ddp137

Dominissini, D., and Rechavi, G. (2017). Loud and clear epitranscriptomic m(1)A signals: now in single-base resolution. *Mol. Cell* 68, 825–826. doi: 10.1016/j.molcel.2017.11.029

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., et al. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530, 441–446. doi: 10.1038/nature16998

Du, T., Rao, S., Wu, L., Ye, N., Liu, Z., Hu, H., et al. (2015). An association study of the m6A genes with major depressive disorder in Chinese Han population. *J. Affect. Disord.* 183, 279–286. doi: 10.1016/j.jad.2015.05.025

Duan, J., Li, L., Lu, J., Wang, W., and Ye, K. (2009). Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol. Cell* 34, 427–439. doi: 10.1016/j.molcel.2009.05.005

Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., et al. (2015). Genome-wide identification of zero nucleotide recursive splicing in Drosophila. *Nature* 521, 376–379. doi: 10.1038/nature14475

Duker, A. L., Ballif, B. C., Bawle, E. V., Person, R. E., Mahadevan, S., Alliman, S., et al. (2010). Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur. J. Hum. Genet.* 18, 1196–1201. doi: 10.1038/ejhg.2010.102

Durdevic, Z., Mobin, M. B., Hanna, K., Lyko, F., and Schaefer, M. (2013). The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. *Cell Rep.* 4, 931–937. doi: 10.1016/j.celrep.2013.07.046

Edupuganti, R. R., Geiger, S., Lindeboom, R. G. H., Shi, H., Hsu, P. J., Lu, Z., et al. (2017). N(6)-methyladenosine (m(6)A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* 24, 870–878. doi: 10.1038/nsmb.3462

Fähling, M., Mrowka, R., Steege, A., Kirschner, K. M., Benko, E., Förstera, B., et al. (2009). Translational regulation of the human achaete-scute homologue-1 by fragile X mental retardation protein. *J. Biol. Chem.* 284, 4255–4266. doi: 10.1074/jbc.M807354200

Falaleeva, M., Welden, J. R., Duncan, M. J., and Stamm, S. (2017). C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: old dogs show new tricks. *BioEssays* 39:1600264. doi: 10.1002/bies.201600264

Fernandez-Funez, P., Nino-Rosales, M. L., de Gouyon, B., She, W. C., Luchak, J. M., Martinez, P., et al. (2000). Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature* 408, 101–106. doi: 10.1038/35040584

Fernandez-Vizarra, E., Berardinelli, A., Valente, L., Tiranti, V., and Zeviani, M. (2007). Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by myopathy, lactic acidosis and sideroblastic anaemia (MLASA). *J. Med. Genet.* 44, 173–180. doi: 10.1136/jmg.2006.045252

Fisher, O., Siman-Tov, R., and Ankri, S. (2006). Pleiotropic phenotype in Entamoeba histolytica overexpressing DNA methyltransferase (Ehmeth). *Mol. Biochem. Parasitol.* 147, 48–54. doi: 10.1016/j.molbiopara.2006.01.007

Fitzek, E., Joardar, A., Gupta, R., and Geisler, M. (2018). Evolution of eukaryal and archaeal pseudouridine synthase pus10. *J. Mol. Evol.* 86, 77–89. doi: 10.1007/s00239-018-9827-y

Flores, J. V., Cordero-Espinoza, L., Oeztuerk-Winder, F., Andersson-Rolf, A., Selmi, T., Blanco, S., et al. (2017). Cytosine-5 RNA methylation regulates neural stem cell differentiation and motility. *Stem Cell Rep.* 8, 112–124. doi: 10.1016/j.stemcr.2016.11.014

Foster, T. C. (2012). Role of estrogen receptor alpha and beta expression and signaling on cognitive function during aging. *Hippocampus* 22, 656–669. doi: 10.1002/hipo.20935

Franke, B., Vermeulen, S. H., Steegers-Theunissen, R. P., Coenen, M. J., Schijvenaars, M. M., Scheffer, H., et al. (2009). An association study of 45 folate-related genes in spina bifida: Involvement of cubilin (CUBN) and tRNA aspartic acid methyltransferase 1 (TRDMT1). *Birth Defects Res. A Clin. Mol. Teratol.* 85, 216–226. doi: 10.1002/bdra.20556

Freude, K., Hoffmann, K., Jensen, L. R., Delatycki, M. B., des Portes, V., Moser, B., et al. (2004). Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine- binding protein cause nonsyndromic X-linked mental retardation. *Am. J. Hum. Genet.* 75, 305–309. doi: 10.1086/422507

Froyen, G., Bauters, M., Boyle, J., Van Esch, H., Govaerts, K., van Bokhoven, H., et al. (2007). Loss of SLC38A5 and FTSJ1 at Xp11.23 in three brothers with non-syndromic mental retardation due to a microdeletion in an unstable genomic region. *Hum. Genet.* 121, 539–547. doi: 10.1007/s00439-007-0343-1

Gamerdinger, M., Manthey, D., and Behl, C. (2006). Oestrogen receptor subtype-specific repression of calpain expression and calpain enzymatic activity in neuronal cells - implications for neuroprotection against Ca^{2+-}mediated excitotoxicity. *J. Neurochem.* 97, 57–68. doi: 10.1111/j.1471-4159.2006.03675.x

Ganot, P., Bortolin, M. L., and Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* 89, 799–809. doi: 10.1016/S0092-8674(00)80263-9

Genenncher, B., Durdevic, Z., Hanna, K., Zinkl, D., Mobin, M. B., Senturk, N., et al. (2018). Mutations in cytosine-5 tRNA methyltransferases impact mobile element expression and genome stability at specific DNA repeats. *Cell Rep.* 22, 1861–1874. doi: 10.1016/j.celrep.2018.01.061

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., et al. (2015). m(6)A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* 347, 1002–1006. doi: 10.1126/science.1261417

Giorda, R., Bonaglia, M. C., Beri, S., Fichera, M., Novara, F., Magini, P., et al. (2009). Complex segmental duplications mediate a recurrent dup(X)(p11.22-p11.23) associated with mental retardation, speech delay, and EEG anomalies in males and females. *Am. J. Hum. Genet.* 85, 394–400. doi: 10.1016/j.ajhg.2009.08.001

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., et al. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398. doi: 10.1126/science.1120976

Gong, P., Li, J., Dai, L., Zhang, K., Zheng, Z., Gao, X., et al. (2008). Genetic variations in FTSJ1 influence cognitive ability in young males in the chinese han population. *J. Neurogenet.* 22, 277–287. doi: 10.1080/01677060802337299

Grosjean (2009). *DNA and RNA Modification Enzymes*. Austin, TX: Landes Bioscience. 289–302.

Gu, T., Zhao, T., Kohli, U., and Hewes, R. S. (2017). The large and small SPEN family proteins stimulate axon outgrowth during neurosecretory cell remodeling in Drosophila. *Dev. Biol.* 431, 226–238. doi: 10.1016/j.ydbio.2017.09.013

Guo, J., Tang, H. W., Li, J., Perrimon, N., and Yan, D. (2018). Xio is a component of the *Drosophila* sex determination pathway and RNA N^6-methyladenosine methyltransferase complex. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3674–3679. doi: 10.1073/pnas.1720945115

Guo, M., and Schimmel, P. (2013). Essential nontranslational functions of tRNA synthetases. *Nat. Chem. Biol.* 9, 145–153. doi: 10.1038/nchembio.1158

Guo, W., Polich, E. D., Su, J., Gao, Y., Christopher, D. M., Allan, A. M., et al. (2015). Fragile X proteins FMRP and FXR2P control synaptic GluA1 expression and neuronal maturation via distinct mechanisms. *Cell Rep.* 11, 1651–1666. doi: 10.1016/j.celrep.2015.05.013

Guy, M. P., and Phizicky, E. M. (2015). Conservation of an intricate circuit for crucial modifications of the tRNAPhe anticodon loop in eukaryotes. *RNA* 21, 61–74. doi: 10.1261/rna.047639.114

Guy, M. P., Podyma, B. M., Preston, M. A., Shaheen, H. H., Krivos, K. L., Limbach, P. A., et al. (2012). Yeast Trm7 interacts with distinct proteins for critical modifications of the tRNAPhe anticodon loop. *RNA* 18, 1921–1933. doi: 10.1261/rna.035287.112

Guy, M. P., Shaw, M., Weiner, C. L., Hobson, L., Stark, Z., Rose, K., et al. (2015). Defects in tRNA Anticodon Loop 2'-O-methylation are implicated in nonsyndromic X-linked intellectual disability due to mutations in FTSJ1. *Hum. Mutat.* 36, 1176–1187. doi: 10.1002/humu.22897

Haag, S., Warda, A. S., Kretschmer, J., Günnigmann, M. A., Höbartner, C., and Bohnsack, M. T. (2015). NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA* 21, 1532–1543. doi: 10.1261/rna.051524.115

Hagerman, R. J., and Polussa, J. (2015). Treatment of the psychiatric problems associated with fragile X syndrome. *Curr. Opin. Psychiatry* 28, 107–112. doi: 10.1097/YCO.0000000000000131

Hamel, B. C., Smits, A. P., van den Helm, B., Smeets, D. F., Knoers, N. V., van Roosmalen, T., et al. (1999). Four families (MRX43, MRX44, MRX45, MRX52) with nonspecific X-linked mental retardation: clinical and psychometric data and results of linkage analysis. *Am. J. Med. Genet.* 85, 290–304. doi: 10.1002/(SICI)1096-8628(19990730)85:3<290::AID-AJMG21>3.0.CO;2-H

Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A., and Stamm, S. (1999). The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Mol. Biol. Cell* 10, 3909–3926. doi: 10.1091/mbc.10.11.3909

Haussmann, I. U., Bodi, Z., Sanchez-Moran, E., Mongan, N. P., Archer, N., Fray, R. G., et al. (2016). m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* 540, 301–304. doi: 10.1038/nature20577

He, C. (2010). Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* 6, 863–865. doi: 10.1038/nchembio.482

Heiss, N. S., Bächner, D., Salowsky, R., Kolb, A., Kioschis, P., and Poustka, A. (2000). Gene structure and expression of the mouse dyskeratosis congenita gene, Dkc1. *Genomics* 67, 153–163. doi: 10.1006/geno.2000.6227

Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., et al. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene encoding a protein with putative nucleolar functions. *Nat. Genet.* 19, 32–38. doi: 10.1038/ng0598-32

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* 34, 721–733. doi: 10.1093/nar/gkj471

Hermann, A., Schmitt, S., and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.* 278, 31717–31721. doi: 10.1074/jbc.M305448200

Hess, M. E., Hess, S., Meyer, K. D., Verhagen, L. A., Koch, L., Brönneke, H. S., et al. (2013). The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nat. Neurosci.* 16, 1042–1048. doi: 10.1038/nn.3449

Higa-Nakamine, S., Suzuki, T., Uechi, T., Chakraborty, A., Nakajima, Y., Nakamura, M., et al. (2012). Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.* 40, 391–398. doi: 10.1093/nar/gkr700

Honda, S., Hayashi, S., Imoto, I., Toyama, J., Okazawa, H., Nakagawa, E., et al. (2010). Copy-number variations on the X chromosome in Japanese patients with mental retardation detected by array-based comparative genomic hybridization analysis. *J. Hum. Genet.* 55, 590–599. doi: 10.1038/jhg.2010.74

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., et al. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* 17, 1265–1272. doi: 10.1016/j.cub.2007.06.030

Hussain, S., Sajini, A. A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., et al. (2013). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* 4, 255–261. doi: 10.1016/j.celrep.2013.06.029

Iijima, T., Wu, K., Witte, H., Hanno-Iijima, Y., Glatter, T., Richard, S., et al. (2011). SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. *Cell* 147, 1601–1614. doi: 10.1016/j.cell.2011.11.028

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi: 10.1016/j.cell.2014.11.035

Jacob-Hirsch, J., Eyal, E., Knisbacher, B. A., Roth, J., Cesarkas, K., Dor, C., et al. (2018). Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. *Cell Res.* 28, 187–203. doi: 10.1038/cr.2018.8

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887. doi: 10.1038/nchembio.687

Jurkowski, T. P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., et al. (2008). Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA* 14, 1663–1670. doi: 10.1261/rna.970408

Kan, L., Grozhik, A. V., Vedanayagam, J., Patil, D. P., Pang, N., Lim, K. S., Huang, Y. C., et al. (2017). The m(6)A pathway facilitates sex determination in Drosophila. *Nat. Commun.* 8:15737. doi: 10.1038/ncomms15737

Kaya, Y., and Ofengand, J. (2003). A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* 9, 711–721. doi: 10.1261/rna.5230603

Keller, L., Xu, W., Wang, H. X., Winblad, B., Fratiglioni, L., and Graff, C. (2011). The obesity related gene, FTO, interacts with APOE, and is associated with Alzheimer's disease risk: a prospective cohort study. *J. Alzheimer's Dis.* 23, 461–469. doi: 10.3233/JAD-2010-101068.

Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., et al. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 90, 856–863. doi: 10.1016/j.ajhg.2012.03.023

Khoddami, V., and Cairns, B. R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31, 458–464. doi: 10.1038/nbt.2566

Kishore, S., and Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311, 230–232. doi: 10.1126/science.1118265

Knuckles, P., Lence, T., Haussmann, I. U., Jacob, D., Kreim, N., Carl, S. H., Masiello, I., et al. (2018). Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d. *Genes Dev.* 32, 415–429. doi: 10.1101/gad.309146.117

Koonin, E. V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res.* 24, 2411–2415. doi: 10.1093/nar/24.12.2411

Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W. W., Morrill, K., et al. (2017). Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLoS Genet.* 13:e1006635. doi: 10.1371/journal.pgen.1006635

Kuchino, Y., Borek, E., Grunberger, D., Mushinski, J. F., and Nishimura, S. (1982). Changes of post-transcriptional modification of wye base in tumor-specific tRNAPhe. *Nucleic Acids Res.* 10, 6421–6432. doi: 10.1093/nar/10.20.6421

Kumar, S., Mapa, K., and Maiti, S. (2014). Understanding the effect of locked nucleic acid and 2'-O-methyl modification on the hybridization thermodynamics of a miRNA-mRNA pair in the presence and absence of AfPiwi protein. *Biochemistry* 53, 1607–1615. doi: 10.1021/bi401677d

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development* 130, 5083–5090. doi: 10.1242/dev.00716

Kurth, H. M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. *RNA* 15, 675–685. doi: 10.1261/rna.1455509

Lacoux, C., Di Marino, D., Boyl, P. P., Zalfa, F., Yan, B., Ciotti, M. T., et al. (2012). BC1-FMRP interaction is modulated by 2'-O-methylation: RNA-binding activity of the tudor domain and translational regulation at synapses. *Nucleic Acids Res.* 40, 4086–4096. doi: 10.1093/nar/gkr1254

Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* 12, 527–537. doi: 10.1101/gad.12.4.527

Lee, S. H., Kim, I., and Chung, B. C. (2007). Increased urinary level of oxidized nucleosides in patients with mild-to-moderate Alzheimer's disease. *Clin. Biochem.* 40, 936–938. doi: 10.1016/j.clinbiochem.2006.11.021

Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M., et al. (2017). Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* 27, 1589–1596. doi: 10.1101/gr.210666.116

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C. H., et al. (2016). m(6)A modulates neuronal functions and sex determination in Drosophila. *Nature* 540, 242–247. doi: 10.1038/nature20568

Lence, T., Soller, M., and Roignant, J. Y. (2017). A fly view on the roles and mechanisms of the m(6)A mRNA modification and its players. *RNA Biol.* 14, 1232–1240. doi: 10.1080/15476286.2017.1307484

Leschziner, G. D., Coffey, A. J., Andrew, T., Gregorio, S. P., Dias-Neto, E., Calafato, M., et al. (2011). Q8IYL2 is a candidate gene for the familial epilepsy syndrome of partial epilepsy with pericentral spikes (PEPS). *Epilepsy Res.* 96, 109–115. doi: 10.1016/j.eplepsyres.2011.05.010

Leygue (2007). Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl. Recept. Signal.* 5:e006. doi: 10.1621/nrs.05006

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.* 15, 1501–1507. doi: 10.1016/j.cub.2005.07.029

Li, L., Zang, L., Zhang, F., Chen, J., Shen, H., Shu, L., et al. (2017a). Fat mass and obesity-associated (FTO) protein regulates adult neurogenesis. *Hum. Mol. Genet.* 26, 2398–2411. doi: 10.1093/hmg/ddx128

Li, T. R., Jia, Y. J., Wang, Q., Shao, X. Q., and Lv, R. J. (2017b). Circular RNA: a new star in neurological diseases. *Int. J. Neurosci.* 127, 726–734. doi: 10.1080/00207454.2016.1236382

Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., et al. (2013). Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat Neurosci* 16, 529–531. doi: 10.1038/nn.3368

Li, X., Zhu, P., Ma, S. , Song, J., Bai, J., Sun, F., et al. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597. doi: 10.1038/nchembio.1836

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., et al. (2016). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.* 12, 311–316. doi: 10.1038/nchembio.2040

Li, X., Xiong, X., Zhang, M., Wang, K., Chen, Y., Zhou, J., et al. (2017c). Base-resolution mapping reveals distinct m(1)A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* 68, 993.e9–1005. e9. doi: 10.1016/j.molcel.2017.10.019

Liang, B., Zhou, J., Kahen, E., Terns, R. M., Terns, M. P., and Li, H. (2009). Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat. Struct. Mol. Biol.* 16, 740–785. doi: 10.1038/nsmb.1624

Lin, M. J., Tang, L. Y., Reddy, M. N., and Shen, C. K. (2005). DNA methyltransferase gene dDnmt2 and longevity of Drosophila. *J. Biol. Chem.* 280, 861–864. doi: 10.1074/jbc.C400477200

Liu, J. Z., Yue, Y. N., Han, D. L., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N-6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95. doi: 10.1038/nchembio.1432

Lovejoy, A. F., Riordan, D. P., and Brown, P. O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS ONE* 9:e110799. doi: 10.1371/journal.pone.0110799

Lubs, H. A., Stevenson, R. E., and Schwartz, C. E. (2012). Fragile X and X-linked intellectual disability: four decades of discovery. *Am. J. Hum. Genet.* 90, 579–590. doi: 10.1016/j.ajhg.2012.02.018

Luo, S., and Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13834–13839. doi: 10.1073/pnas.1412742111

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., et al. (2013). MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.* 41, D262–D267. doi: 10.1093/nar/gks1007

Martinez, F. J., Lee, J. H., Lee, J. E., Blanco, S., Nickerson, E., Gabriel, S., et al. (2012). Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *J. Med. Genet.* 49, 380–385. doi: 10.1136/jmedgenet-2011-100686

Martinez, G., Choudury, S. G., and Slotkin, R. K. (2017). tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res.* 45, 5142–5152. doi: 10.1093/nar/gkx103

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., et al. (2017). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* 541, 371–375. doi: 10.1038/nature21022

Melka, M. G., Gillis, J., Bernard, M., Abrahamowicz, M., Chakravarty, M. M., Leonard, G. T., et al. (2013). FTO, obesity and the adolescent brain. *Hum. Mol. Genet.* 22, 1050–1058. doi: 10.1093/hmg/dds504

Meola, G., and Cardani, R. (2015). Myotonic dystrophies: an update on clinical aspects, genetic, pathology, and molecular pathomechanisms. *Biochim. Biophys. Acta* 1852, 594–606. doi: 10.1016/j.bbadis.2014.05.019

Metodiev, M. D., Spåhr, H., Loguercio Polosa, P., Meharg, C., Becker, C., Altmueller, J., et al. (2014). NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. *PLoS Genet.* 10:e1004110. doi: 10.1371/journal.pgen.1004110

Meyer, K. D., and Jaffrey, S. R. (2017). Rethinking m(6)A readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.* 33, 319–342. doi: 10.1146/annurev-cellbio-100616-060758

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. doi: 10.1016/j.cell.2012.05.003

Minnerop, M., Weber, B., Schoene-Bake, J. C., Roeske, S., Mirbach, S., Anspach, C., et al. (2011). The brain in myotonic dystrophy 1 and 2: evidence for a predominant white matter disease. *Brain* 134, 3527–3543. doi: 10.1093/brain/awr299

Mitchell, J. R., Wood, E., and Collins, K. (1999). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402, 551–555. doi: 10.1038/990141

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P. J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10756–10761. doi: 10.1073/pnas.0402560101

Motorin, Y., and Grosjean, H. (1999). Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast *Saccharomyces cerevisiae*: identification of the gene and substrate specificity of the enzyme. *RNA* 5, 1105–1118. doi: 10.1017/S1355838299982201

Motorin, Y., and Helm, M. (2010). tRNA stabilization by modified nucleotides. *Biochemistry* 49, 4934–4944. doi: 10.1021/bi100408z

Motorin, Y., and Helm, M.(2011). RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* 2, 611–631. doi: 10.1002/wrna.79

Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 26, 2536–2540. doi: 10.1093/nar/26.11.2536

Oldmeadow, C., Mossman, D., Evans, T.-J., Holliday, E. G., Tooney, P. A., Cairns, M. J., et al. (2014). Combined analysis of exon splicing and genome wide polymorphism data predict schizophrenia risk loci. *J. Psychiatr. Res.* 52, 44–49. doi: 10.1016/j.jpsychires.2014.01.011

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Res*. doi: 10.1038/s41422-018-0013-y. [Epub ahead of print].

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537, 369–373. doi: 10.1038/nature19342

Pergolizzi, R. G., and Grunberger, D. (1980). The effect of exogenous nutrients on the biosynthesis of Y base in tRNAPHe from Ehrlich ascites carcinoma. *Cancer Lett.* 8, 329–333. doi: 10.1016/0304-3835(80)90149-4

Perrat, P. N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., et al. (2013). Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* 340, 91–95. doi: 10.1126/science.1231965

Peters, J. (2008). Prader-willi and snoRNAs. *Nat. Genet.* 40, 688–689. doi: 10.1038/ng0608-688

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.* 41, 696–702. doi: 10.1038/ng.360

Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* 24, 177–189. doi: 10.1038/cr.2014.3

Pintard, L., Lecointe, F., Bujnicki, J. M., Bonnerot, C., Grosjean, H., and Lapeyre, B. (2002). Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. *EMBO J.* 21, 1811–1820. doi: 10.1093/emboj/21.7.1811

Raddatz, G., Guzzardo, P. M., Olova, N., Fantappié, M. R., Rampp, M., Schaefer, M., et al. (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8627–8631. doi: 10.1073/pnas.1306723110

Rai, K., Chidester, S., Zavala, C. V., Manos, E. J., James, S. R., Karpf, A. R., et al. (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.* 21, 261–266. doi: 10.1101/gad.1472907

Ramser, J., Winnepenninckx, B., Lenski, C., Errijgers, V., Platzer, M., Schwartz, C. E., et al. (2004). A splice site mutation in the methyltransferase gene FTSJ1 in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). *J. Med. Genet.* 41, 679–683. doi: 10.1136/jmg.2004.019000

Rebane, A., Roomere, H., and Metspalu, A. (2002). Locations of several novel 2'-O-methylated nucleotides in human 28S rRNA. *BMC Mol. Biol.* 3:1. doi: 10.1186/1471-2199-3-1

Reitz, C., Tosto, G., Mayeux, R., Luchsinger, J. A., Group, N.-L. N. F. S., and Alzheimer's Disease Neuroimaging, I. (2012). Genetic variants in the Fat and Obesity Associated (FTO) gene and risk of Alzheimer's disease. *PLoS ONE* 7:e50354. doi: 10.1371/journal.pone.0050354

Rintala-Dempsey, A. C., and Kothe, U. (2017). Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.* 14, 1185–1196. doi: 10.1080/15476286.2016.1276150

Roignant, J. Y., and Soller, M. (2017). m(6)A in mRNA: an ancient mechanism for fine-tuning gene expression. *Trends Genet.* 33, 380–390. doi: 10.1016/j.tig.2017.04.003

Ross, R. J., Weiner, M. M., and Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* 505, 353–359. doi: 10.1038/nature12987

Roth, A., Schaffner, W., and Hertel, C. (1999). Phytoestrogen kaempferol (3,4',5,7-tetrahydroxyflavone) protects PC12 and T47D cells from beta-amyloid-induced toxicity. *J. Neurosci. Res.* 57, 399–404. doi: 10.1002/(SICI)1097-4547(19990801)57:3<399::AID-JNR12>3.0.CO;2-W

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi: 10.1016/j.cell.2017.05.045

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., et al. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551, 251–255. doi: 10.1038/nature24456

Sahoo, T., del Gaudio, D., German, J. R., Shinawi, M., Peters, S. U., Person, R. E., et al. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85C/D box small nucleolar RNA cluster. *Nat. Genet.* 40, 719–721. doi: 10.1038/ng.158

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M. C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2 '-O-methylation of PIWI-interacting RNAs at their 3 ' ends. *Genes Dev.* 21, 1603–1608. doi: 10.1101/gad.1563607

Schaefer, M., and Lyko, F. (2010a). Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation. *Nat. Genet.* 42, 920–921. doi: 10.1038/ng1110-920

Schaefer, M., and Lyko, F. (2010b). Solving the Dnmt2 enigma. *Chromosoma* 119, 35–40. doi: 10.1007/s00412-009-0240-6

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., et al. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 24, 1590–1595. doi: 10.1101/gad.586710

Schöller, E., Weichmann, F., Treiber, T., Ringle, S., Treiber, N., Flatley, A., et al. (2018). Interactions, localization and phosphorylation of the m6A generating METTL3-METTL14-WTAP complex. *RNA* 24, 499–512. doi: 10.1261/rna.064063.117

Schorn, A. J., Gutbrod, M. J., LeBlanc, C., and Martienssen, R. (2017). LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* 170, 61.e11–71.e11. doi: 10.1016/j.cell.2017.06.013

Schosserer, M., Minois, N., Angerer, T. B., Amring, M., Dellago, H., Harreither, E., et al. (2015). Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan. *Nat. Commun.* 6, 6158. doi: 10.1038/ncomms7158

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., et al. (2014a). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162. doi: 10.1016/j.cell.2014.08.028

Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., et al. (2014b). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* 8, 284–296. doi: 10.1016/j.celrep.2014.05.048

Shaheen, R., Han, L., Faqeih, E., Ewida, N., Alobeid, E., Phizicky, E. M., et al. (2016). A homozygous truncating mutation in PUS3 expands the role of tRNA modification in normal cognition. *Hum. Genet.* 135, 707–713. doi: 10.1007/s00439-016-1665-7

Shao, Y., and Chen, Y. (2016). Roles of circular RNAs in neurologic disease. *Front. Mol. Neurosci.* 9:25. doi: 10.3389/fnmol.2016.00025

Sharma, S., Yang, J., Watzinger, P., Kötter, P., and Entian, K. D. (2013). Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic Acids Res.* 41, 9062–9076. doi: 10.1093/nar/gkt679

Sibley, C. R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., et al. (2015). Recursive splicing in long vertebrate genes. *Nature* 521, 371–375. doi: 10.1038/nature14466

Sledz, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife* 5:e18434. doi: 10.7554/eLife.18434

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Somme, J., Van Laer, B., Roovers, M., Steyaert, J., Versées, W., and Droogmans, L. (2014). Characterization of two homologous 2'-O-methyltransferases

showing different specificities for their tRNA substrates. *RNA* 20, 1257–1271. doi: 10.1261/rna.044503.114

Sridhar, P., Gan, H. H., and Schlick, T. (2008). A computational screen for C/D box snoRNAs in the human genomic region associated with Prader-Willi and Angelman syndromes. *J. Biomed. Sci.* 15, 697–705. doi: 10.1007/s11373-008-9271-x

Takano, K., Nakagawa, E., Inoue, K., Kamada, F., Kure, S., Goto, Y., et al. (2008). A loss-of-function mutation in the FTSJ1 gene causes nonsyndromic X-linked mental retardation in a Japanese family. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 479–484. doi: 10.1002/ajmg.b.30638

Terrazas, M., and Kool, E. T. (2009). RNA major groove modifications improve siRNA stability and biological activity. *Nucleic Acids Res.* 37, 346–353. doi: 10.1093/nar/gkn958

Theler, D., Dominguez, C., Blatter, M., Boudet, J., and Allain, F. H. (2014). Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res.* 42, 13911–13919. doi: 10.1093/nar/gku1116

Torres, A. G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. *Trends Mol. Med.* 20, 306–314. doi: 10.1016/j.molmed.2014.01.008

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., et al. (2012). RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* 19, 900–905. doi: 10.1038/nsmb.2357

Uttara, B., Singh, A. V., Zamboni, P., and Mahajan, R. T. (2009). Oxidative stress and neurodegenerative diseases: a review of upstream and downstream antioxidant therapeutic options. *Curr. Neuropharmacol.* 7, 65–74. doi: 10.2174/157015909787602823

van Rossum, D., Verheijen, B. M., and Pasterkamp, R. J. (2016). Circular RNAs: novel regulators of neuronal development. *Front. Mol. Neurosci.* 9:74. doi: 10.3389/fnmol.2016.00074

Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaillé, J., et al. (2005). ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.* 169, 745–753. doi: 10.1083/jcb.200411129

Wang, C. C., Lo, J. C., Chien, C. T., and Huang, M. L. (2011). Spatially controlled expression of the Drosophila pseudouridine synthase RluA-1. *Int. J. Dev. Biol.* 55, 223–227. doi: 10.1387/ijdb.103112cw

Wang, P., Doxtader, K. A., and Nam, Y. (2016a). Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell* 63, 306–317. doi: 10.1016/j.molcel.2016.05.041

Wang, T., Bray, S. M., and Warren, S. T. (2012). New perspectives on the biology of fragile X syndrome. *Curr. Opin. Genet. Dev.* 22, 256–263. doi: 10.1016/j.gde.2012.02.002

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., et al. (2016b). Structural basis of N-6-adenosine methylation by the METTL3-METTL14 complex. *Nature* 534, 575–578. doi: 10.1038/nature18298

Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z. L., and Zhao, J. C. (2014). N-6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16, 191–198. doi: 10.1038/ncb2902

Wang, Y., Li, Y., Yue, M., Wang, J., Kumar, S., Wechsler-Reya, R. J., et al. (2018). N(6)-methyladenosine RNA modification regulates embryonic neural stem cell self-renewal through histone modifications. *Nat. Neurosci.* 21, 195–206. doi: 10.1038/s41593-017-0057-1

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., et al. (2018). Zc3h13 regulates nuclear RNA m(6)A methylation and mouse embryonic stem cell self-renewal. *Mol. Cell* 69, 1028.e6–1038.e6. doi: 10.1016/j.molcel.2018.02.015

Weng, Y. L., Wang, X., An, R., Cassin, J., Vissers, C., Liu, Y., et al. (2018). Epitranscriptomic m(6)A regulation of axon regeneration in the adult mammalian nervous system. *Neuron* 97, 313.e6–325.e6. doi: 10.1016/j.neuron.2017.12.036

Widagdo, J., Zhao, Q. Y., Kempen, M. J., Tan, M. C., Ratnu, V. S., Wei, W., et al. (2016). Experience-dependent accumulation of N6-methyladenosine in the prefrontal cortex is associated with memory processes in mice. *J. Neurosci. Off. J. Soc. Neurosci.* 36, 6771–6777. doi: 10.1523/JNEUROSCI.4053-15.2016

Wilkinson, C. R., Bartlett, R., Nurse, P., and Bird, A. P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. *Nucleic Acids Res.* 23, 203–210. doi: 10.1093/nar/23.2.203

Willems, P., Vits, L., Buntinx, I., Raeymaekers, P., Van Broeckhoven, C., and Ceulemans, B. (1993). Localization of a gene responsible for nonspecific mental retardation (MRX9) to the pericentromeric region of the X chromosome. *Genomics* 18, 290–294. doi: 10.1006/geno.1993.1468

Wilusz, J. E. (2015). Controlling translation via modulation of tRNA levels. *Wiley Interdiscipl. Rev. RNA* 6, 453–470. doi: 10.1002/wrna.1287

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., et al. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10, 927–929. doi: 10.1038/nchembio.1654

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res.* 27, 626–641. doi: 10.1038/cr.2017.31

Yao, P., and Fox, P. L. (2013). Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol. Med.* 5, 332–343. doi: 10.1002/emmm.201100626

Yildirim, I., Kierzek, E., Kierzek, R., and Schatz, G. C. (2014). Interplay of LNA and 2'-O-methyl RNA in the structure and thermodynamics of RNA hybrid systems: a molecular dynamics study using the revised AMBER force field and comparison with experimental results. *J. Phys. Chem. B* 118, 14177–14187. doi: 10.1021/jp506703g

Yoon, K. J., Ringeling, F. R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., et al. (2017). Temporal control of mammalian cortical neurogenesis by m(6)A methylation. *Cell* 171, 877–899. doi: 10.1016/j.cell.2017.09.003

Yu, A. T., Ge, J., and Yu, Y. T. (2011). Pseudouridines in spliceosomal snRNAs. *Protein Cell* 2, 712–725. doi: 10.1007/s13238-011-1087-1

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* 307, 932–935. doi: 10.1126/science.1107130

Yu, J., Chen, M., Huang, H., Zhu, J., Song, H., Zhu, J., et al. (2018). Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Res.* 46, 1412–1423. doi: 10.1093/nar/gkx1182

Yue, Y., Liu, J., 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. doi: 10.1038/s41421-018-0019-0

Zahn, L. M. (2017). ERVs affect brain gene expression. *Science* 355, 491–492. doi: 10.1126/science.355.6324.491-d

Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di Penta, A., Reis, S., et al. (2003). The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112, 317–327. doi: 10.1016/S0092-8674(03)00079-5

Zhang, Y., Shi, J., and Chen, Q. (2017). tsRNAs: new players in mammalian retrotransposon control. *Cell Res.* 27, 1307–1308. doi: 10.1038/cr.2017.109

Zhao, X., Patton, J. R., Davis, S. L., Florence, B., Ames, S. J., and Spanjaard, R. A. (2004). Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* 15, 549–558. doi: 10.1016/j.molcel.2004.06.044

Zhao, Y., Mo, B., and Chen, X. (2012). Mechanisms that impact microRNA stability in plants. *RNA Biol.* 9, 1218–1223. doi: 10.4161/rna.22034

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015

Zhou, C., Molinie, B., Daneshvar, K., Pondick, J. V., Wang, J., Van Wittenberghe, N., et al. (2017). Genome-wide maps of m6A circRNAs Identify Widespread and cell-type-specific methylation patterns that are distinct from mRNAs. *Cell Rep.* 20, 2262–2276. doi: 10.1016/j.celrep.2017.08.027

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter III: Aims of the PhD project

1. A genomic screen unravels a novel tRNA methylase involved in the sncRNA pathways

My team is interested in the molecular mechanisms and the connections between the small non-coding RNA pathways. In the purpose of identifying new actors involved in the miRNA pathway, a self-silencing gene reporter (automiG) was constructed and characterized (Carre et al., 2013). Under the control of a single copper-inducible promoter, the system expresses the GFP protein as well as two reprogramed miRNAs, miG-1 and miG-2, forming perfect dsRNA duplexes and perfectly matched to 2 distinct sites in the GFP coding sequence for maximizing GFP silencing. For the construction of the automiG reporter the first coding nucleotides of exon-2, the second intron and the first nucleotides of exon-3 from the *Drosophila RpL17* gene was fused to the coding sequence of the GFP. Then, the sequences of the pre-miRNAs of the two synthetic miRNAs were inserted within the RpL17 intron. The construct relies on the property of Ago-2 to load miRNA duplexes having perfect complementarity Figure 11.

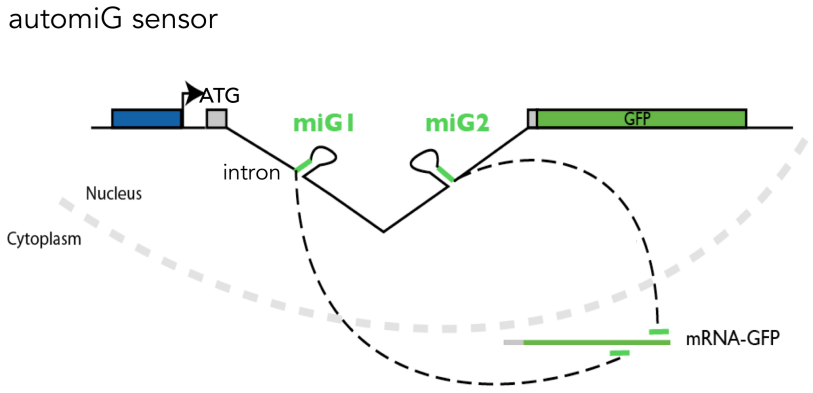

Figure 11. Schematic representation of the automiG sensor.

My team characterized that the self-silencing of the system does not involve Ago-1, the mediator of the miRNA pathway. They found that the self-silencing of the automiG gene requires Drosha, Pasha, Dcr-1, Dcr-2 (for its loading activity) and Ago-2 loaded with the anti-GFP miRNAs miG1 and miG2. Thus, the resulting sensor is specific in the same time for miRNA biogenesis and for Ago-2-mediated silencing and can be used to report *in vivo* for deregulations in both of these mechanisms (Carre et al., 2013).

automiG was then used as a biosensor to screen a chemical library in the search of compounds capable to inhibit miRNA silencing (Carre et al., 2013) as well as more recently in a genomic screen (unpublished),

aiming to identify new actors involved in the miRNA pathway or in Ago-2-mediated silencing. Seventeen new genes, uncharacterized before as involved in the miRNA- or Ago-2-mediated silencing mechanisms, scored positively in the screen. *CG7009*, the central subject of my PhD research, was one of these genes.

2. CG7009 is a novel actor of the Ago-2-dependent miRNA pathway

In the beginning of my project *CG7009* was known as encoding a protein, whose amino acid sequence presented high homology with two known tRNA 2'-O methyltransferases: Trm7 in yeast and its human ortholog FTSJ1. We thus wondered if we discovered a 2'-O tRNA methylase implicated in the miRNA-Ago-2 pathway. In regards to the novel role of RNA modifying enzymes and tRNA fragments in the regulation of the sncRNA pathways this gene was interesting to us. Furthermore, the *CG7009* gene was predicted to catalyze 2'-O methylation and, as described before, the sncRNAs loaded in Ago-2 are 2'-O-methylated at their 3'-terminal nucleotides. As discussed, this modification is catalyzed in *Drosophila* by DmHen-1/ Pimet. Surprisingly, in flies, mutant for *pimet* are viable and fertile and do not present typical sncRNA phenotypes, contrary to what was expected if the sncRNAs are destabilized. We hypothesized that maybe we had uncovered a redundant protein of DmHen-1, which could compensate Hen-1 loss-of-function and thus explain the rather mild phenotypes of DmHen-1 mutant flies.

During my Master 2 we favored two hypotheses: (i) *CG7009* encodes an tRNA 2'-O-methyltransferase, which prevents tRNA fragmentation: in the absence of CG7009 an abnormal amount of tRFs are produced, the resulting tRFs are loaded into key proteins involved in the sncRNAs pathways, *i.e*. Dcr-2 or Ago-2 proteins and titrate them. This tRFs loading inhibits the sncRNAs biogenesis and/or silencing activity, leading to sncRNA pathways' failure. This hypothesis evokes an indirect modulation of the sncRNA silencing pathways by CG7009, similarly to Dnmt2 function in the siRNA pathway (Durdevic et al., 2013b). (ii) *CG7009* encodes an tRNA 2'-O-methyltransferase which can also methylate the sncRNAs and CG7009 absence more directly affects the stability or the silencing activity of the sncRNAs.

Was it possible that *CG7009* encodes an tRNA-methyltransferase catalyzing the methylation of both tRNAs and sncRNAs and/ or their precursors in order to maintain their stability and protect them against endo- or exonucleases?

RESULTS and DISCUSSION

Chapter I: The *CG7009* and *CG5220* genes – tools and characterization of mutants

1. Amino acid alignment and conservation

We discovered *CG7009* in a genomic screen for factors involved in the Ago-2 miRNA pathway in flies. Until our research, this gene had no known function. Preliminary data, received before my arrival in the laboratory implicated this gene in the miRNA-Argonaute 2-mediated mechanism of RNA silencing in *Drosophila*. Amino Acid (AA) sequence analysis indicates that the polypeptide encoded by *CG7009* possesses a methyltransferase domain belonging to a conserved methyltransferase subfamily (TRM7 family). Aligning the sequence of the CG7009 protein with the flies' reference proteome leads to another protein with unknown function, the CG5220 gene product, which has a 63 % AA sequence identity with the CG7009 protein (Figure 12). We were later able to also implicate CG5220 in the regulation of the sncRNA pathways of the fly. Moreover, according to *FlyBase*, the richest public online database for genetic and molecular data on *Drosophila melanogaster*, both CG7009 and CG5220 have the same predicted orthologs in yeast and in human.

The putative yeast ortholog of CG7009 and of CG5220 is Trm7. Alignment of CG7009 with Trm7 shows 52 % AA sequence identity, with nearly perfect homology in the methyltransferase domain and 66 % of coverage (Figure 12). Similarly, alignment of CG5220 with Trm7 shows high scores – 48 % identity and 83 % coverage (Figure 12). In yeast *S. cerevisiae* Trm7 catalyzes Nm deposition at positions 32 and 34 in the anti-codon (AC) loop $tRNA^{Phe}$, $tRNA^{Trp}$ and $tRNA^{LeuTAA}$ (Pintard et al., 2002). Trm7 uses S-adenosyl methionine (SAM) as donor of methyl groups for the methylation of its RNA substrates and has a SAM-binding domain. Lack of Trm7 function in the AC loop on its substrate tRNAs has been linked to growth and translational defects (Pintard et al., 2002).

FTSJ1 is the human ortholog of Trm7 (Guy et al., 2015) and thus the potential ortholog of CG7009 and CG5220. CG7009 shares 51 % of AA identity and 86 % coverage with FTSJ1, CG5220 shares 58 % of AA identity and 82 % coverage (Figure 12). Another interesting fact is the implication of FTSJ1 in human pathology. FTSJ1 lack of function is associated with the development of non-syndromic X-linked intellectual disability (NSXLID) (Freude et al., 2004), a common neurodegenerative disease. FTSJ1 is also

reported as differentially expressed in cancer in public databases (http://www.proteinatlas.org/ENSG00000068438-FTSJ1/cancer).

The existence of these two potential paralogs in *Drosophila* complicated the analysis of their intrinsic function(s) as it was possible that they have redundant functions. In order to study the function(s) of these two proteins, we first needed to obtain or engineer proper tools, allowing their molecular characterization.

In this first part of the results, obtained during my PhD, I will describe the characterized and/or obtained *CG7009* and *CG5220* mutants and tools that allowed us to preform further research on the two genes. I will also characterize part of the phenotypes, associated with CG7009 loss-of-function, that we uncovered during the mutants' characterization.

Figure 12. Conservation between CG7009, CG5220, Trm7 And FTSJ1. A. Multiple sequence alignment of the amino acid sequences of CG7009, CG5220, Trm7, and FTSJ1 (www.ebi.ac.uk/Tools/msa/kalign/) visualized in Unipro UGENE 1.26. Red squares indicate the predicted catalytic tetrad K-D-K-E (marked with *). **B.** Amino acid identity and coverage between CG7009, CG5220, Trm7, and FTSJ1 (www.ensembl.org). Ident. – identity; Cov. – coverage

2. The *CG7009* gene

According to *FlyBase*, *CG7009* is located on the right arm of the 3rd chromosome (3R) and its genomic sequence location is 3R:16958456...16959603[-] (cytological map location 93C1). According to the report in FlyBase, *CG7009* has one annotated transcript and encodes for two reported polypeptides: one of 320 amino acids (35.2 kDa, UniProt, Q9VDD9), and one of 324 amino acids (35.64 kDa, UniProt, Q8MRW7). Analysis of sequences homology suggests that the product of *CG7009* belongs to the conserved RlmE family and Trm7 subfamily of the methyltransferase superfamily. In *FlyBase* no experimental data are available on *CG7009* mutant phenotype(s), 2 mutant stocks are reported, associated with "viable and fertile" phenotypic classes. Our experimental data agrees with this statement, but we do however find that homozygous and trans-heterozygous mutants of *CG7009* have several interesting phenotypes, *i.e.* the aforementioned involvement in the Ago-2-dependent miRNA pathway, as well as other phenotypes that we discovered during my PhD and that I will present in the following sections.

3. Molecular characterization of the *CG7009* mutation

Strain *CG7009*^{e02001}, mutant for *CG7009*, originates from *The Drosophila Gene Disruption Project* (GDP, http://flypush.imgen.bcm.tmc.edu/pscreen/about.html) mutant strains collection. The collection comprises mutant strains that have single transposon insertions associated with different genes, which were created by *P-* and *piggyBac* transposable element-mediated mutagenesis. In order to map the insertions in the fly genome, small amounts of gDNA sequences from the insertion region have been recovered by inverse PCR (iPCR), sequenced and mapped to the reference genome. Thus, for the majority of the mutant strains generated by the GDP, including the *CG7009*^{e02001} mutant strain, the site of transposon insertion and its flanking genomic sequences were already determined.

The BL18008 strain (Bloomington stock) bears the mutant allele *CG7009*^{e02001}, generated by the insertion of a PBac{RB} transposon in the middle of the CDS (Figure 13A). Confirmation of the reported insertion was performed by PCR, followed by sequencing. gDNA from homo- and heterozygous *CG7009*^{e02001} flies was amplified by PCR with CG7009-dTOPO FW and CG7009-dTOPO Rev primers (Methods) that hybridize respectively at the 5'- and 3'-end of the *CG7009* gene (Figure 13A). After electrophoretic separation of the PCR products, two bands were detected for the heterozygous mutant migrating at approximately 7kb and 1.1kb which correspond to the size of the PBac{RB} insertion allele

CG7009^{e02001} and the wild type alleles of the gene, respectively. Only the band of higher molecular weight was detected in homozygous flies (Figure 13B lane -/-). This confirms that the BL18008 strain carry an allele of the gene *CG7009* that has an inserted element of around 6kb, consistent with the reported size of the PBac{RB} transposon.

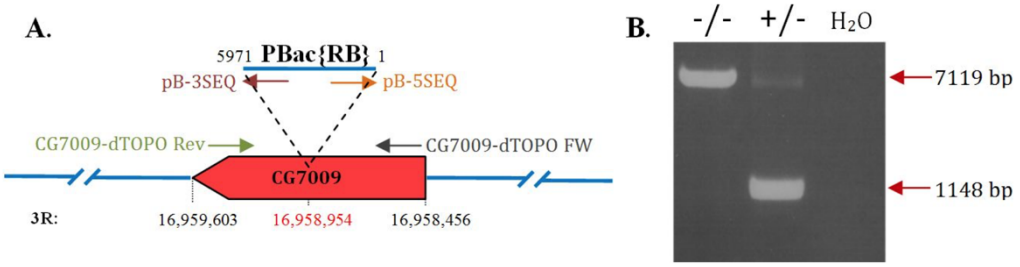

Figure 13. Molecular characterization of the *CG7009*^{e02001} mutation. A. Schematic representation of the *CG7009* gene showing the genomic location of the insertion and the size of the PBac{RB} transgenic transposon. The hybridizing four oligonucleotides used in the PCR analysis are represented with arrows designating their annealing locations and orientation. **B.** Electrophoretic separation of the PCR reaction made on gDNA of flies homozygous (-/-) or heterozygous (+/-) for the mutant allele *CG7009*^{e02001} or H_2O (negative control, no DNA) using CG7009-dTOPO FW and CG7009-dTOPO Rev primers.

In order to precisely determine the insertion site, gDNA of homozygous mutants *CG7009*^{e02001} / *CG7009*^{e02001} was sequenced with divergent primers, aligning either at the 5'-end (pB-5SEQ) or at the 3'-end (pB-3Seq) of the piggyBac transposon (Figure 13A and Methods). Thus, the sequences obtained from this reaction started inside the piggyBac and extended towards the *CG7009* gene sequence. This experiment allowed me to precisely map the insertion site at genomic location 3R:16,958,954, which is consistent with the position reported in *FlyBase* (Figure 13A). Additionally, by re-sequencing the *CG7009* gDNA region in our strain, I was able to verify that the entire coding sequence of *CG7009* is identical to the sequence of the gene in the reference genome, whose translation gives the protein of 32 kDa from the two reported polypeptides for this gene (data not shown).

The reported insertion site of the piggyBac in the middle of the CDS strongly suggests that the *CG7009*^{e02001} allele is null mutant of *CG7009*. In order to test this, a reverse transcription PCR (RT-PCR) on total RNA ovaries extract, using internal primers to detect the mRNA of *CG7009*, was performed in heterozygous or homozygous mutants. This experiment clearly showed that in contrast to the heterozygous mutant, no *CG7009* mRNA is detected in the *CG7009*^{e02001} homozygous mutant ovaries

(Figure 14). These RT-PCR results strongly suggest that the *CG7009*^{e02001} allele is a transcriptional null mutant of *CG7009*.

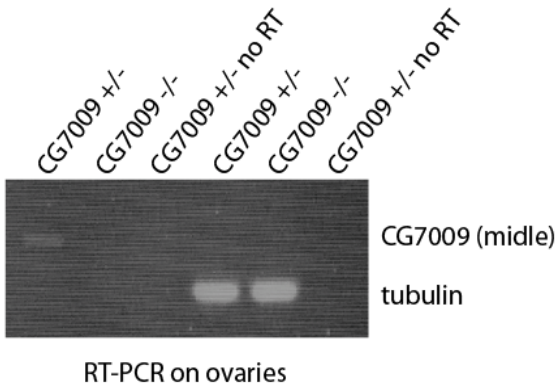

Figure 14. Reverse Transcription PCR (RT-PCR) on total RNA extracts of ovaries. Electrophoretic separation of the RT-PCR reaction made from total RNA of flies homozygous (-/-) or heterozygous (+/-) for the mutant allele *CG7009*^{e02001} or no RT (negative control, no reverse transcriptase in the RT reaction) using CG7009-middle Rev and CG7009-dTOPO FW primers for the PCR reaction (expected size 500nt) and tubulin RT-qPCR primers (expected size 150nt).

4. Construction of a *Drosophila* strain carrying an additional WT *CG7009* locus

To be able to perform genetic rescue experiments on the loss of CG7009 function phenotypes (see below) we wanted to obtain a *Drosophila* strain homozygous for the *CG7009*^{e02001} allele and carrying a wild type allele inserted in an ectopic genomic location. Our goal was to express the wild type gene under the control of its endogenous regulatory elements in a *CG7009* mutant background and observe if, in this genomic context, we will be able to rescue the phenotypes observed in the homozygous mutant. We took advantage of the P[acman] transgenesis platform for *Drosophila* (Venken et al., 2009; Venken et al., 2006). AttB-P[acman] is an enhanced system derived from bacteriophage **Φ**C31 attP/attB-mediated specific recombination. This system adapted to *Drosophila* allows site-specific insertional transformation with large DNA fragments. AttP/ attB recombination in the presence of **Φ**C31 recombinase mediates the insertion of the desired BAC at a specific attP-landing site engineered in the fly's genome. This system overcomes size limitations posed by the classically used P-transposase-mediated transformation (Venken et al., 2006). We used a P[acman] clone (*CH322-117K12*) from a *Drosophila* bacterial artificial chromosome genomic library which encompasses 21948bp of the chromosome arm 3R (from 16954560 to 16976507), including the *CG7009* gene as well as 16960 nt of the chromosomal sequence upstream of its initial ATG and 3915 nt downstream of its stop codon (Figure 15). We reasoned that a larger part of the genomic sequence upstream of the gene's coding sequence as well as a smaller, but still important, part of the sequence downstream of its coding sequence increases the possibility to retrieve if not all, most of the endogenous

cis-regulatory elements of the *CG7009* gene. The construction of this transgenic line was a prerequisite for concluding on the function of our candidate gene in rescue experiments.

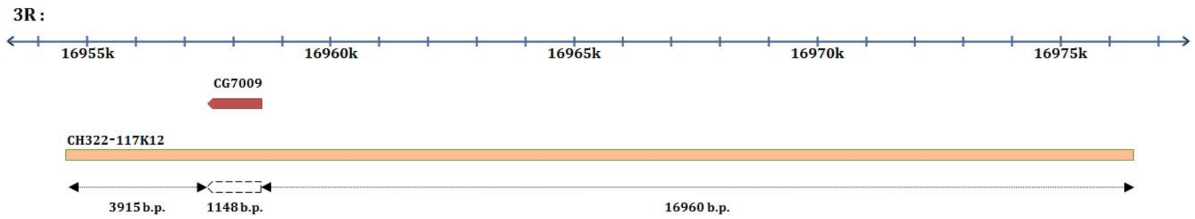

Figure 15. Schematic representation of the BAC clone CH322-117K12. Shown are the mapping to the genome (blue double arrow) of the P[acman] clone CH322-117K12 depicted in beige. *CG7009* gene loci is depicted in red.

A. Verification of the BAC clone

The full-length sequence of the BAC has not been confirmed when the BAC library containing clone CH322-117K12 was constructed by the P[acman] consortium (*http://www.pacmanfly.org/credits.html*). Only the terminal sequences of the clone have been aligned to the reference genome. Thus, the presence of the full sequence of *CG7009* in the CH322-117K12 clone had to be validated. I first verified if clone CH322-117K12 contains the right genomic sequence of *CG7009* using Sanger sequencing. In order to obtain a coverage encompassing the entire CDS of *CG7009* a combination of four oligonucleotides was used. Two of them, which I already presented above are hybridizing at the two extremities of the *CG7009* gene*:* CG7009-dTOPO FW and CG7009-dTOPO Rev (Methods) and a pair of two other divergent oligonucleotides, complementary to sequences in the middle of *CG7009*: CG7009-middle FW and CG7009-middle Rev (Methods). This strategy allowed me to confirm that the entire coding sequence of *CG7009* is indeed contained in the BAC clone. Importantly, the sequencing results (data not shown) confirmed that the CDS of the gene in the reference genome published in *FlyBase* and in the BAC clone are identical.

B. Mating scheme for introducing the BAC transgene in a *CG7009* Mutant Background

An important feature of *Drosophila* genetics is the absence of recombination in males. Recombination in female flies can be controlled in laboratory conditions by one of the most distinguishable tools in fly genetics, the balancer chromosomes. These chromosomes result from induced irradiation inversions. The resulting scrambling/ inversion of the balancer chromosomes sequences leads to a loss of their ability to recombine with their normal homolog during meiosis. In addition, the inheritance or not of balancer chromosomes can easily be followed as they carry dominant marker mutation. Thus, their transmission to the progeny can be traced. Since they block recombination, their un-recombined normal homolog (without dominant marker) can also be tracked.

This illustrates one of the most important principles in fly mating schemes, which is that if the progeny did not get the balancer it must have gotten the homolog unbalanced chromosome. This incapability for recombination proves to be very useful equally for stocking lethal mutant strains. Since *CG7009* is located in chromosome III, we chose an attB-docking site located on chromosome II to simplify future genetic crosses during the rescue experiment. Figure 16A represents the scheme of the realized crosses for the establishment of a *CG7009* rescue line. A heterozygous male from the BAC line (received by the company who made the transgenesis for us (*BestGene®*) was crossed with a double balanced line called CC2. The CC2 line is a double balanced homemade line. The T(2;3)ap[Xa] translocation forces the segregation of CyO-GFP(w-) and TM6C balancers always together. Using this capacity and by following the appropriate dominant mutations present in the balanced chromosomes, I was able to build the following line: w-; BAC; *CG7009^{e02001}/ TM6, Tb, Sb.*

The final crosses represented in Figure 16B (green and blue boxes) allowed to realize the rescue experiment by obtaining a strain homozygous for the *CG7009^{e02001}* mutant allele or trans-heterozygote for a deficiency of the *CG7009* genomic region that carry in the same time the wild type allele of *CG7009* inserted in the second chromosome. This allowed me to observe if the BAC transgene is able to rescue the phenotypes observed in the homozygous CG7009^{e02001} mutant, *i.e.* its aforementioned involvement in the Ago-2-miRNA pathway, providing the final proof that the phenotypes we observe are actually due to the loss-of-function of the *CG7009* gene.

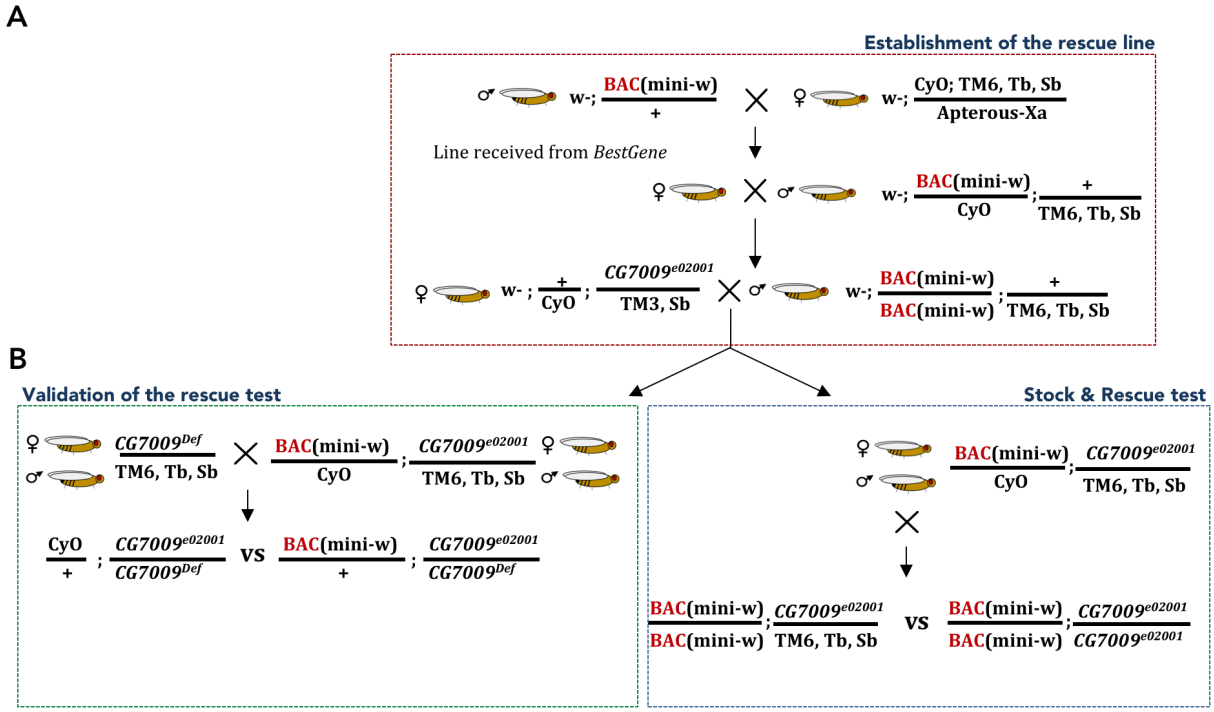

Figure 16. Mating scheme for the construction of the transgenic rescue line. A. Schematic representation of the genetic crosses performed for the establishment of the line w-; BAC; *CG7009*^{e02001}/ TM6, Tb, Sb **B.** Illustration of the genetic rescue experiment. *CG7009*^{e02001}/ *CG7009*^{Def} corresponds to trans-heterozygous *CG7009* mutant, containing one insertional mutant allele and one deficiency. BAC, Bacterial Artificial Chromosome containing the *CG7009* genomic region; w-, mutant *white*; mini-w, mini-*white* marker; Apterous-Xa, CyO, TM6, Tb, Sb and TM3: balancer chromosomes.

C. Validation of the expression of *CG7009* mRNA from the BAC by RT-qPCR

I then analyzed by RT-qPCR the expression of *CG7009* in females (Figure 17A), as well as in the ovaries (Figure 17B) of several fly strains used in our research. For both whole females and ovaries heterozygotes carrying the allele *CG7009*^{e02001} served as control (Figure 17A and B, black bars). In whole females I tested the three available mutants: homozygous (*CG7009*^{e02001} / *CG7009*^{e02001}) and trans-heterozygous (*CG7009*^{e02001} / Def3340 and *CG7009*^{e02001} / Def9487) *CG7009* mutant flies (Figure 17A, red bars). The expression of *CG7009* transcript in the ovaries was assayed only in trans-heterozygous mutants *CG7009*^{e02001} / Def9487 (Figure 17B, red bars). I also analyzed CG7009 transcript expression in the rescue line BAC / CyO ; *CG7009*^{e02001} / Def9487 in both females and ovaries (Figure 17A and B, green graphs). This experiment allowed me to confirm that the three mutants are transcriptional null mutants for *CG7009*. It also allowed me to validate that *CG7009* expression levels in the rescue line is comparable to the gene expression level in heterozygous individuals in females (Figure 17A) and in ovaries (Figure 17B).

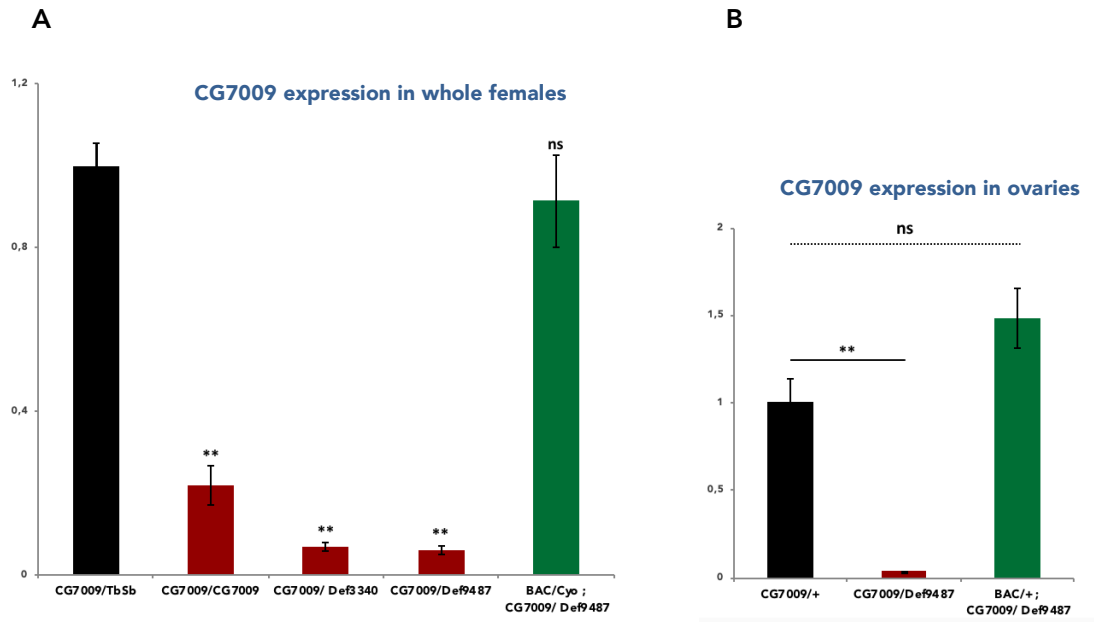

Figure 17. Characterization of *CG7009* expression levels. Determination by RT-qPCR of the expression level of *CG7009* in comparison to its expression in 4 days old heterozygous females "CG7009/TbSb" (**A**) or in the ovaries of 4 days old flies "CG7009/+" (**B**) (black graphs). Indicated in red are the expression levels of CG7009 transcripts in three null mutants: homozygous mutants "CG7009/CG7009" (**A**) and trans-heterozygous mutants "CG7009/Def3340" (**A**) and "CG7009/Def9487" (**A & B**). The green chart represents the expression of CG7009 from one dose of the rescue allele "BAC" in "CG7009/Def9487" mutant background, by using the strain "BAC/Cyo; CG7009/Def9487" in the whole females (**A**) and the strain "BAC/+; CG7009/Def9487" in the ovaries (**B**). Represented are the average values of two independent biological replicates. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. All alleles denoted as CG7009, correspond to the mutant allele *CG7009*^{e02001}; BAC: Bacterial Artificial Chromosome containing the *CG7009* genomic region; Tb, Sb, CyO: balancer chromosomes; Def: Deficiency.

5. Characterization by PCR of *CG7009* trans-heterozygous mutants

During my PhD we also worked with hetero-allelic mutants for *CG7009* using two deficiencies, Def9487 and Def3340. The deficiencies are chromosomal deletions encompassing more than one, usually tens of genes. Def9487 (*Df(3R)ED10845*) is a deficiency corresponding to stock BL9487 of the Bloomington collection. Its ID in *FlyBase* is FBab0044383. The full genotype of Def9487 is w^{1118}; Df(3R)ED10845, P{3'.RS5+3.3'}ED10845/TM6C, cu^1 Sb^1. It was generated in the terms of the *DrosDel* deletion project (Ryder et al., 2007) and affects the expression of more than 40 genomic loci, including *CG7009* locus.

The second deficiency, Def3340 corresponds to a deletion of a region on the 3R chromosome removing several genes including *CG7009*. Its stock number in the Bloomington database is 3340 and it is also known as Df(3R)e-R1. Def3340 has the following genotype: Df(3R)e-R1, Ki^1/TM3, Sb^1 Ser^1 and *FlyBase* ID: FBab0002783.

We established trans-heterozygous mutants *CG7009*^{e02001} / Def9487 and *CG7009*^{e02001} / Def3340 by crossing the balanced stocks of these 3 mutants and recovering the combinations of interest in the first progeny. We then verified the presence and absence of the *CG7009* locus or mutated locus (containing the PiggyBac transposon) in both deficiencies (Figure 18). We equally established the strain BAC / Cyo; *CG7009*^{e02001} / Def9487 and re-verified the presence of the WT *CG7009* locus in this rescue line (Figure 18, lane 7).

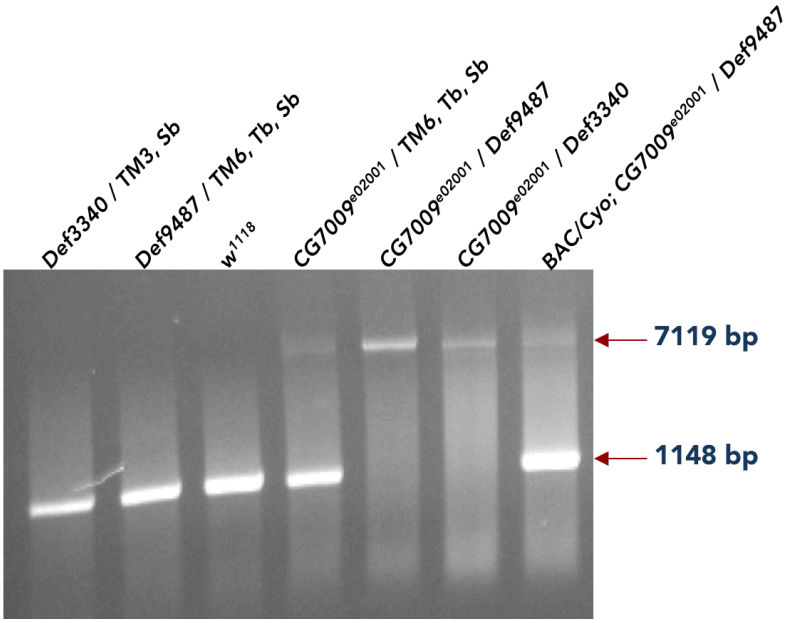

Figure 18. Validation by PCR of Def9487, Def3340 and BAC / CyO ; *CG7009*^{e02001} / Def9487. PCR on gDNA, extracted from flies with the indicated genotypes using primers CG7009-dTOPO FW and CG7009-dTOPO Rev (Methods and Figure 13). The band at 1148 bp corresponds to the WT CG7009 locus, the band at 7119 bp corresponds to the mutant allele *CG7009*^{e02001} with the inserted PiggyBac transposon (Figure 13). BAC: Bacterial Artificial Chromosome containing the *CG7009* genomic region; CyO, TM6, Tb, Sb and TM3: balancer chromosomes; bp: base pairs.

In Figure 18, the first 3 lanes are PCR control conditions, where we confirm only the detection of the smaller band of 1148 bp, corresponding to the WT *CG7009* locus in the reference strain used (w^{1118}, Figure 18, lane 3). The 1148 bp signal is also present in the strains containing the two deficiencies tested under wild type for *CG7009* balancer chromosomes, TM3, Sb for Def3340 and TM6, Tb, Sb for Def9487, respectively (lanes 1 and 2). In the heterozygous mutant *CG7009*^{e02001} / TM6, Tb, Sb (lane 4) we detect the WT allele (1148 bp) amplified from the balancer chromosome TM6, Tb, Sb and the insertional mutant allele (7119 bp), amplified from *CG7009*^{e02001}. Lanes 5 and 6 of Figure 18, present only the band of higher molecular weight that is amplified from *CG7009*^{e02001}. This demonstrates that the signal detected in the two heterozygous individuals Def3340 / TM3, Sb and Def9487 / TM6, Tb, Sb (lanes 1 and 2), *CG7009*

genomic sequence is amplified from the two balancers, confirming that the two deficiencies are deleted for the *CG7009* locus. In the last lane (lane 7) we amplified *CG7009* from strain BAC / CyO; Def9487 / TM6, Tb, Sb. The reappearance of the smaller band of the WT *CG7009* allele in this line (compare lanes 6 and 7), confirmed that during the crosses for the establishment of this trans-heterozygous rescue line the BAC-carried *CG7009* WT allele was not lost.

6. Identification of a second hit mutation on the *CG7009^{e02001}* allele

In the beginning of the project, another reason for our interest in the *CG7009* gene, apart its afore mentioned implication in the Ago-2-miRNA pathway, was the observation that we weren't able to maintain a homozygous stock with the Bloomington stock BL18008, bearing the null mutant allele *CG7009^{e02001}*. There were two possible hypotheses to explain this incapacity. Either *CG7009* loss-of-function was associated to semi-lethality or it was provoking fertility problems in flies (semi-fertility).

As the BL18008 stock was generated *via* insertional mutagenesis and this mutant was obtained by a consortium generating a big collection of mutants, the obtained stocks were not verified in detail. We were thus aware of the possibility that our allele is not "clean" and that it could carry a second insertion or a natural mutation disrupting the function of another gene in the 3rd chromosome (a second hit mutation). In order to be sure that we were not working with a mutant carrying additional mutations we had two possibilities: (i) We could "clean" the potential second hit by backcrossing the flies. This is accomplished by crossing heterozygous females over a sufficient number of generations with males with a controlled genetic background (*i.e.*, w^{1118}) until the putative second hit mutation is "cleaned" by recombination with wild type allele. (ii) Another option was to do a genetic analysis on the potential presence of a second hit mutation in the BL18008 stock.

As we had the two afore described publicly available deficiencies for *CG7009* we first undertook this strategy. It is not probable that a PiggyBac or another natural transposon inserts in the region encompassed by the deficiency of *CG7009*, thus working with a hetero-allelic mutant combination will most certainly put the potential second hit mutation in heterozygous state. Thus, we compared if the phenotypes of semi-lethality and/or semi-sterility of the homozygous mutant *CG7009^{e02001}* were maintained in the trans-heterozygous mutants *CG7009^{e02001}* / Def9487. If the studied phenotypes were maintained in both homozygous and trans-heterozygous mutants, it is strongly suggested that they were

due to the *CG7009* mutation. On the contrary, if the phenotypes observed in homozygous *CG7009* mutants were lost in the *CG7009* trans-heterozygous mutants, it would clearly demonstrate that the semi-lethality and/or semi-sterility are not linked to *CG7009* lack of function.

A. Sub-lethality test – rationale and result

In a first experiment we addressed the potential semi-lethality thought to be a possible cause of the incapacity to maintain a stock $CG7009^{e02001}$ / $CG7009^{e02001}$. We crossed either males and females $CG7009^{e02001}$/ *TM6,* Tb, Sb together or females $CG7009^{e02001}$/ *TM6,* Tb, Sb with males Def9487 / TM6, Tb, Sb, and looked in the progeny for the proportion of mutant descendants CG7009^{e02001} / $CG7009^{e02001}$ or $CG7009^{e02001}$ / Def9487, respectively (Figure 19A).

The expected proportion for both mutants in the F1 generation is 33%, as balancer / balancer is a known lethal combination. Three independent crosses were done for both conditions and the proportion of the descendent with the corresponding mutant genotypes was calculated as a function of the total number of progenies for a given cross (Figure 19B). The bars on Figure 19B represent the average proportion of mutants in F1 from three biological replicates. The result indicates that the proportion of $CG7009^{e02001}$ / Def9487 individuals (34.75 %, Figure 19B, light blue bar) is corresponding to the theoretical expected 33%. In contrast, the descendants with genotype $CG7009^{e02001}$/$CG7009^{e02001}$ represented 18.58 % and not 33% of the F1 generation (Figure 19, dark blue bar), indicating that there is a lethality problem associated with this homozygous mutant. Thus, the result with the trans-heterozygous mutant demonstrated that the semi-lethal phenotype, associated originally with the stock BL18008 ($CG7009^{e02001}$/$CG7009^{e02001}$), is not due to the loss-of-function of the *CG7009* gene but to another mutation on the third chromosome.

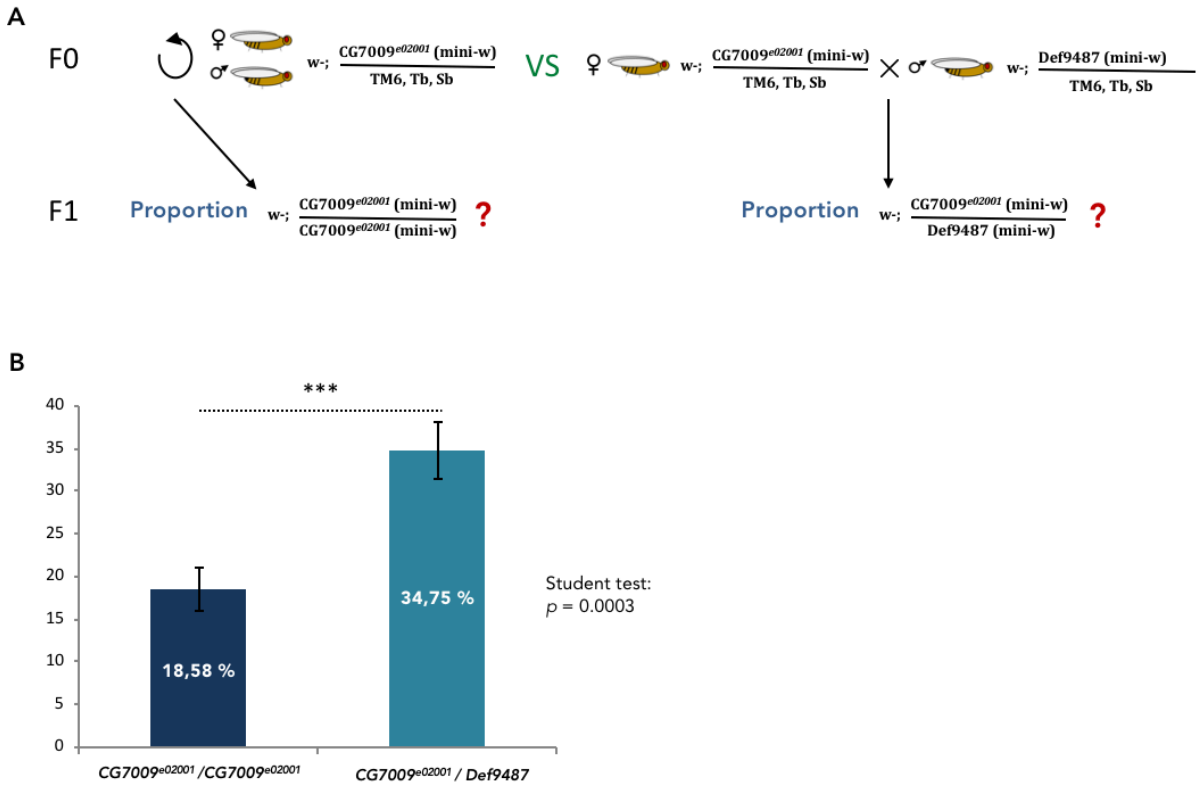

Figure 19. The function of *CG7009* is not linked to a sub-lethality phenotype. A. Schematic representation of the analysis of semi-lethality, associated with the Bloomington stock 18008 (allele *CG7009*^{e02001}). Comparison of the proportion of descendants arising from the cross of heterozygous individuals *CG7009*^{e02001} / TM6, Tb, Sb (left) to the proportion of descendants arising from the cross of heterozygous females *CG7009*^{e02001} / TM6, Tb, Sb and trans-heterozygous males *CG7009*^{e02001} / Def9487. **B.** Proportion of the descendants in F1, obtained from the described crosses, calculated as ratio between the number of the descendants with mutant homozygous *CG7009*^{e02001} / *CG7009*^{e02001} (dark blue graph) or trans-heterozygous *CG7009*^{e02001} / Def9487 genotypes (light blue graph) to the total number of descendants in the corresponding cross. Represented are the average values of three independent crosses for the indicated experiment conditions. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001.

B. Female fertility test – rationale and result

We also wondered whether the incapacity to maintain a homozygous stock *CG7009*^{e02001} / *CG7009*^{e02001} was caused by a semi-sterility phenotype in males or in females. For the male fertility we crossed homozygous *CG7009*^{e02001} / *CG7009*^{e02001} or trans-heterozygous *CG7009*^{e02001} / Def9487 males with females w^{1118} and analyzed the number of the progeny in F1. No fertility problems were detected for both homozygous and trans-heterozygous mutants compared to control crossings (data not shown).

A similar test was performed for the female sterility by crossing homozygous *CG7009*^{e02001} / *CG7009*^{e02001} or trans-heterozygous *CG7009*^{e02001} / Def9487 female mutants to w^{1118} males and determining the number of descendants in the progeny for both crosses (Figure 20A, red panel). For both

conditions, two crosses between 8 females and 3 males were performed and the number of descendants was determined in 4 flips of these parents. The total number of F1 progeny from these crosses was 486 for the homozygous female parents and 916 for the trans-heterozygous female parent, respectively. As control, we counted the number of descendants from crosses of w^{1118} flies or of w^{1118} males and females *CG7009*^{e02001} / TM6, Tb, Sb (Figure 20A, blue panel). For both of the control crosses only one tube of 8 females and 3 males was started. The result in Figure 20B is thus represented by the average of the progeny per female, because at the beginning of the experiment, not enough control crosses were initiated in order to perform a relevant comparison or statistical analysis.

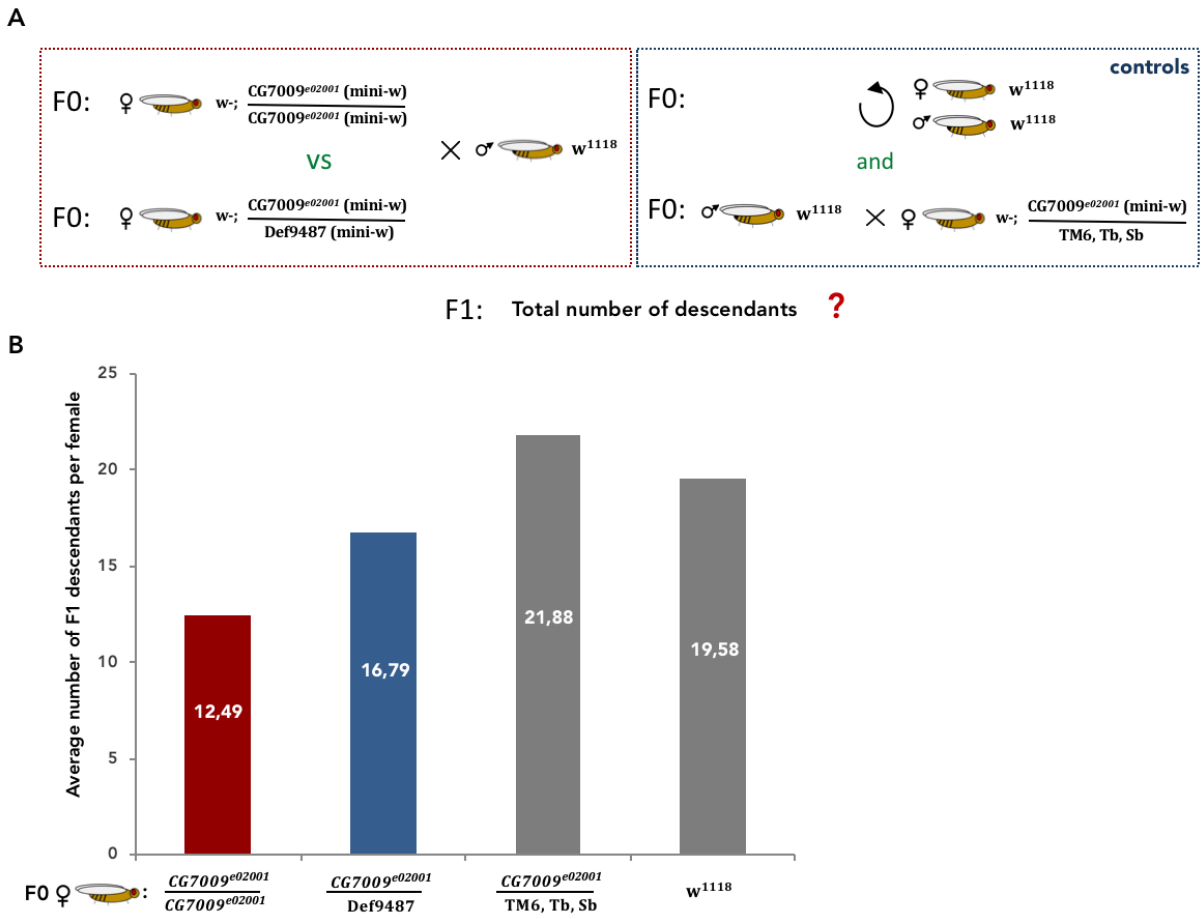

Figure 20. Effect of *CG7009* mutation on the fertility in females. **A.** Rationale for the experiment on the semi-sterility in females from Bloomington stock 18008 (allele *CG7009*^{e02001}). The experiment consisted in comparing the total number of descendants arising from the cross of homozygous females *CG7009*^{e02001} / *CG7009*^{e02001} with males w^{1118} (red panel, upper cross) and the total number of descendants arising from the cross of trans-heterozygous females *CG7009*^{e02001} / Def9487 with males w^{1118} (red panel, lower cross). As control the total number of descendants arising from the cross of w^{1118} flies (blue panel, upper cross) of females w^{1118} and trans-heterozygous males *CG7009*^{e02001} / Def9487 were to be determined. **B.** The result is represented as an average of the descendants *per* female from the following crosses: 2 crosses between 3 males w^{1118} and 8 females *CG7009*^{e02001} / *CG7009*^{e02001} (red graph); 2 crosses between 3 males w^{1118} and 8 females *CG7009*^{e02001} / Def9487 (blue graph); 1 control cross between 3 males w^{1118} and 8 females *CG7009*^{e02001} / TM6, Tb, Sb (first grey graph); and 1 control cross between 3 males w^{1118} and 8 females w^{1118} (second grey graph). The number of descendants was counted for 4 flips of parents in the indicated crossings and determined first as an average result at each flip by taking into account the dead parents. The final results represent another average value determined from the number of descendants per female for each flip and cross.

A first observation on these results is that there seems to be no difference between the two control crosses used in the experiment (Figure 20B, grey bars). The average number of descendants per homozygous female *CG7009*^{e02001} / *CG7009*^{e02001} is approximately two times lower than in the controls (12.49 descendants versus 21.88 or 19.58 per 1 female, Figure 20B compare the red bar to the two grey bars). This indicated that the *CG7009*^{e02001} allele might be linked to female semi-sterility. *CG7009* trans-heterozygous females have more progeny than the homozygous females, indicating that the observed

semi-sterility of the allele *CG7009*^{e02001} is not uniquely linked to CG7009 loss-of-function (Figure 20B, compare the blue and the red bars). However, this simplified analysis suggests that the function of *CG7009* might have an effect on female fertility, because the trans-heterozygous mutants do present less progeny on average per female than the two controls (Figure 20B, compare the blue to the grey bars).

The conclusion of this experiment was that the observed female semi-sterility is majorly not associated with CG7009 loss-of-function.

The result with the female fertility is also interesting because *CG7009* mutants have reduced ovaries size, which might explain at least partly the fertility problems (Figure 21).

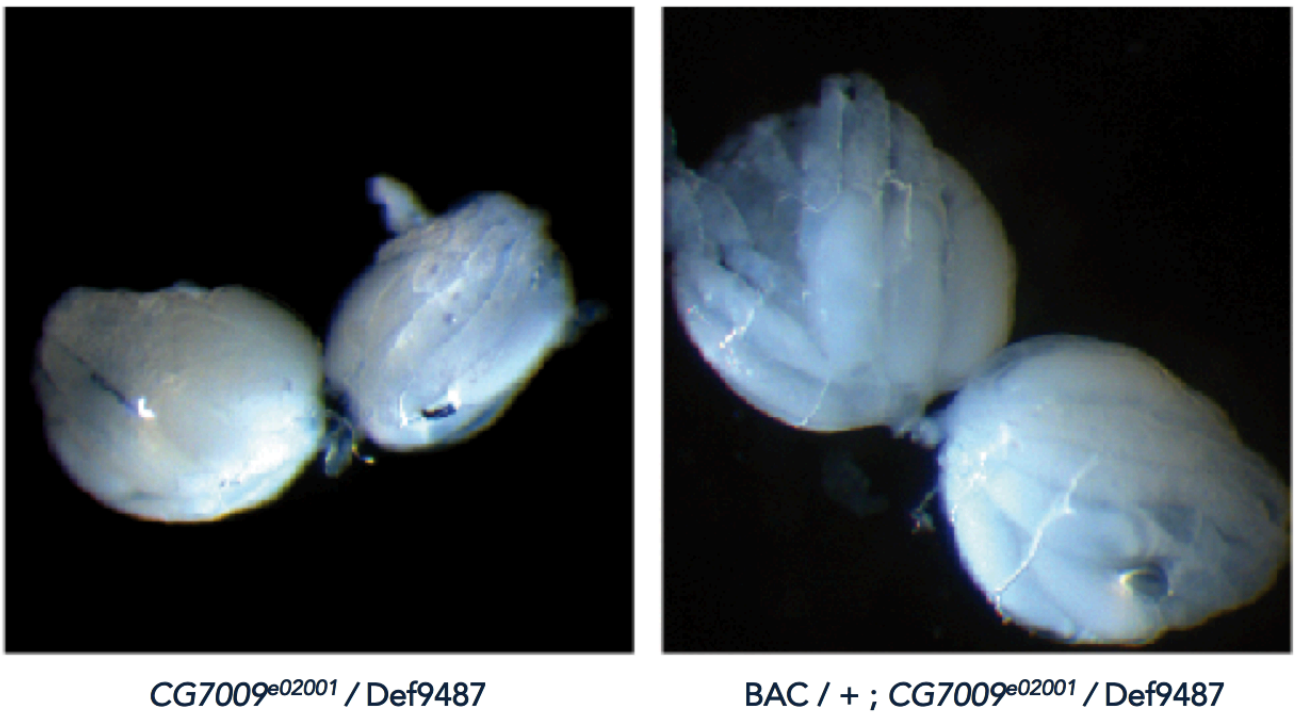

Figure 21. *CG7009* mutation is associated with ovarian size reduction. The image represents ovaries from 4 days old female of trans-heterozygous CG7009 mutants *CG7009*^{e02001} / Def9487 and the genetic rescue with the line BAC / + ; *CG7009*^{e02001} / Def9487.

7. Establishment and verification of the *CG7009*$^{e02001\text{-}G10}$ line

In order to remove the second hit mutation from chromosome 3 of the *CG7009*^{e02001} allele, we cleaned the chromosome by backcrossing heterozygous females with males with a controlled genetic background (w^{1118}) for 10 generations. This is how we established the mutant *CG7009*$^{e02001\text{-}G10}$ allele where G10 stands for backcross generation #10. It is at this point that we realized that we manage to maintain homozygous

stock with this new "cleaned" mutant stock *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$. This indicated that the second hit mutation allele was changed by recombination with a wild-type allele as we lost the phenotype of semi-sterility and or semi-lethality, associated with the homozygous stock *CG7009*^{e02001} / *CG7009*^{e02001}.

We then verified that the *PiggyBac* transposon was still present in the stock *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ by PCR on homozygous *versus* heterozygous individuals for this allele (data not shown), as described for the mutant *CG7009*^{e02001} / *CG7009*^{e02001} (Figure 13). I also determined the expression of *CG7009* transcript from the new homozygous cleaned G10 stock by RT-qPCR (Figure 22).

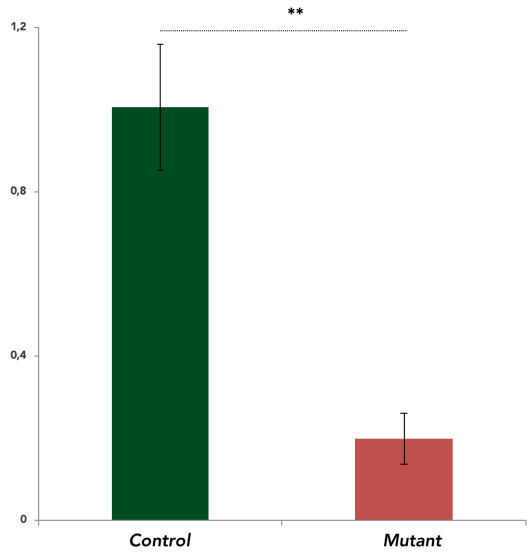

Figure 22. Expression of *CG7009* from the allele *CG7009*$^{e02001\text{-}G10}$. Determination by RT-qPCR of the expression level of *CG7009* in the 4 days old homozygous mutants *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$, denoted as Mutant (red graph) relative to *CG7009* expression in 4 days old heterozygous females *CG7009*^{e02001}/TM6, Tb, Sb, denoted as Control (green graph). Represented are the average values of two independent biological replicates. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05; Tb, Sb, balancer chromosomes.

8. Analysis of the effect of *CG7009* mutations on the longevity of *Drosophila*

During the characterization of the potential semi-lethal phenotype linked to the *CG7009*^{e02001} allele, we also performed a survival test in 3 biological replicates, using 10 females for each replicate. This result (not shown) indicated that the mutant *CG7009*^{e02001}/*CG7009*^{e02001} lives on average 24 days, and the mutant *CG7009*^{e02001}/*Def9487* survives on average between 60 and 70 days (67 days). Thus, the earlier mortality observed in the stock BL18008, bearing the allele *CG7009*^{e02001}, was not primarily caused by *CG7009* loss-of-function, but to the aforementioned second hit mutation. However, this first survival experiment was interesting, because the trans-heterozygous mutants seemed to live less longer than a heterozygous individual *CG7009*^{e02001}/ + (data not shown).

Determining if CG7009 has a role for normal longevity in *Drosophila* was also interesting. On one hand CG7009 is a predicted Nm tRNA methylase, and m^5C tRNA methylase Dnmt2 has also been implicated in *Drosophila* longevity (Durdevic et al., 2013a). Importantly, both Dnmt2 (Durdevic et al., 2013b) and CG7009 affect the efficiency of the siRNA pathway activity in *Drosophila*. On the other hand, CG7009 is a predicted 2'-O-methyltransferase and DmHen1, the only known 2'-O-methyltransferase, involved in the sncRNA pathways (Horwich et al., 2007; Li et al., 2005; Park et al., 2002; Saito et al., 2007; Yu et al., 2005) is also implicated in flies longevity, as it has been shown that *dmHen1* mutants have a shorter lifespan (Abe et al., 2014).

By the time this first survival experiment was performed, the BAC line was not yet established. We thus decided to reproduce the survival test, by including the rescue line for *CG7009*. For this survival experiment we worked with the second deficiency for *CG7009*, Def3340 (Figure 23). As for the previous experiments we analyzed the longevity of three females replicates. The flies were of the following genotypes: mutants $CG7009^{e02001}$ / Def3340 (Figure 23, red curve), heterozygotes $CG7009^{e02001}$ / TM6, Tb, Sb (Figure 23, black curve), and rescue for *CG7009*: BAC / BAC, $CG7009^{e02001}$ / Def3340 (Figure 23, green curve).

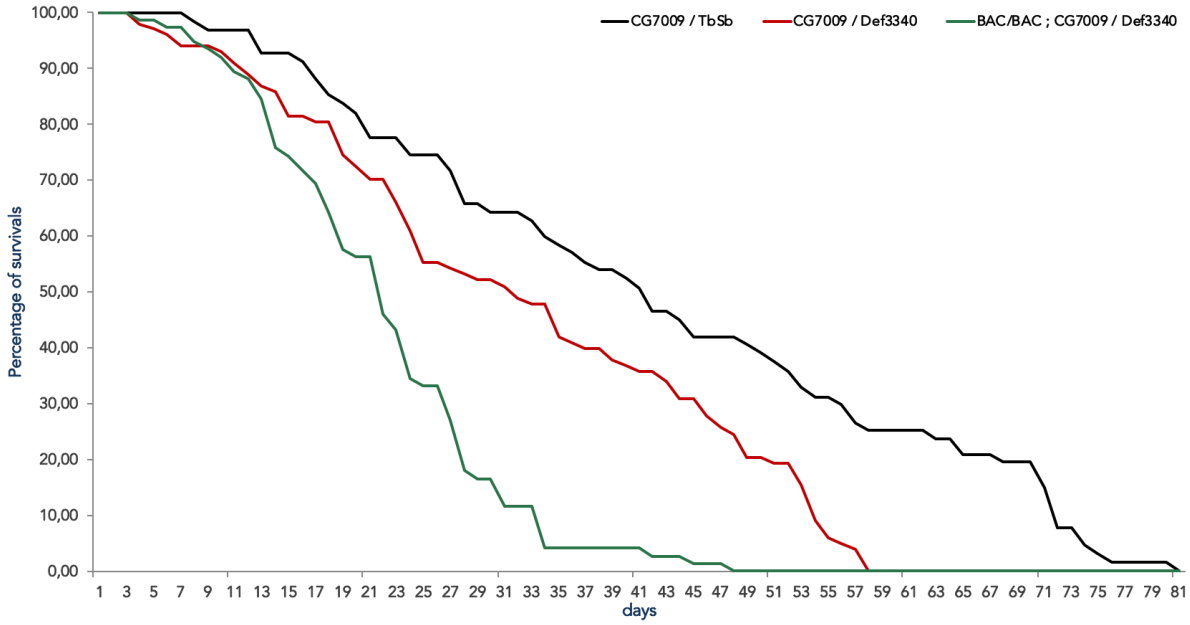

Figure 23. Survival experiment. Life span curves, representing the percentage of survival of heterozygous individuals $CG7009^{e02001}$ / TM6, Tb, Sb, denoted as "CG7009 /TbSb" (**black curve**), trans-heterozygous CG7009 mutants $CG7009^{e02001}$ / Def3340, denoted as CG7009/Def3340 (**red curve**) and recue line BAC/BAC; $CG7009^{e02001}$ / Def3340, denoted as BAC/BAC; CG7009/Def3340 (**green curve**). The curves represent the average values of at least three biological replicates. The flies of the corresponding genotypes were aged under similar conditions and the experiment was performed at 25°C.

The result indicated that trans-heterozygous mutants *CG7009*^{e02001}/Def3340 die approximately 20 days earlier than heterozygous individuals *CG7009*^{e02001}/TM6, Tb, Sb (Figure 23, compare the black and the red curves). However, in this particular experiment, the individuals with genotype BAC / BAC ; *CG7009*^{e02001} / Def3340 (Figure 23, green curve), which correspond to the *CG7009* rescue line reach 0% of survival first. The BAC was thus not only incapable to rescue the observed phenotype of reduced lifespan, but its presence provoked a stronger survival defect.

In another independent survival experiment with triplicates of the same genotypes, *CG7009*^{e02001} / *Def3340* mutants attained 0% of survival 19 days earlier than the heterozygote *CG7009*^{e02001} / *TM6, Tb, Sb* (data not shown). This seems consistent with the result represented in Figure 23 if we compare the black and the red curves. The result was thus reproducible in two independent survival experiments, and *CG7009*^{e02001} / *Def3340* mutants had a lifespan of 20 days shorter than *CG7009*^{e02001} / *TM6, Tb, Sb* flies. However, as for the BAC line, in this second survival experiment (not shown) the phenotype of reduced longevity was partially rescued. Flies from the rescue line attained 0% of survival 9 days after the mutant and 10 days before the heterozygotes. Thus, the BAC line behaved oppositely in regards to the mutant in this rescue test (not shown) when compared to the result on Figure 23.

I decided to present the survival experiment represented in Figure 23 in order to introduce several important observations concerning the BAC line. All results obtained with the BAC line are to be considered with caution. Despite the correct RNA expression of *CG7009* from this line, as determined by RT-qPCR (Figure 22), we noticed that the insertion of the BAC renders the flies "sick", as the BAC line stocks seem to have higher occurrence of lethality as well as higher occurrence of infections in the stock tubes.

We never studied these observations in detail. It is possible that the insertion of the *CG7009* locus at the ectopic genomic location on the second chromosome has some additional consequences for the flies' fitness. It is also possible that despite the correct levels of *CG7009* mRNA expression from the BAC genomic clone, the presence of *CG7009* on its genomic location assures supplementary functions, unlinked to the expression of the gene.

In conclusion the BAC line either does not restitute all of the *CG7009* functions properly or it triggers supplementary effects in flies not linked to CG7009. However, as described below, the expression of *CG7009* from the BAC seems capable to reconstitute a very important property of CG7009, notably its

methylation activity (second part of the Results section, Figure 40). We are currently constructing a UAS-CG7009 rescue line in order to reproduce several crucial experiments, as survival test.

Determining the implication of the *CG7009* gene in *Drosophila* longevity turned out to be very complicated. A very recent survival test was also performed with mutants homozygous for the cleaned allele $CG7009^{e02001\text{-}G10}$ / $CG7009^{e02001\text{-}G10}$ (data not shown). Briefly, the mutants had a better fitness than the heterozygous individuals $CG7009^{e02001\text{-}G10}$ / TM6, Tb, Sb, which served as control. The rescue line used in this experiment was BAC / BAC ; $CG7009^{e02001\text{-}G10}$ / $CG7009^{e02001\text{-}G10}$. These flies still died before the homozygous mutants but outlived the control heterozygous individuals used in the experiment. Thus, it is very complicated to deduce if and how exactly CG7009 impacts ageing in flies. It is possible that one dose of the second hit, that remains in the $CG7009^{e02001}$ allele of heterozygous individuals is sufficient to impact flies' longevity. As a last option it is possible to reproduce the survival experiment by using as control heterozygous females $CG7009^{e02001\text{-}G10}$ / + or balancer, after homogenizing the genetic background variations by backcrossing the flies.

9. The *CG5220* gene

According to *FlyBase*, *CG5220* is located on the right arm of the 3rd chromosome (3R) and its genomic sequence location is 3R:17040682…17041895 [+] (cytological map location 89E10). *CG5220* has one annotated transcript and encodes for one reported polypeptide of 302 amino acids (33.22 kDa, UniProt, Q9VEP1). Analysis of sequences homology suggests that the product of *CG5220* also belongs to the conserved RlmE family and Trm7 subfamily of the methyltransferase superfamily (Figure 12). No experimental data are available on *CG5220* mutant phenotypes. *CG5220* was identified in a study pertaining to its potential yeast ortholog Trm7 (Guy and Phizicky, 2015). trm7Δ mutants are characterized with severe growth defects and the group discovered in a rescue experiment that expressing the fly *CG5220* in trm7Δ yeast mutants is capable to suppress the growth defects observed in trm7Δ yeast (Guy and Phizicky, 2015). For *CG5220* no mutants are available in the public databases. We generated several mutants for the gene by using CRISPR/Cas9 (see below) and showed that the mutants are, like for *CG7009*, viable and fertile.

A. Generation of catalytically inactive *CG5220* mutants by CRISPR/Cas9

Understanding the functional importance of genes is linked to characterizing the phenotypes triggered by their mutation, inactivation, or overexpression. Flies are a developed model for genetic research and numerous strategies have been developed to modify their genome from random gene disruption to insertional mutagenesis (St Johnston, 2013).

Over the years, research on *Drosophila* has yielded a large collection of mutants. However, for *CG5220* no mutants were available in public databases. When RNAi-mediated inactivation of *CG5220* was used to study its functions, another complication raised from its high genomic sequence homology with the gene *CG7009*. This complicated the design of efficient and specific siRNA-producing dsRNAs that will not inhibit simultaneously at least partially the expression of *CG7009*. In order to unambiguously determine the function(s) of this gene we decided to construct mutants, allowing a complete disruption of it function.

As of today, the ability to alter genomes has gone one step further with the recent advances in the CRISPR/Cas9 technology (Bassett et al., 2015; Gratz et al., 2013b; Jinek et al., 2012; Ren et al., 2014). This tool was initially discovered in bacteria and later adapted in eukaryotes (Jinek et al., 2012; Mali et al., 2013). CRISPR/Cas9-mediated homologous recombination requires a single guide RNA (sgRNA), a

template for recombination (ssODN: single strand oligodeoxynucleotide), and a Cas9 nuclease. CRISPR/Cas9 advantages are its remarkable efficiency, its requirement of limited number of upstream reagents a well as the rapidity in the generation of precise mutations *via* nucleotide replacements. Pioneer works on adapting CRISPR/Cas9 through Non-Homologous End Joining (NHEJ) or Homologous Directed Repair (HDR) in *Drosophila* research was accomplished by Gratz and colleagues (Gratz et al., 2013a; Gratz et al., 2014).

As a new emerging method assuring high mutagenesis rates, CRISPR/Cas9 seemed to be a suitable strategy for accomplishing mutagenesis on *CG5220*. We decided to replace two conserved amino-acids in the predicted catalytic domain of CG5220, which were reported as important for the catalytic function of its presumable yeast and human orthologs Trm7 and FTSJ1, respectively (Guy et al., 2015).

For reminder, FTSJ1 is known to methylate both positions C_{32} and G_{34} in $tRNA^{Phe}$ by interacting with two different partner proteins – THADA for C_{32} and WDR6 for G_{34}. In their study the authors showed that the yeast *trm7-K28A* substitution of a lysine codon at position 28 with an alanine codon (K28A) abolished the catalytic activity of Trm7 for Nm deposition on C_{32} and G_{34} of $tRNA^{Phe}$. They also described a distinct mutation of human FTSJ1 (FTSJ1-p.A26P missense allele) and the corresponding yeast engineered mutant *trm7-A26P*. In this second mutant version of FTSJ1 an alanine codon at position 26 is substituted with a proline codon (A26P). This A26P form of FTSJ1 disrupts its catalysis of Nm deposition on G_{34} of $tRNA^{Phe}$ but leaves the C_{32} position normally methylated in the same tRNA. The group discovered that human patients, bearing a substitution A26P in their FTSJ1 gene, develop the NSXLID. This suggested that the Nm deposition on G_{34} of $tRNA^{Phe}$ is the crucial modification that might be linked to NSXLID development in human (Guy et al., 2015). Thus, obtaining an A26P mutated form of CG5220 was interesting for us because, if the described methylation specific defect was conserved in flies, it could provide a catalytically inactive protein for only one of the methylations (Gm_{34}) whose absence was shown as sufficient for pathology development in human. Likewise, we reasoned that this mutant could enable us to determine on the molecular level the importance of Cm_{32} and Gm_{34} modifications for the fitness of the flies, by analyzing if lack of methylation on G_{34} position (located in the AC loop) in *Drosophila* is sufficient to trigger the phenotypes associated with *CG5220* loss of function.

In order to obtain the described catalytically inactive CG5220 we used CRISPR/Cas9-mediated one-step homologous recombination without insertion of a phenotypic marker at the *CG5220* locus. All of

the experiments on the generation of the *CG5220* mutants were performed by our collaborator Dr. Matthias Schaefer from the Medical University of Vienna, Austria.

B. Mutagenesis of the CG5220 lysine 28 and alanine 26 codons in *Drosophila*

Choosing the right single guide RNA (sgRNA) is important for the mutagenesis efficiency, as it has been shown that the cleavage site should be as close as possible to the sequence to be edited. The two corresponding sgRNAs (Methods) were cloned for *Drosophila* mutagenesis into the pDCC6 vector allowing their expression in *Drosophila* embryos (Gokcezade et al., 2014). To substitute the target codons, we generated a 136 nt long ssODN template carrying the lysine (AAG) to alanine (GCG) substitution (K28A) and a 136 nt long ssODN template carrying the alanine (GCC) to proline (CCA) substitution (A26P). The whole ssODN was homologous to CG5220 locus except the single base pair changes for the mutations plus the mutation of the PAM domain. *Drosophila* w^{1118} embryos were injected with a mix of ssODN and the bicistronic vector pDCC6, allowing the simultaneous expression of the corresponding sgRNA and the Cas9 as described in (Gokcezade et al., 2014) (Figure 24).

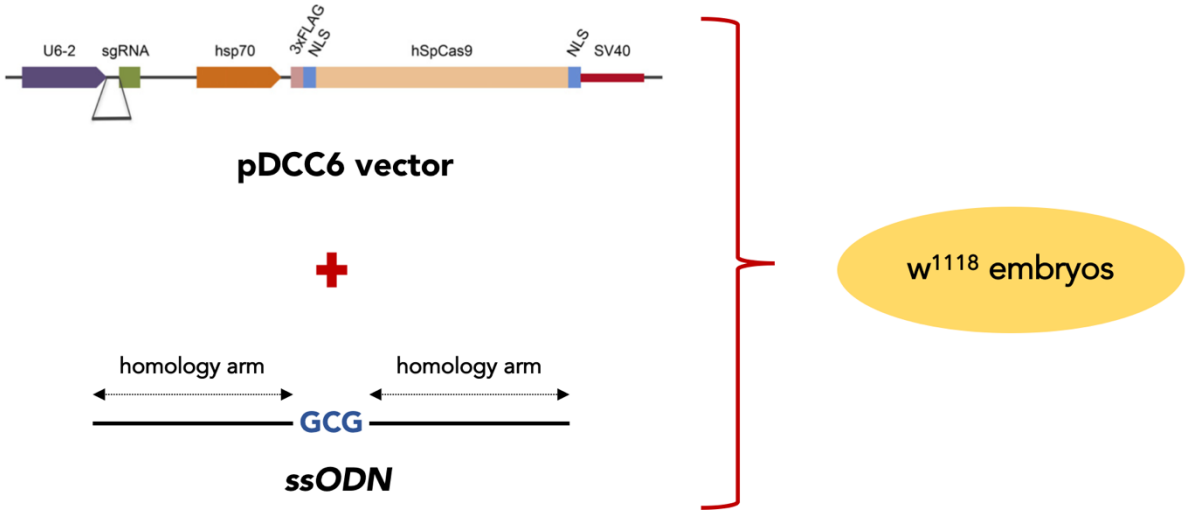

Figure 24. Mutagenesis reagents. Represented are the three components required for CRISPR/Cas9-mediated homologous recombination: single guide RNA (sgRNA), template for recombination (ssODN), and Cas9 nuclease. The gRNA, corresponding to PAM sequences "A-P" and "K-A", were subcloned into the bicistronic *Drosophila* CRISPR/ Cas9 vector pDCC6, expressing sgRNA cassette under the control of the *Drosophila* U6:96Ab (U6-2) promoter as well as an hsp70Bb promoter driving Cas9 expression Recombination template was provided as a 136 nt long ssODN containing the desired mutation (GCG alanine codon for the substitution K28A and CCA for the substitution A26P). These two components were injected in embryos from the w^{1118} strain. Figure adapted from (Gokcezade et al., 2014).

C. Screen for *CG5220* mutations

After injection, single wings or whole adult males were tested for editing events. Interestingly, the *CG5220* locus turned out to be hardly accessible for PCR on small amounts of wing tissues, which might have to do with the chromatin structure. That is why Matthias Schaefer had to cross out the F1 generation to balancers before sacrificing the males for molecular analysis. F0 flies were individually crossed with a TM3, Ser balancer strain. Then, F1 flies bearing the Ser balancer were individually crossed with the same balancer strain. Once larvae emerged, the F1 parents were collected for genotyping. A first step of the phenotype-independent molecular screening consisted in T7 endonuclease I (T7E1) assays for mismatches at the locus of *CG5220* (Vouillot et al., 2015). T7E1 digestion assays allow to identify mutational events by recognizing a subset of SNPs (single nucleotide polymorphisms), depending on the base substitution (Tsuji and Niida, 2008) that form upon amplifying heterozygous regions by PCR. Once the editing events (punctual mutations) were detected by T7EI digestion, an allele specific PCR using primers matching the two mutations was performed (Figure 25).

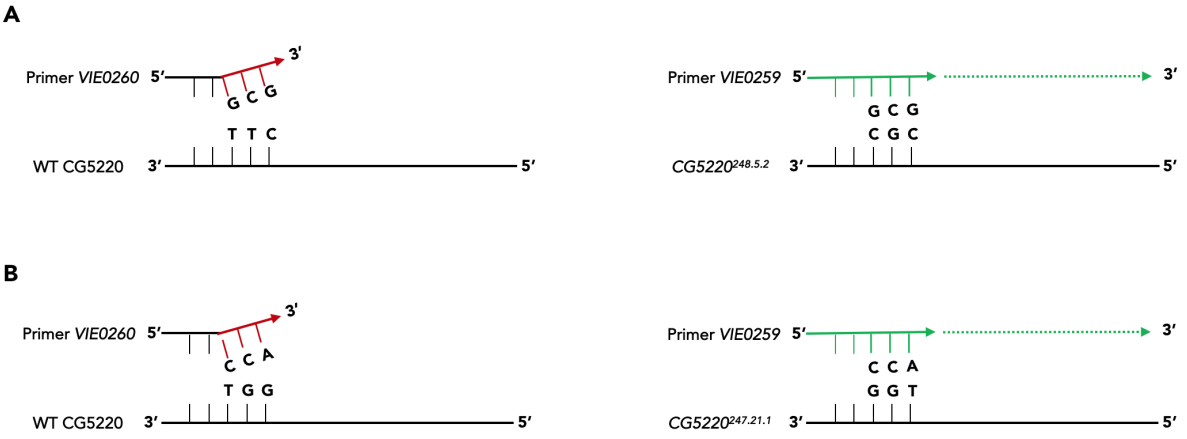

Figure 25. Molecular screening for the CG5220 K28A allele by PCR. A. Rationale for the discriminative PCR for constructing the mutant allele *CG5220*$^{248.5.2}$ bearing the substitution K28A.The primers VIE0260 ending with the alanine GCG codon which correspond to the desired mutation do not anneal to the WT version of CG5220 (*left*) but hybridize with the edited allele *CG5220*$^{248.5.2}$ (*right*). **B.** Rationale for the discriminative PCR for constructing the mutant allele *CG5220*$^{247.21.1}$ bearing the substitution A26P.The primers VIE0259 ending with the proline CCA codon which correspond to the desired mutation do not anneal to the WT version of CG5220 (*left*) but hybridize with the edited allele *CG5220*$^{247.21.1}$ (*right*).

About five single-male F2 progeny from 10 injected founder flies were screened for the presence of the recombined A to P or K to A mutations in the catalytic center by locus-specific PCR amplification, using PCR oligos that anneal to the mutated codon. For increasing the PCR specificity, the discriminative allele-specific primers were designed to match the edited codon at their 3′-end, as it is reported that the 3′-end base pairing is the most important for polymerase elongation (Simsek and Adnan, 2000). Then a screen for the allele *CG5220*$^{248.5.2}$ (substitution K28A) consisted in amplifying the *CG5220* edited locus with an alanine codon matching primer (VIE0260, Methods section) and a screen for the allele *CG5220*$^{247.21.1}$ consisted in amplifying the *CG5220* edited locus with the proline codon matching primer (VIE0259, Methods section), Figure 25. Five lines with A to P mutations, and three lines with K to A mutations were recovered. The editing events (mutations) were then crossed to obtain a balanced fly line and their viability was confirmed.

Few mismatches in the primer binding sequence can lead to delayed PCR amplification instead of its complete abolishment potentially giving rise to false positives. In order to exclude this, as well as to confirm the presence of the desired point mutations, the PCR amplicons were gel purified and subjected to Sanger sequencing. Sequencing of the PCR amplicons confirmed that the established fly lines contain either the A to P or the K to A mutations in *CG5220*. Flies carrying these mutations are homozygous viable

and look phenotypically normal (whether these are escapers or really can live without the balancer has to be seen over generations).

CRISPR/Cas9-mediated mutagenesis is known to produce off-target mutations that can be rather unpredictable. In order to assure that the phenotypes associated with *CG5220* disruption are indeed due to the loss of function of this gene it will be necessary to clean the potential off-targets in the three mutant fly stains generated by crossing them over several generations with a controlled genetic background (w^{1118}). This can be done by crossing heterozygous females with w^{1118} males for several generations. Since the *CG5220* alleles are not associated to phenotypic markers, females should be crossed individually at each generation, then genotyped.

In conclusion, we generated two different codon substitutions in the *CG5220* gene and obtained two new mutant alleles for this gene in *Drosophila melanogaster*. We managed to edit the alanine 26 to proline (A26P) and the lysine 28 to alanine (K28A) by single step CRISPR/Cas9 mediated HDR using a single strand DNA probe as template (ssODN).

D. Generation of a null mutant of *CG5220* allele (*CG5220*$^{-2nt}$) by CRISPR/Cas9

CRISPR/Cas9-induced double strand breaks are highly mutagenic and promote both HDR and NHEJ induced mutations at the desired loci. While we introduced a template to promote homologous directed repair, it was likely that NHEJ-mediated also occurred in the injected flies. The endonuclease activity of Cas9 generates double strand breaks at predictable sites, depending on the positioning of the used sgRNAs. Cas9 also possess an exonuclease activity, which was shown to promote insertion/deletion events (indels) occurring at the double strand break site and extending on 5' direction of the PAM containing strand (Bassett et al., 2015; Jinek et al., 2012).

Therefore, it was conceivable that some F1 flies could bear indels in the edited codons. Such mutated alleles would not be detected with our allele-specific method. Thus, we sequenced DNA sample extracted from F1 flies. We individually analyzed 10 F1 flies. Among them, we detected a mutant, bearing a deletion of two nucleotides, corresponding to nucleotides 75 and 76 of the reference CDS of CG5220. Analysis showed that this deletion, leading to a frameshift mutation, provokes an appearance of a premature TGA stop codon at nucleotides 211of the mutated CDS, giving rise to a protein of only 8 kDa, instead of the 33.22 kDa of the WT CG5220 protein. This is most probably a null mutant for *CG5220*, however more

functional studies are required in order to unambiguously confirm this. For establishing a balanced Indel containing line the flies were crossed with balancer TM3, Ser, Sb leading to event/TM3 Ser. The ensuing stock was used to sequence the *CG5220* locus from a homozygous fly. This revealed that the *CG5220* locus had a 2 bp deletion.

In conclusion, using CRISPR/Cas9-mediated mutagenesis, we obtained, in collaboration with Dr. Schaefer, a third new allele of *CG5220* containing one deletion of two nucleotides (-2nt) and thus potentially giving rise to a null mutant.

10. Establishment and characterization of *CG7009*, *CG5220* double mutants

A. Fly crosses for *CG7009*, *CG5220* double mutant recombinants recovery

CG7009, CG5220 double mutants were obtained by homologous recombination of chromosome 3R during meiosis. In order to obtain the *CG7009, CG5220* double mutants the mating scheme represented in Figure 26 was used.

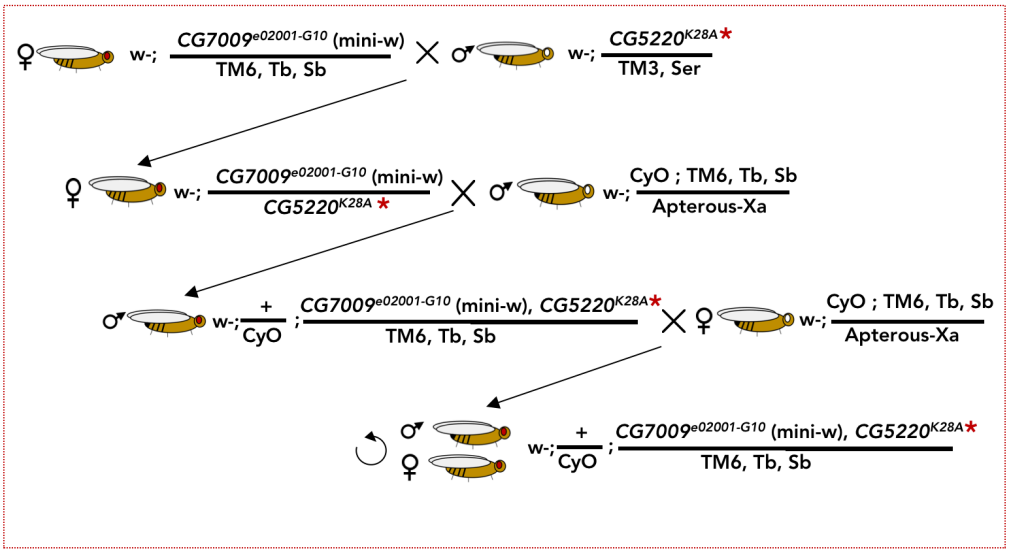

Figure 26. Mating scheme for the construction of the double mutant recombinants *CG7009*$^{18008\ G10}$, *CG5220*^{K28A}. Schematic representation of the genetic crosses performed for the establishment of the line w-; +/ Cyo; *CG7009*$^{18008\text{-}G10}$, *CG5220*^{K28A} / TM6, Tb, Sb. w-, mutant *white*; mini-w, mini-white marker, Apterous-Xa, CyO, TM6, Tb, Sb: balancer chromosomes; *CG7009*$^{18008\text{-}G10}$ is the cleaned by backcross mutant allele *CG7009*^{e02001}; *CG5220*^{K28A} corresponds to the allele *CG5220*$^{248.5.2}$. * The same mating scheme was used to obtain the mutants *CG5220*^{A26P} (not shown).

A. PCR Screen for *CG7009*, *CG5220* double mutant recombinants

The potential double mutant recombinant males with the expected recombinant genotypes (w- ; + / CyO ; $CG7009^{e02001\text{-}G10}$ (w+) , $CG5220^{K28A}$ / TM6, Tb, Sb) or (w- ; + / CyO ; $CG7009^{e02001\text{-}G10}$ (w+) , $CG5220^{A26P}$ / TM6, Tb, Sb, were selected on the (*w+*) of the $CG7009^{e02001\text{-}G10}$ allele and screened for the presence of the recombined alleles $CG5220^{K28A}$ or $CG5220^{A26P}$ by PCR with discriminative primers for the two mutants, as for the generation of the corresponding *CG5220* simple mutants by CRISPR/Cas9 (Figure 25 and Methods).

On the results on Figure 27, amplification with the *CG5220* mutant alleles primers on flies w^{1118} served as negative controls, and no amplification was detected. PCR on $CG5220^{K28A}$ single mutants (positive control) yielded a band at 284 bp, which is the expected size of the PCR product, for the mutation AAG→ GCG (Figure 27A, lanes "$CG5220^{K28A}$"). PCR on $CG5220^{A26P}$ single mutants (positive control) lead to the detection of a band at 279 bp, which is the expected size of the PCR product, for the mutation GCC→ CCA (Figure 27B, lanes "$CG5220^{A26P}$"). We detected 3 recombination events validating the establishment of the double mutants $CG7009^{e02001\text{-}G10}$, $CG5220^{K28A}$ (Figure 27A, lanes 1, 4 and 7) and 2 recombination events validating the establishment of double mutants $CG7009^{e02001\text{-}G10}$, $CG5220^{A26P}$ (Figure 27B, lanes J and R).

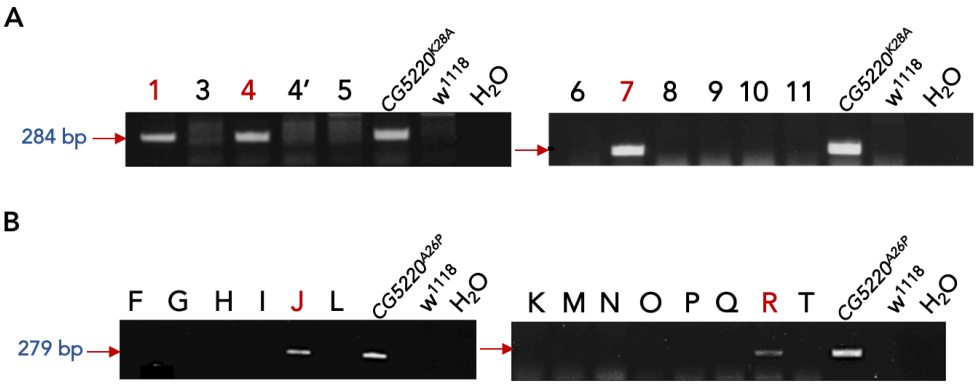

Figure 27. PCR Identification of CG7009, CG5220 double mutants. PCR was performed on gDNA from flies with the following phenotype [w+; Ap] for which they inherited the *CG7009* mutation (giving the [w+] phenotype) from the male parent and Ap from the female parent. To check if these flies have also inherited from the male parent a mutation $CG5220^{248.5.2}$ ($CG5220^{K28A}$) **(A)** or $CG5220^{247.21.1}$ ($CG5220^{A26P}$) (**B**), PCR using reverse primers specific to the mutations (VIE0260 and VIE0259) was performed (Methods). The corresponding simple mutants *CG5220* were used as the positive control, flies w^{1118} were used as negative control. Numbers 1-11 designate the crossings number of the tested for mutation $CG5220^{K28A}$ flies, similarly for the letters from crossing F-T for the screen of the $CG5220^{A26P}$ mutants. $CG5220^{248.5.2}$ PCR fragment is and 284 bp, and $CG5220^{A26P}$ is 279 bp. PCR products were loaded to 1.4% agarose gels, stained with ethidium bromide. bp: base pairs.

Then the double mutants corresponding to recombination events 1 and 4 for *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} and to the event J for CG*7009*$^{e02001\text{-}G10}$, *CG5220*^{A26P} were validated by Sanger sequencing (*GATC* Biotech, Figure 28). The sequences in Figure 28 represent antisense strands, as they were sequenced with the VIE0198 reverse primer (Methods).

To detect heterozygous *CG5220* mutations in the double mutant with *CG7009*, we examined each chromatogram and compared it with the reference sequence from a single *CG5220* mutant (heterozygote for K28A or homozygote for A26P). In addition, flies not bearing *CG5220* mutation, identified in the PCR on Figure 27 were used as negative controls: Rec6 for *CG5220*^{K28A} and RecF for *CG5220*^{A26P} (Figure 27A and B, lower panels).

Sequencing the edited loci confirmed that the positive genomic DNA PCR amplicons contained the recombinant alleles at the heterozygous state for *CG5220*^{K28A} (Figure 28A, second and third panels) and at the homozygous state for *CG5220*^{A26P} (Figure 28B, second panel).

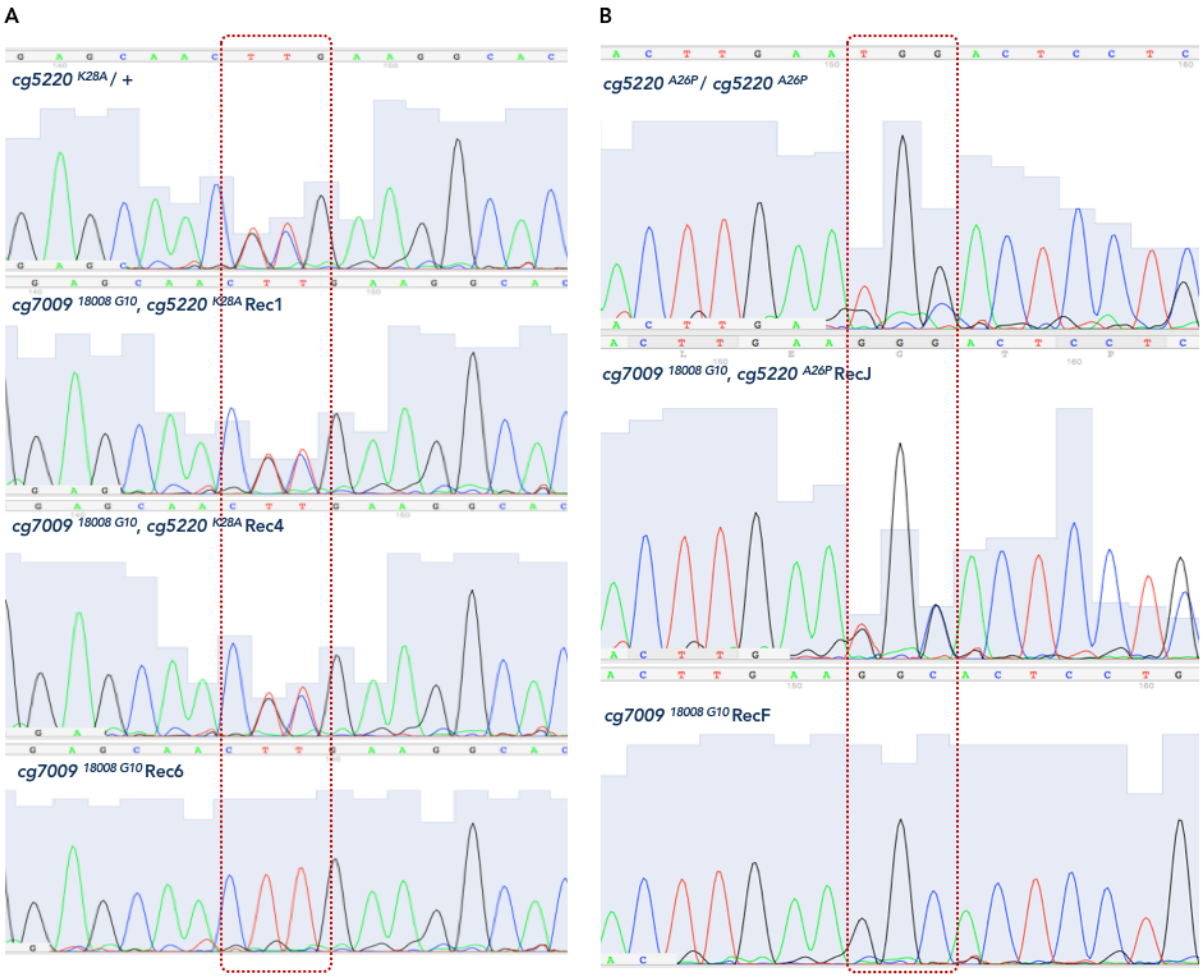

Figure 28. Validation by sequencing of the double mutants CG7009, CG5220. *CG5220* PCR fragments were amplified by PCR from flies bearing the *CG7009*$^{e02001\text{-}G10}$ mutant allele (giving the [w+] phenotype) and the mutations *CG5220*^{K28A} (**A**: #1 and #4) or *CG5220*^{A26P} (**B**: #J) determined previously by PCR were sequenced. Their corresponding simple mutant *CG5220* (a heterozygous for *CG5220*$^{248.5.2}$ and a homozygous for *CG5220*^{A26P}) were used as positive controls and flies characterized with no *CG5220* mutation were used as negative controls (**A**: #6 for *CG5220*^{K28A} and **B**: #F for *CG5220*^{A26P}). The results were generated with *4Peaks* and correspond to the PCR fragments sequenced with the reverse primer VIE0198. The substituted nucleotides are indicated in the red squares. Nucleotides color code: in blue: cytosine, red: thymine; green: adenine; black: guanine.

Chapter II: Discovering the molecular functions of CG7009 and CG5220

1. Construction of CG7009 recombinant proteins for *in vitro* methylation assays

An important line in my project that started from its very beginning was to characterize the molecular function, first of CG7009 and later also of CG5220. The molecular function of DmHen1/Pimet was originally described by *in vitro* methylation assays on small non-coding RNAs (si- and piRNAs) with purified recombinant DmHen1 and ^{14}C-SAM as methyl group donor (Horwich et al., 2007; Saito et al., 2007).

Likewise, in a first strategy we decided to construct and to produce a recombinant CG7009-tagged protein, a tool which could allow us to biochemically characterize its function. For this purpose, full-length N-terminally tagged recombinant FLAG_CG7009 and GST_CG7009 proteins were cloned and produced in *E. coli* following a routine procedure in the laboratory (Carre and Shiekhattar, 2011). The purified and intermediates products were analyzed using SDS-PAGE and protein gel coloration (Figure 29 and Methods).

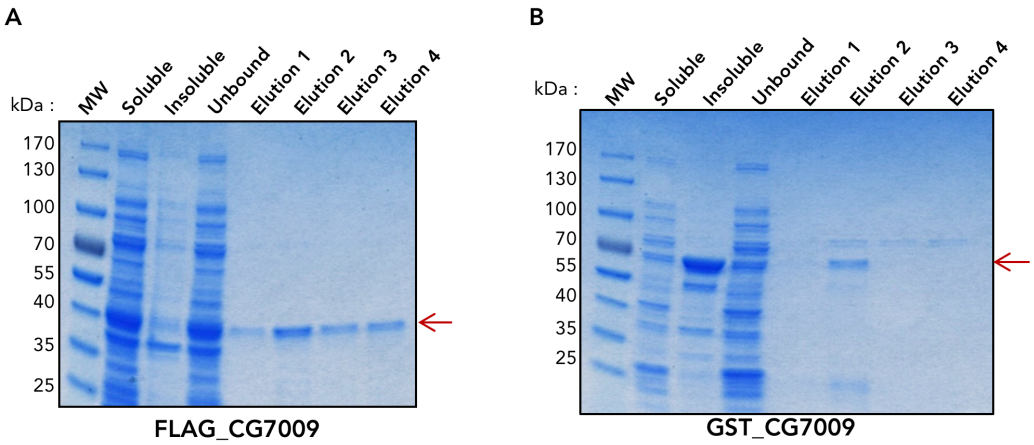

Figure 29. Purification of FLAG_CG7009 and GST_CG7009. Shown is the migration of the indicated fractions of FLAG_CG7009 (**A**) and GST_CG7009 (**B**) recombinant proteins in 4-20% SDS-PAGE. MW, Molecular Weight marker. Gels staining were made using the Simply Blue Safe Stain from Invitrogen.

For the eluted fractions of the FLAG_CG7009 purification (Figure 29A) we observe one band which appears to be pure and migrates at the expected molecular weight of ~38 kDa (size of the protein plus FLAG peptide size). However, the recombinant protein was not ideally bound to the anti-FLAG antibody-containing beads during the purification as a part of the protein seems to remain in the unbound fractions (Figure 29A, lane *unbound*). Reduction of salt concentration (here 500 mM KCl) in the washing steps and

increase of beads concentration might help to avoid this problem in the future. For the GST_CG7009 recombinant protein (61 kDa) we can see that an important proportion of the protein remains in the insoluble fraction, probably due to aggregation (red arrow, Figure 29B, lane *Insoluble*).

In conclusion, I managed to clone and establish the condition of the production and purification for the FLAG_CG7009 recombinant protein. I also constructed a recombinant GST_CG7009 protein. However, additional optimization of the experimental procedure might be needed for a proper production of the CG7009_GST recombinant protein to avoid aggregation problems encountered during this experiment, in particular for our ongoing collaboration to crystalize and resolve the structure of this protein with our crystallographer collaborators.

2. Expression of recombinant DmHen1 for *in vitro* methylation assays

As a positive and negative control for the *in vitro* methylation assays, aiming to test the putative CG7009 methyltransferase activity, we decided to produce recombinant GST_DmHen1 and GST_ΔDmHen1 respectively (74 kDa). The plasmids for these two constructions were already available (Saito et al., 2007). The experimental protocol for the production of these two proteins was the same as the one used for producing the CG7009-tagged recombinant proteins (Carre and Shiekhattar, 2011).

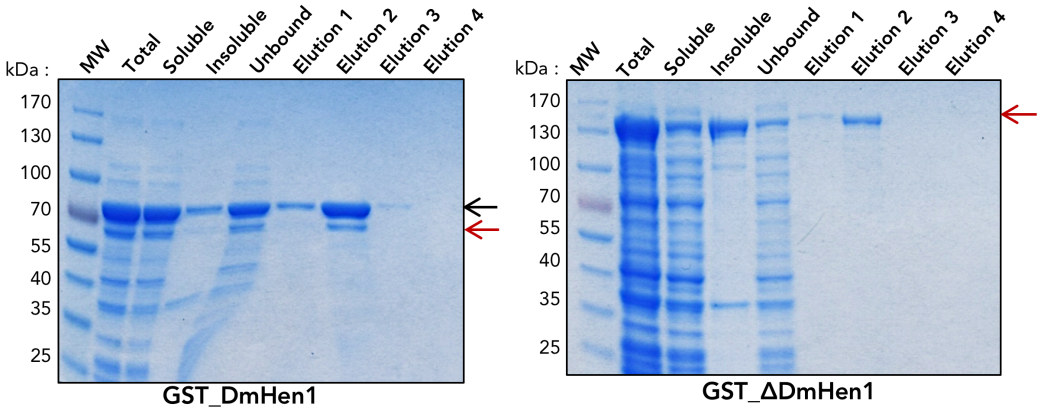

Figure 30. Purification of GST_DmHen1 and GST_ΔDmHen1. Purification of GST_DmHen1 (**A**) and GST_ΔDmHen1 (**B**). Shown is the migration of the indicated fractions of the respective tagged recombinant protein in 4-20% SDS-PAGE. MW, Molecular Weight marker. Gels staining were made using the Simply Blue Safe Stain from Invitrogen.

On the GST_DmHen-1 protein gel coloration we observe one band that migrates at the expected size for this fusion protein (74 kDa) (black arrow, Figure 30A). The protein seems to be unstable as in the eluted purified fractions a second band, migrating at a lower molecular weight appears (red arrow, Figure 30A). However, the decent amount of full-length protein in Elution 2 (black arrow) as well as the small portion of the supposedly degraded protein (red arrow) let us think that the methylation test could be carried out with this Elution 2.

For the mutant form ΔDmHen1 purification we observe one band migrating around 150 kDa which corresponds to the double size of the one expected for the recombinant ΔDmHen1 protein (Figure 30B, red arrow). This is probably due to the dimerization of the GST fusion protein in this condition of purification. Thus, for the production of the recombinant ΔDmHen1 protein, additional optimization of the experimental conditions will be needed.

3. *In vitro* methylation

A. *In vitro* methylation assays with recombinant CG7009-FLAG and CG7009-GST

The purified recombinant CG7009 proteins were used to test their capacity to methylate tRNAs, miRNAs and piRNAs *in vitro* in the presence of the methyl group donor SAM (S-adenosyl-L-[methyl-^{3}H] Methionine). We performed multiple *in vitro* methylation assays with modifications of the experimental conditions in the attempt to uncover if CG7009 has a methyltransferase activity.

The experiment on Figure 31 represents an *in vitro* methylation assays, performed with recombinant CG7009 proteins by following the protocol described in (Saito et al., 2007). The only difference with the protocol is that we used as a methyl group donor SAM marked with ^{3}H isotopes instead of ^{14}C-SAM (Methods). We assayed the methylation activity of FLAG_CG7009, GST_CG7009, GST_DmHen1 (positive control) and ΔGST_DmHen1 (negative control). Substrates for methylation were a single-stranded small RNA of 26 nt (synthetic piRNA, left part of Figure 31) and a 3'-terminally 2'-O-methylated 26 nt small RNA, which served as specificity control for the activity of GST_DmHen1, as DmHen1 is known to be able to transfer a methyl group from SAM to the 3'-terminal 2' hydroxyl group of the RNA and thus should not be capable to methylate this substrate. The sequences of the two small RNA substrates are available from (Saito et al., 2007).

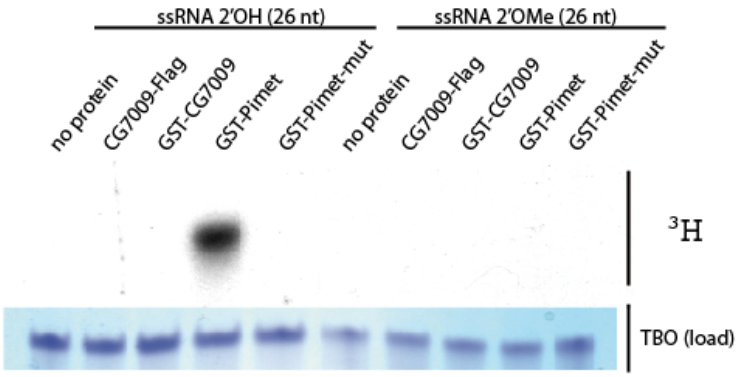

Figure 31. *In-vitro* methylation assay. Left part: *(ssRNA 2'OH (26nt):* Single-stranded small RNAs (26 nt, 0.04 nmol/50 µL) were incubated with GST_DmHen1 (GST-Pimet), ΔGST_DmHen1 (GST-Pimet-mut), GST_CG7009 or FlAG_CG7009 in the presence of ^{3}H-labeled SAM. Resultant small RNAs were run on a denaturing acrylamide gel. The *top* panel shows the autoradiography of ^{3}H-labeled RNAs. Only GST-Pimet (positive control) is able to transfer methyl groups from SAM to the single-stranded small RNAs (lane 4). The *bottom* panel shows a staining image (TBO, Toluidine Blue O) of the top gel, which indicates that comparable amounts of single-stranded small RNAs were used in all the lanes. **Right part:** *(ssRNA 2'OMe (26nt):* 3'-terminal 2'-*O*-modified 26-nt ssRNAs: the 3'-terminal 2'-*O*-modified 26-nt ssRNAs is not methylated by DmHen1 (Pimet). Specificity test of the DmHen-1 protein.

In the result on Figure 31, we detected methylation activity only with the positive control GST_DmHen1. Unfortunately, for the proteins of interest no signal was detected on the autoradiography. We varied multiple conditions, including the proteins concentration, the substrate concentration, the buffers and the times of incubation with different substrates for methylation (tRNAs, siRNA, miRNAs). In any of the tested experimental conditions no activity was detected (data not shown).

B. *In vitro* methylation assays with recombinant CG7009 and its potential partner

a. Identification of CG7009 partners

It has been described that many eukaryotic post-transcriptional tRNA modification enzymes require binding partners (Guy and Phizicky, 2014). Trm7, the predicted ortholog of CG7009, associates with two distinct partners for its catalytic activity (Guy et al., 2012). Trm7 requires binding with Trm732 to catalyze the addition of Nm_{32} and interaction with Trm734 for addition of Nm_{34} on its tRNA substrates: $tRNA^{Phe}$, $tRNA^{Trp}$ and $tRNA^{Leu\text{-}TAA}$. These interactions are also conserved in human cells and THADA and WDR6 are

the likely orthologs of Trm732 and Trm734 (Guy and Phizicky, 2015). THADA is able to supplement the growth defect of Δtrm732 mutants, despite the low sequence homology between THADA and Trm732 (~23% overall identity) (Guy and Phizicky, 2015). The amino acid sequences conservation of Trm7's interacting partners, Trm732 and Trm734 and of the THADA and WDR6 proteins, respectively, point to two potential *Drosophila* orthologs: CG15618, also known as DmTHADA, is the potential ortholog of Trm732 (29 % AA coverage, 26 % AA identity; E value: $5e^{-32}$) and THADA (17 % AA coverage, 35 % AA identity; E value: $3e^{-39}$). CG33172 is the potential ortholog of Trm734 (84 % AA coverage, 22 % AA identity; E value: $8e^{-17}$) and WDR6 (84 % AA coverage, 23 % AA identity; E value: $2e^{-26}$).

During my PhD I cloned and produced both of these proteins with N-terminal FLAG tag for CG33172 and GST N-terminal tag for CG15618 as described for CG7009 recombinant proteins. I will describe only the result obtained with CG33172, because we didn't perform any experiments yet with CG15618.

CG33172 encodes only one polypeptide of relatively big size (998 amino acids). The protein has an unknown function, it sequence contains WD40-repeat-containing domain. WD40 proteins are characterized to function as platforms for the assembly of multi-protein complexes. The predicted molecular function of CG33172 is RNA binding due to structural similarity with WDR6. This was very encouraging for us, because CG7009 does not possess a proper RNA binding domain. I thus cloned and produced CG33172 in the aim to perform *in vitro* methylation assays with CG7009 in the presence of its predicted partner. I managed to successfully clone and produce FLAG_CG33172 in bacteria (Methods).

We then first tested if we can co-immunoprecipitate (co-IP) the protein FLAG_CG33172 with GST_CG7009.

b. **Co-immunoprecipitation of GST_CG7009 and FLAG_CG33172 co-expressed in bacteria**

In a first experiment GST_CG7009 and its putative protein partner FLAG_CG33172 were synthesized separately in *E. coli*, purified by affinity chromatography and then incubated together before co-IP assays. Different incubation times and quantities of the incubated proteins were tested.

These first co-IPs were not conclusive, as no bands were detected at the expected sizes (not shown).

It is also possible that during the incubation one of the two potential partners gets degraded before the formation of a stable complex. Overall, this negative result suggested that, if there is an interaction between the two proteins, the complex is not sufficiently stable for the interaction to be detected in our co-IP experimental conditions.

In an attempt to stabilize the putative protein complex between GST_CG7009 and FLAG_CG33172, we co-expressed them in *E. coli*. A co-expression of GST alone with FLAG_CG33172 served as control. Then a co-IP of GST_CG7009 and FLAG_CG33172 was assayed by a FLAG-IP, followed by an anti-GST WB (Figure 32A). An anti-FLAG WB served as quality control of the FLAG-IP efficiency (Figure 32B).

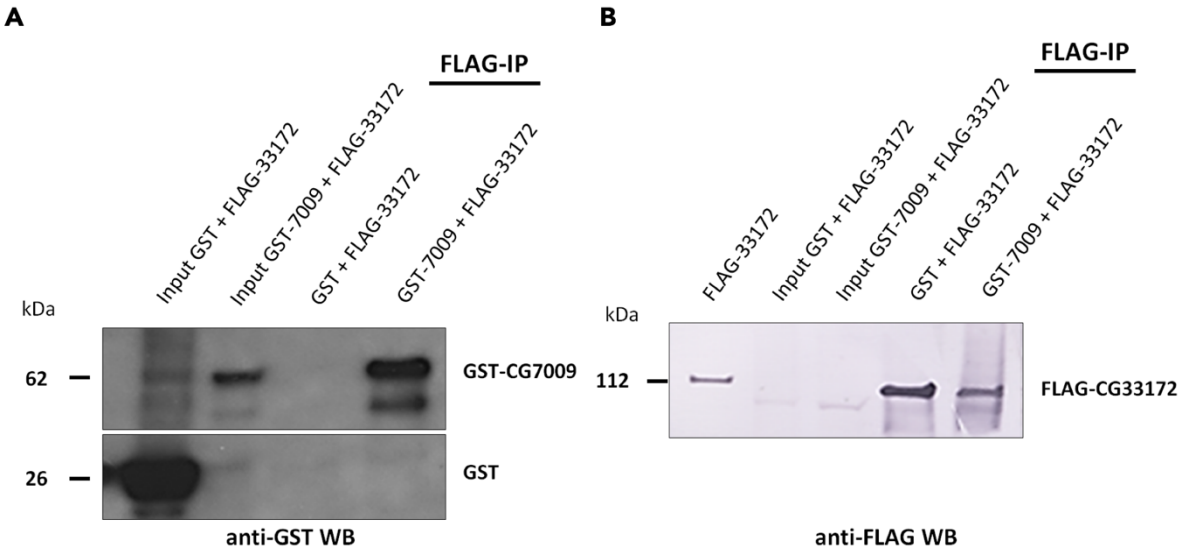

Figure 32. Co-immunoprecipitation of tagged CG7009 and CG33172 recombinant proteins co-expressed in bacteria. A. Upper panel Anti-GST WB on co-expressed GST_CG7009 and FLAG_CG33172, purified by FLAG-IP; **Lower panel** Anti-GST WB on co-expressed GST and FLAG_CG33172, purified by FLAG-IP. **B.** Anti-FLAG WB on co-expressed GST and FLAG_CG33172, purified by FLAG-IP (lanes 2 and 4); and co-expressed GST_CG7009 and FLAG_CG33172, purified by FLAG-IP (lanes 3 and 5); 1 µg recombinant FLAG_CG33172 was loaded as size control (lane 1). Inputs correspond to 5-10% of 10µg of protein eluates. The expected proteins sizes are GST (26 kDa), GST_CG7009 (62 kDa), FLAG_CG33172 (112 kDa); WB – Western blot; kDa – kilo Daltons.

In the control WB (Figure 32B), we detect FLAG_CG33172 in the purification from its co-expression with GST (input lane 2 and IP lane 4), as well as from its co-expression with GST_CG7009 (input lane 3 and IP lane 5). This demonstrated that the FLAG-IP was successful. The anti-GST WB that followed the FLAG_CG33172 IP (Figure 32A) demonstrated that the GST alone (lane 3, lower panel) does not co-IP with FLAG_CG33172, as no signal is detected. This also demonstrates that the anti-FLAG antibody used

in the IP does not interact with the GST tag. On the upper panel of Figure 32A a band at the expected size of GST_CG7009 was detected in the input, demonstrating that the protein was efficiently co-expressed with CG33172 (Figure 32A lane 2). Most importantly, a band corresponding to GST_CG7009 was also detected in the eluates of the immunoprecipitation of FLAG_CG33172 (Figure 32A lane 4). This showed that CG7009 co-immunoprecipitated with CG33172, suggesting that the two predicted partners are part of a common protein complex. However, we cannot conclude whether the interaction is direct or mediated by other binding partners.

The result of this co-IP of CG7009 and CG33172 suggests that CG33172 is an interacting partner of CG7009. However, the complex does not seem to be very stable, because both the GST and GST_CG7009 seem to be degraded, as smaller bands are detected (Figure 32A).

c. Methylation assays with the CG7009-CG33172 complex

In an attempt to optimize the *in vitro* methylation assay with recombinant CG7009 proteins, we performed the methylation reaction with the co-IP purified complexes of recombinant CG7009-CG33172. We reasoned that it is possible that CG7009 functions similarly to Trm7, which requires binding with Trm734 in order to catalyze Gm_{34} on its substrate tRNAs (Guy and Phizicky, 2015; Guy et al., 2012). If this is the case, supplementing CG7009 with CG33172 could be sufficient to recover the putative methylase activity of CG7009 *in vitro*. However, no signal was detected on the autoradiography of the gels in the results of the *in vitro* methylation assays carried out with the purified CG7009-CG33172 complex (data not shown). This could mean that either the presence of CG33172 is not sufficient to restore CG7009 enzymatic activity *in vitro,* as the protein could be part of a bigger complex, or that the protein was not stable enough (as seen from its purification) and is degraded during the *in vitro* methylation assay. It is important to note that, to our knowledge, no successful *in vitro* activity assays of recombinant TRM7 family proteins have been reported so far, suggesting that obtaining a functional methylase complex *in vitro* is more challenging that initially expected.

C. *In vitro* methylation assays with immunopurified from S2 cells tagged CG7009 and CG5220 proteins

As a consequence of the negative results of the *in vitro* methylation assays with the recombinant proteins, as well as with the *in vitro* purified GST_CG7009-FLAG_CG33172 complex, we decided to try another strategy.

We purified endogenous interacting partners of CG7009 and CG5220 by IPs, in the hope of recovering all the necessary proteins for their methylation activity. We then performed *in vitro* methylation assays with the co-IPed complexes.

For this purpose, with the help of Dilyana Dimitrova, (a second year PhD student), we cloned and expressed tagged versions of CG7009 in S2 cells of *Drosophila*. CG7009 was cloned using the Gateway system vectors pAWF and pMT-DEST48, allowing the production of FLAG_CG7009 or CG7009_V5_His, respectively (Methods). CG5220 expression clone, allowing the production of CG5220_FLAG_HA from a pMK33-CFH-BD was publicly available (Methods).

The CG7009 and CG5220 endogenous interacting complexes were then co-immunoprecipitated from S2 cell lysates and used in *in vitro* methylation assays on *in vitro* transcribed $tRNA^{Phe}$ (see Methods). Unfortunately, the results on the autoradiography remained negative (data not shown). We thus considered other strategies in order to decipher what the molecular function of CG7009 and CG5220 are.

4. Characterization of the molecular function of CG7009 by RTLP

We decided to try to determine whether CG7009 and CG5220 are involved in the methylation of their predicated target $tRNA^{Phe}$ by analyzing the presence of 2'-O-methylated nucleotides (Nm) in the anti-codon (AC) loop of $tRNA^{Phe}$ of *Drosophila*. As $tRNA^{Phe}$ is a conserved target of Trm7, the putative orthologs of CG7009 in yeast (Pintard et al., 2002) and of FTSJ1, the putative orthologs of CG7009 in human (Guy et al., 2015), we tried to detect the presence of a Nm on $tRNA^{Phe}$, using Reverse Transcription at Low dNTP concentration followed by PCR (RTL-P) (Dong et al., 2012). These experiments were performed with the help of Cyrinne Achour, a Master 2 student that worked under my supervision.

In this method, the property of the reverse transcriptase to pause and eventually detach from its template when it encounters RNA modifications at low [dNTPs] concentration is used (Figure 33A). 2'-O-methylation (Nm) is one of the RNA nucleotide modifications known to induce such pausing of the reverse transcriptase. We thus designed specific primers (F_U, and R, Figure 33B) in order to investigate by RTL-P if there will be a decrease in the predicted products of CG7009 and/or CG5220 catalytic activity: Nm_{32} and Nm_{34} of $tRNA^{Phe}$ (marked in red in the depicted tRNA of Figure 33B) in the corresponding mutants. In a first experiment, RTL-P was performed on *in vitro* transcribed $tRNA^{Phe}$ (unmodified tRNA template) in order to determine if the efficiency of cDNA synthesis was similar in the two different dNTPs concentrations (low and high, Figure 33C).

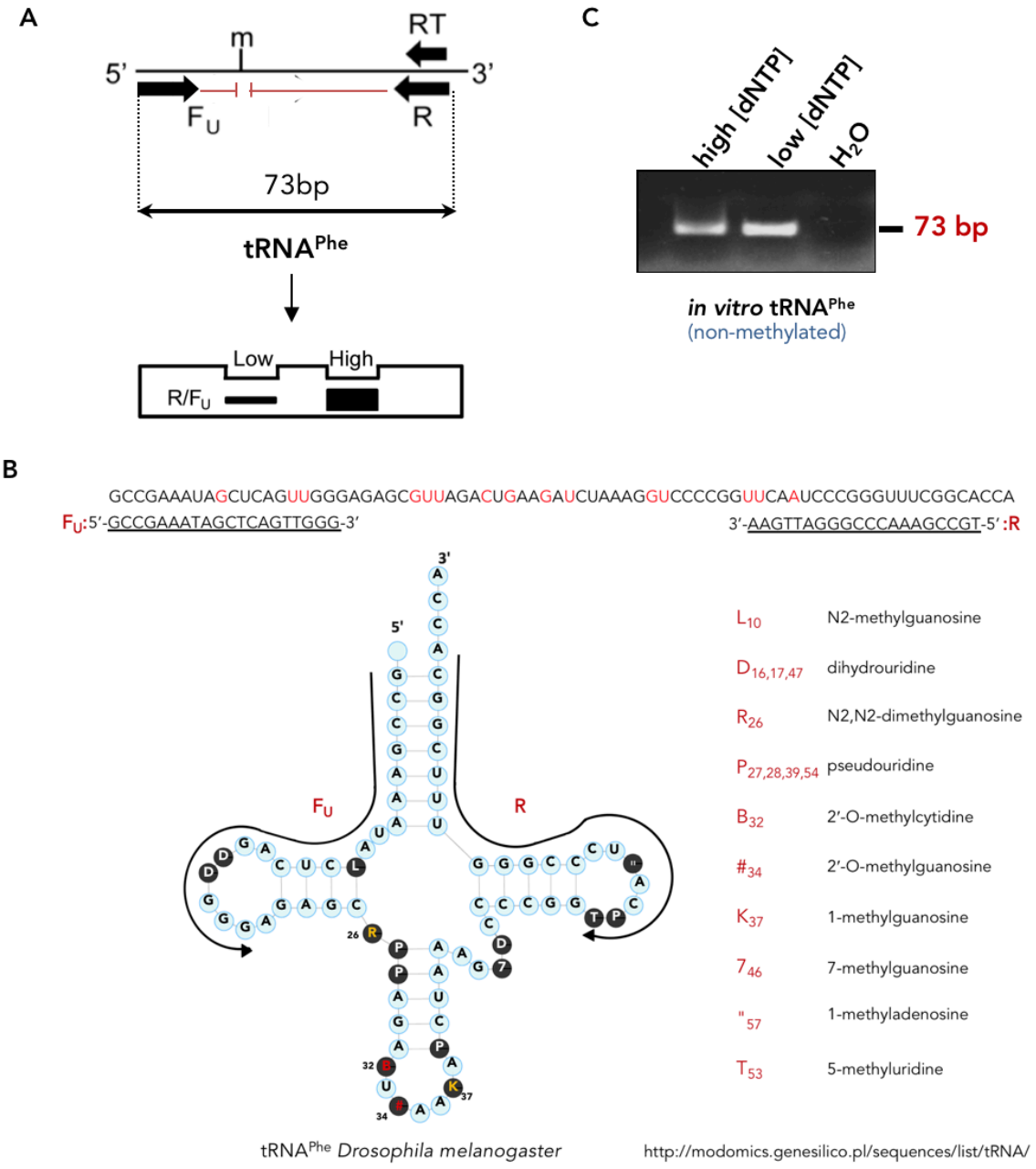

Figure 33. RTLP on *in vitro* transcribed tRNA^{Phe}. A. Rationale of the RTL-P experiment. In low [dNTPs] concentration the reverse transcriptase is blocked by the presence of Nm on the RNA template, yielding lower cDNA synthesis and lower amounts of PCR products after amplification on the obtained cDNA. **B.** Schematic representation of tRNA^{Phe} of *Drosophila*. The position of the primers used in the RT and the PCR (F_U and R) are indicated. Primer R was used both for the RT and PCR. The two Nm present on tRNA^{Phe} Cm_{32} (B, according to the Modomics nomenclature) and Gm_{34} (# according to the Modomics nomenclature) are indicated in red. Other modifications in the amplified region, known to be able to block the RT at low [dNTPs] are $m^2{}_2G_{26}$ and m^1G_{37}, represented in yellow. **C.** RTL-P on *in vitro* transcribed tRNA^{Phe}. tRNA^{Phe} (73 bp) was *in vitro* transcribed using with chimeric forward primer specific to tRNA^{Phe} and fused to the T7 bacteriophage promoter (Methods). Equal volumes of the RTL-P PCR products were loaded on 2.2% agarose TAE gel electrophoresis, stained with ethidium bromide. High [dNTP] concentration corresponds to 1 mM; low [dNTP] concentration corresponds to 1 µM. bp: base pairs, dNTP: deoxyribonucleotide triphosphate, Nm: 2'-O-methylation.

Figure 33C shows that the efficiency of the reverse transcriptase on *in vitro* transcribed $tRNA^{Phe}$ is, as expected, not affected by the low [dNTPs] condition (Figure 33C, lane 2).

We then investigated possible modifications on $tRNA^{Phe}$ by comparing the low and high [dNTPs] conditions in RTL-P experiments performed on total RNAs purified from Def3340 / TM3, Sb heterozygous individuals (Figure 34, lanes 3 and 4) *versus* $CG7009^{e02001}$ / Def3340 hetero-allelic mutants (Figure 34, lanes 8 and 9). The Rp49 mRNA served as an endogenous control of the experiment, as its transcript levels are constant between the tested genotypes and because the efficiency of its detection is not affected by high and low [dNTPs] concentrations (Figure 34, lanes 1-2 and 6-7).

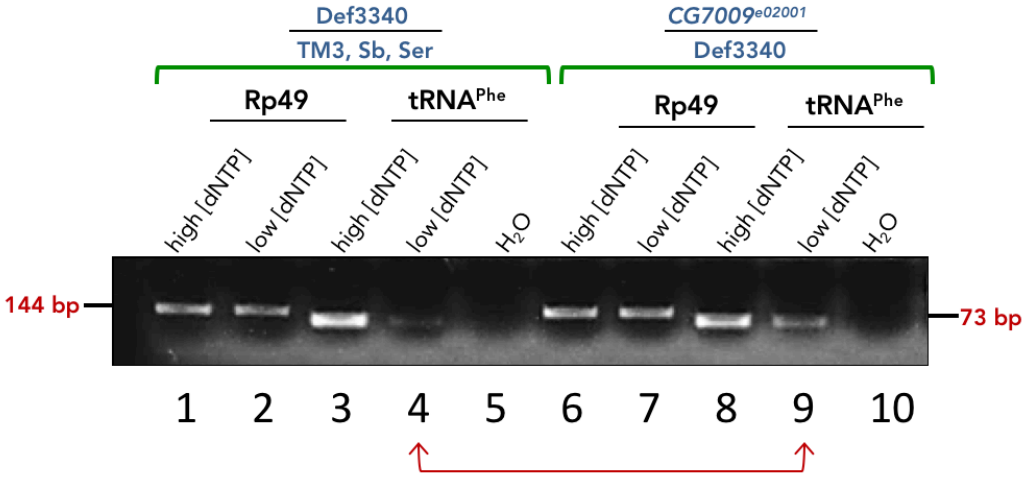

Figure 34. RTL-P detection of modified nucleotides in tRNA-Phe^{GAA} of *Drosophila*. RTL-P on total RNA from heteroallelic *CG7009* mutants (18008/def3340) and heterozygous flies (def3340/TbSb) with primers specific for $tRNA^{Phe}$ (73 bp). Rp49 (144 bp) was used as an endogenous control of the RT and PCR steps of the RTL-P (Methods). The high concentration of dNTPs is 1 mM, the low concentration is 1 µM. Equal volume of PCR products were loaded to a 2.2% agarose gel, stained with ethidium bromide. This experiment was repeated 3 times.

In heterozygous flies, lower amounts of PCR products were detected in low compared to high [dNTPs] concentration (Figure 34, compare lanes 3 and 4). This indicated that the reverse transcriptase was blocked at one, or several, modified nucleotide(s) in the heterozygotes when the concentration of [dNTPs] was low, as expected. Rp49 amplification from the same flies showed no difference between low and high [dNTPs] concentration (Figure 34, lanes 1 and 2). Lower amount of PCR products was also detected in low compared to high [dNTPs] concentration when $tRNA^{Phe}$ extracted from trans-heterozygous mutants was amplified (lanes 8 and 9). However, this decrease in the obtained PCR product in low [dNTPs]

concentration was less important in the mutants than in the heterozygous control flies (compare lanes 3 and 4 with lanes 8 and 9).

In addition, the PCR product obtained in the mutant for the low [dNTPs] condition is notably stronger than the product obtained in the heterozygotes (Figure 34, lanes 4 and 9, red double arrow). These results suggest that lower amounts of RNA modifications on $tRNA^{Phe}$ are detected by RTL-P in the absence of CG7009 when compared to heterozygous flies. Similar amounts of the Rp49 PCR product were detected in *CG7009* mutants, demonstrating that the experiment functioned in both [dNTPs] conditions (Figure 34, lanes 6 and 7).

The RTL-P result indirectly provided us with the first clue that the function of CG7009 is linked to deposition of a modification on $tRNA^{Phe}$, as predicted from its homology with Trm7 and FTSJ1. However, $tRNA^{Phe}$ bears multiple other modifications and the RTL-P cannot prove with certainty that CG7009 is specifically catalyzing a Nm deposition, as predicted, and not another modification, known to block the reverse transcription at low [dNTPs] and present in the AC loop of the tested tRNA.

However, in *CG7009* mutants the amplification of $tRNA^{Phe}$ was not equally efficient in low *versus* high [dNTPs] conditions, as seen for the *in vitro* transcribed $tRNA^{Phe}$ (Figure 34). This could be due to some remaining modification(s) on $tRNA^{Phe}$ in *CG7009* mutants, capable to block the reverse transcription and provoking a decrease in the amount of amplified $tRNA^{Phe}$ at low [dNTPs] in comparison with high [dNTPs]. Such other modifications in the amplified region of $tRNA^{Phe}$ are m^1G_{37} and $m^2{}_2G_{26}$ (marked in yellow on the $tRNA^{Phe}$ scheme on Figure 33B and (Motorin et al., 2007). It is also possible that CG5220, the second predicted ortholog of Trm7 and FTSJ1, is capable to partially modify $tRNA^{Phe}$ in *CG7009* mutants. Finally, it is possible that CG5220 and CG7009 have a redundant function and target the same nucleotide positions: when one of the putative methyltransferases is mutated the other could still be functional and perform its activity, hence the blocking of the reverse transcriptase.

5. Characterization of the molecular function of CG7009 by LC-MS

The RTL-P result suggested that CG7009 impacts modifications on $tRNA^{Phe}$ but was inconclusive to tell us exactly which modification on the tRNA is occurring. We decided to identify the nature of the modification by Liquid Chromatography coupled to Mass Spectrometry (LC-MS) with the help of our collaborators, Dr. Damien Brégeon and Mrs. Catherine Goyenvalle from the IBPS team of Olivier Jean Jean, as well as Mr. Vincent Guérineau from the ICSN Mass Spectrometry platform at Gif sur Yvette.

LC-MS combines the physical separation by liquid chromatography with analysis by mass spectrometry. Modified or unmodified nucleosides can be identified because they have specific retention times in the chromatography column. After the separation of a given fraction, it is subjected to mass spectrometry analysis providing a quantitative, sensitive and precise method for RNA modifications identification and quantification. For this purpose, total RNA corresponding to a mix of 1000 females and 1000 males of *CG7009*$^{e02001\text{-}G10}$ / Def3340 mutants and *CG7009*^{e02001} / TM6, Tb, Sb heterozygotes (control) were extracted. Then Damien and Catherine performed a column-purification of the small RNAs (size < 200 nt), followed by a specific purification of $tRNA^{Phe}$ with the use of magnetic beads crosslinked to an anti-$tRNA^{Phe}$ DNA oligonucleotide. Purified $tRNA^{Phe}$ were then first digested to the nucleotide level and then dephosphorylated to the nucleoside level. After adjustment of their concentration the nucleosides were injected into a liquid chromatography column, coupled to a mass-spectrometry analyzer. We were particularly interested by the levels of 2'-O-methylation of Nm_{32} and Nm_{34}, the conserved target sites of Trm7 and FTSJ1 (Guy et al., 2015; Pintard et al., 2002).

As individual nucleosides are analyzed in LC-MS, the method cannot give us the exact position of the modification. However, all 8 $tRNA^{Phe}$ genes of *Drosophila* for encode tRNAs with a GAA anticodon (containing a Gm_{34}) and a Cm_{32} (http://modomics.genesilico.pl/ and http://gtrnadb.ucsc.edu/genomes/eukaryota/Dmela6/). Only Cm_{32} and Gm_{34} are present in the $tRNA^{Phe}$ sequence (marked in red in Figure 33B). This meant that if we detect a difference in the Gm or Cm levels between mutants and control flies with the LC-MS analysis, they will correspond to Gm_{34} and Cm_{32}, the two predicted target sites of CG7009.

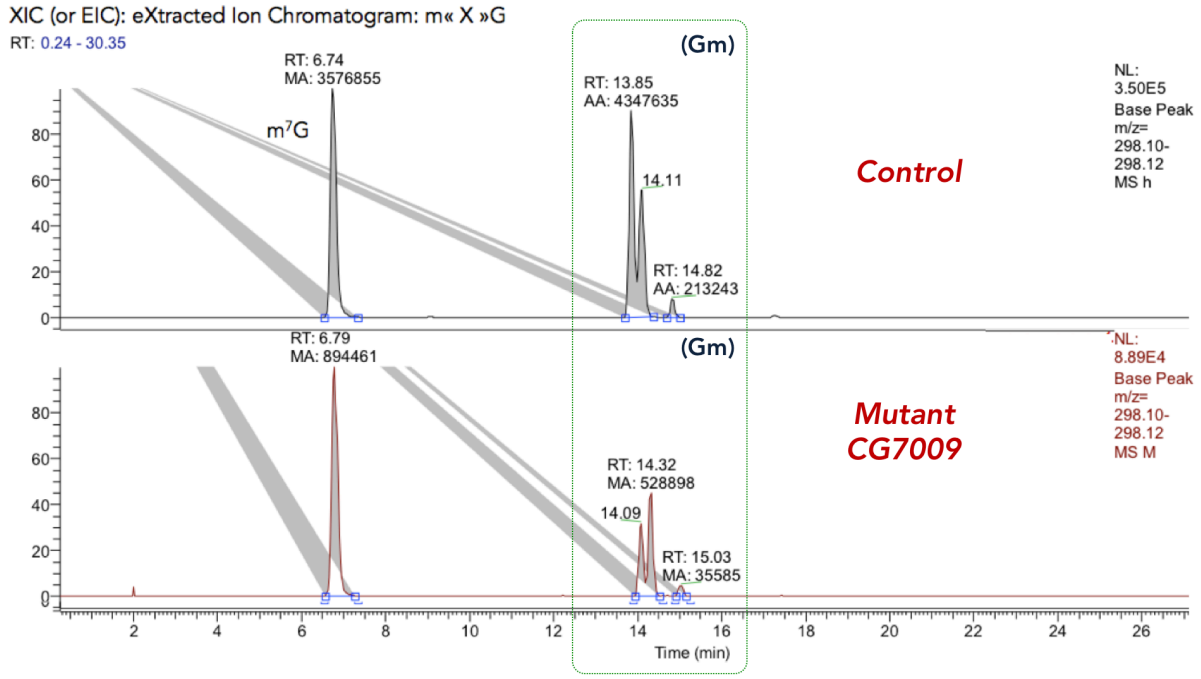

Figure 35. Gm_{34} formation in *D. melanogaster* $tRNA^{Phe}$ is performed by CG7009. LC-MS analysis of the Gm_{34} abundance in $tRNA^{Phe}$ extracted from heterozygous flies *CG7009*^{e02001} / TM6, Tb, Sb (upper panel) or trans-heterozygous individuals *CG7009*^{e02001}/ Def3340 (lower panel).

The LC-MS result on Figure 35 reports on the levels of 2'-O-methylguanosine (Gm, green panel) in $tRNA^{Phe}$ purified from heterozygous flies *CG7009*^{e02001} / TM6, Tb, Sb as control (upper panel) or in homozygous mutants *CG7009*$^{e02001\text{-}G10}$ / Def3340 (lower panel).

We can see that the peak corresponding to Gm is decreasing in the mutant. This result demonstrated that CG7009 is involved in Gm formation on $tRNA^{Phe}$ in *Drosophila*, and as there is only one Gm on this tRNA, it is most probably Gm_{34}. However, the peak did not disappear completely in the mutants. We propose two possibilities to explain this. (i) Either CG5220 has a redundant function with CG7009, partially compensating the loss of CG7009 activity. In order to verify this, we needed to perform the experiment in CG7009, CG5220 double mutants, whose establishment was ongoing by the moment we obtained this result. (ii) A second option was that during the purification of $tRNA^{Phe}$ some other Gm containing tRNA were co-purified and thus the observed remaining peak represents traces of contamination with other tRNA. It is very less probable that only one of the samples is contaminated by other tRNA(s) binding to the anti-$tRNA^{Phe}$ oligo and not the other. The result of the LC-MS strongly suggested that CG7009 is a tRNA methylase of $tRNA^{Phe}$ in flies. In order to validate our results, we decided to perform additional experiments.

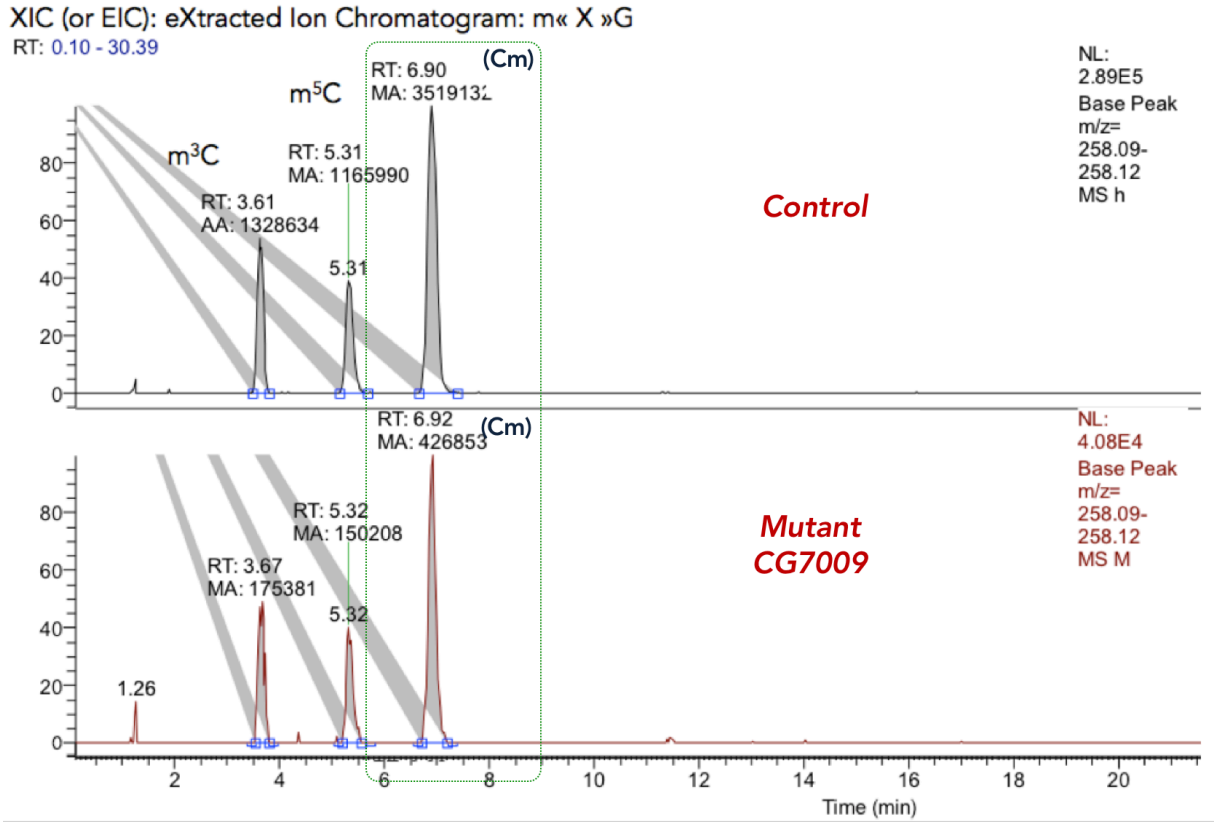

Figure 36. Cm_{32} formation in *D. melanogaster* $tRNA^{Phe}$ seems independent of CG7009. LC-MS analysis on the Cm_{32} abundance in $tRNA^{Phe}$ extracted from heterozygotes *CG7009*^{e02001} / TM6, Tb, Sb (upper panel) or trans-heterozygotes *CG7009*^{e02001}/ Def3340 (lower panel).

On Figure 36, the same analysis was performed for 2'-O-methyl cytidine (Cm) (green doted box). The peak corresponding to Cm is not changed between *CG7009*^{e02001} / TM6, Tb, Sb heterozygous individuals used as control (upper panel) and *CG7009*$^{e02001\text{-}G10}$ / Def3340 homozygous mutants (lower panel). This indicates that CG7009 is not involved in the deposition of Cm_{32} on $tRNA^{Phe}$ in *Drosophila*.

Unexpectedly, we also detect m^3C and m^5C, two modifications that have not been reported as existing on *Drosophila* $tRNA^{Phe}$. This observation is in support of a contamination during the purification of $tRNA^{Phe}$ or another stage of the experiment. We thus had to reproduce the experiment.

6. RNase digestion coupled to MALDI-TOF demonstrate that CG7009 methylates position G_{34} and that CG5220 methylates position C_{32} of $tRNA^{Phe}$

In the search of the molecular function of CG7009 we performed another experiment in collaboration with Dr. Damien Brégeon, Mrs. Catherine Goyenvalle and Mr. Vincent Guérineau: an RNase digestion using either RNase A or RNase T1, coupled to MALDI-TOF analysis, a well-established procedure in the mapping of RNA modification by mass spectrometry. The rationale of this experiment is similar to the one for LC-MS: different nucleoside-specific RNases, *i.e.* RNase T1 being guanosine specific, while RNase A is pyrimidine specific, will generate different products from the digestion of their RNA substrates, depending on the RNA modifications present (Machnicka et al., 2013), which are then analyzed by MS. A major difference with LC-MS is that m/z ratio of the generated RNA fragments are compared and not nucleosides. This allows to exclude the previously observed problems with LC-MS of co-analyzing the levels of contaminant nucleosides with the nucleosides of interest. Deducing the m/z ratio of the generated fragments after MALDI-TOF in this experiment also allows to precisely determine not only the nature, but also the position of the RNA modification, as the modification composition for $tRNA^{Phe}$ was available by the MODOMICs consortium.

We used the property of RNase A to cut after every C or U nucleotide unless they bear a methylation and the property of RNase T1 to digest RNA after every guanosine, unless it is methylated. Thus, for the anti-codon loop of $tRNA^{Phe}$ ($Gm_{34}AA$) and the adjacent Cm_{32}, the combination of these two RNases was a pertinent choice in order to attempt to decipher the exact role of CG7009 on the modification of $tRNA^{Phe}$.

For this experiment I extracted total RNA of approximately the same number of mix of males and females (1000 of each). RNAs smaller than 200 nt were then column purified followed by isolation of $tRNA^{Phe}$ with anti-$tRNA^{Phe}$ oligos as for the LC-MS. The modifications on $tRNA^{Phe}$ were then analyzed by MALDI-TOF MS analysis preceded by digestion either with RNase T1 (Figure 37), or with RNase A (Figure 38).

A. RNase T1 digestion coupled to MALDI-TOF

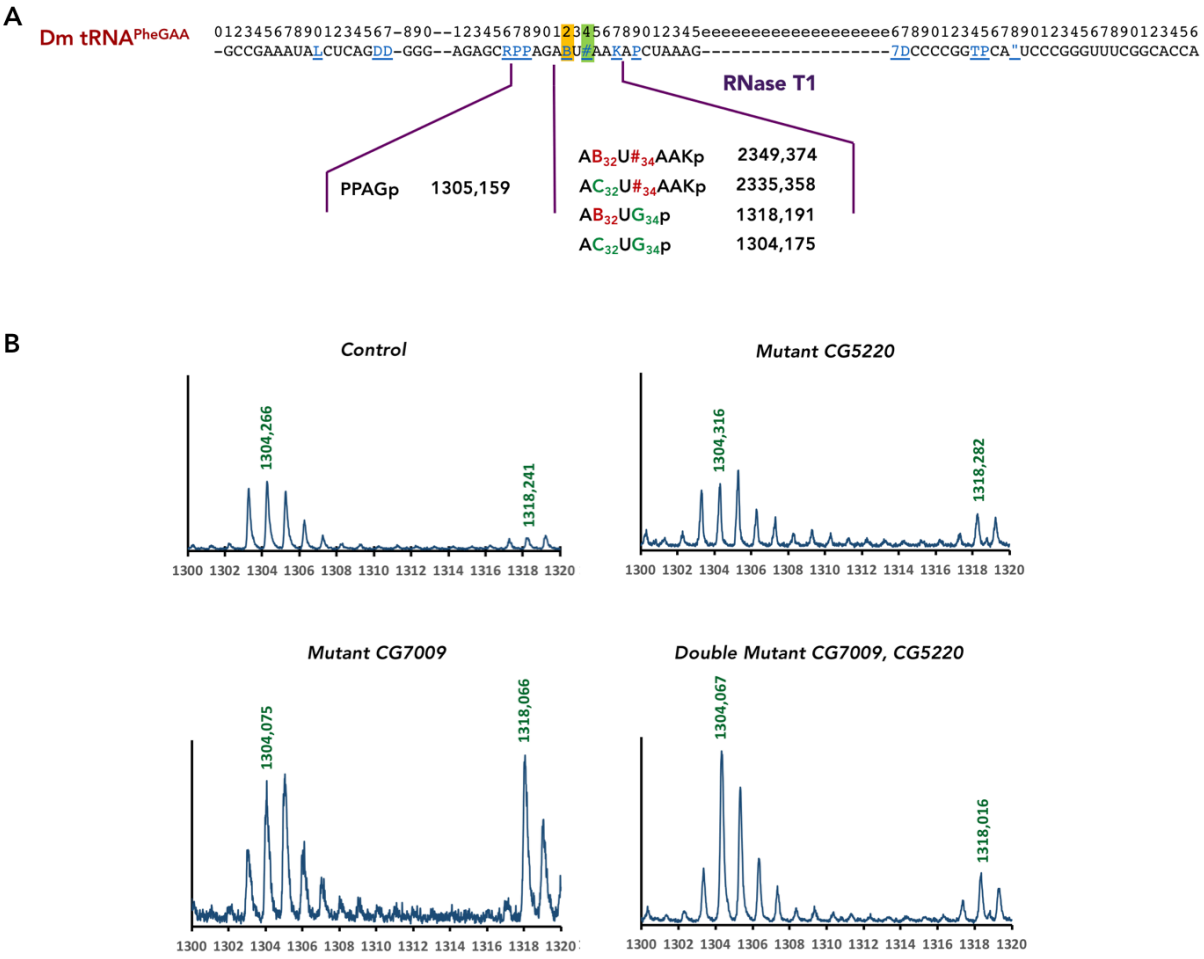

Figure 37. MALDI-TOF analysis of RNase T1 digested tRNA^{Phe}. A. Sequence of *Drosophila* melanogaster tRNA^{Phe} with m/z values of fragments containing 2'-O-methyl-guanylate and/or 2'-O-methyl-cytidilate. L is N2-methylguanosine, D is dihydrouridine, R is N2,N2-dimethylguanosine, P is pseudouridine, B is 2'-O-methylcytidine, # is 2'-O-methylguanosine, K is 1-methylguanosine, 7 is 7-methylguanosine, T is 5-methyluridine (ribothymidine) and " is 1-methyladenosine. **B.** MALDI-MS spectrum of fragments resulting from RNAse T1 digestion of tRNA^{Phe} originated from Control flies: *CG7009$^{e02001\text{-}G10}$* / TM6, Tb, Sb (*top left*); *CG5220* mutants: *CG5220^{K28A}* / *CG5220^{K28A}* (*top right*), *CG7009* mutants: *CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$* (*bottom left*) and double mutant *CG7009, CG5220*: *CG7009$^{e02001\text{-}G10}$*, *CG5220^{K28A}* / *CG7009$^{e02001\text{-}G10}$*, *CG5220^{K28A}* (*bottom right*). Relevant peaks are identified by their m/z values denoted in green.

Figure 37A, represents the sequence of $tRNA^{Phe}$ of *Drosophila*, reported on the MODOMICs project website (http://modomics.genesilico.pl/). The sizes in Daltons (Da) of the different combination of fragments expected to be obtained following $tRNA^{Phe}$ digestion with RNase T1 if the positions C_{32} and G_{34} bear (B_{32} and $\#_{34}$) or not (C_{32} and G_{34}) a Nm modification are indicated below the sequence.

A *PPAGp* fragment will always be obtained from the activity of RNase T1 after the N2,N2-dimethylguanosine at position 26 and the guanosine at position 30. The accumulation of this fragment is seen on Figure 37B, but it is not denoted because it is not linked directly to the modification status of Cm_{32} and Gm_{34} in the AC-loop of tRNA.

The fragments bigger than 2000 Da (between A_{31} and 1-methylguanosine: K_{37}), presented on Figure 37A, will be obtained if position G_{34} keeps its 2'-O methylation modification. These fragments are not shown in Figure 37B.

We can deduce the Gm_{34} modification status from fragment $AB_{32}UG_{34}p$ (1318, 191 Da), which will be obtained if Cm_{32} (B_{32}) remains modified but Gm_{34} ($\#_{34}$) is lost, giving an unmodified G_{34}. We can also deduce the Cm_{32} (B_{32}) modification status by analyzing the accumulation of fragment $AC_{32}UG_{34}p$ (1304,175 13 Da), which will increase if both C_{32} and G_{34} lose their Nm modification.

We analyzed $tRNA^{Phe}$ from *CG7009*^{e02001} / TM6, Tb, Sb heterozygous flies (Figure 37B, *top left*), as control, *CG5220*^{K28A} / *CG5220*^{K28A} (Figure 37, *top right*) and *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ (Figure 37, *bottom left*) homozygous single mutants, as well as *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} / *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} Rec1 double mutants (Figure 37, *bottom right*).

In Figure 37 we see that the peak at 1318 Da increases in *CG7009* mutants: $AB_{32}UG_{34}p$ (1318, 191 Da) (Figure 37, bottom left). This indicates that CG7009 is a Nm methylase of position G_{34} of $tRNA^{Phe}$ in wild-type flies.

The peak at 1304 Da increases in the double mutants: $AC_{32}UG_{34}p$ 1304,175 (Figure 37B, bottom right). It thus seems that *CG5220* is a methylase of position C_{32} of $tRNA^{Phe}$ in wild-type *Drosophila*.

B. RNase A1 digestion coupled to MALDI-TOF

Figure 38A represents the tRNA^{Phe} sequence according to MODOMICs and the expected fragments upon RNase A digestion depending on the modification status of C_{32}. Digestion with RNase A can give us more precise information on the modification of C_{32}, as this nuclease will cut directly after the cytosine if it is not methylated (giving rise to a smaller fragment: AGAC_{32}p – 1327,202 Da) and will cut at the next non-modified pyrimidine if C bears a Nm (B_{32}) (giving rise to a bigger fragment: AGAB_{32}Up - 1647,243 Da). As for the RNase T1 analysis, peak bigger than 2000 Da are not represented in Figure 38.

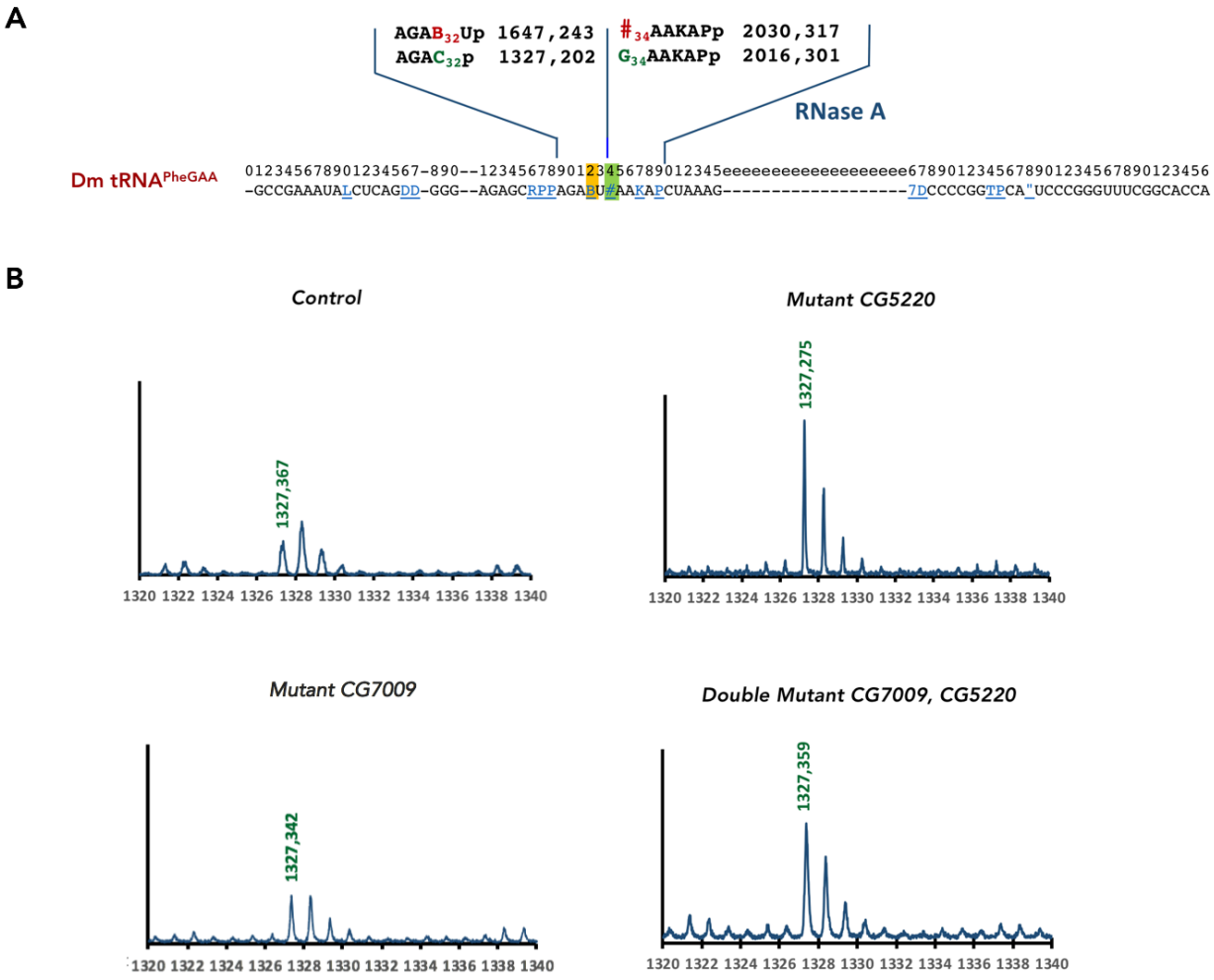

Figure 38. MALDI-TOF analysis of RNase A digested tRNA^{Phe}. A. Sequence of *Drosophila* melanogaster tRNA^{Phe} with m/z values of fragments containing 2'-O-methyl-guanylate and/or 2'-O-methyl-cytidilate. L is N2-methylguanosine, D is dihydrouridine, R is N2,N2-dimethylguanosine, P is pseudouridine, B is 2'-O-methylcytidine, # is 2'-O-methylguanosine, K is 1-methylguanosine, 7 is 7-methylguanosine, T is 5-methyluridine (ribothymidine) and " is 1-methyladenosine. **B.** MALDI-MS spectrum of fragments resulting from RNAse A digestion of tRNA^{Phe} originated from Control flies: *CG7009$^{e02001\text{-}G10}$* / TM6, Tb, Sb (*top left*); *CG5220* mutants: *CG5220^{K28A}* / *CG5220^{K28A}* (*top right*), *CG7009* mutants: *CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$* (*bottom left*) and double mutant *CG7009, CG5220*: *CG7009$^{e02001\text{-}G10}$*, *CG5220^{K28A}* / *CG7009$^{e02001\text{-}G10}$*, *CG5220^{K28A}* (*bottom right*). Relevant peaks are identified by their m/z values denoted in green.

Figure 38B shows that the peak at 1327 Da increases importantly in *CG5220*^{K28A} / *CG5220*^{K28A} mutants (Figure 38B, *top right*), a little in *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ (Figure 38B, *bottom left*) and moderately in the *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} / *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} Rec1 double mutants (Figure 38B, *bottom right*). We can thus conclude that CG5220 is a methylase of position 32 in tRNA^{Phe} of *Drosophila*.

In conclusion we managed to characterize the exact molecular function of two novel 2'-O-methyltransferases in flies in regards to tRNA^{Phe}. It seems that *Drosophila* has evolved 2 paralogues which specialize in the modification of distinct nucleotides in the AC-loop of tRNA^{Phe}. The result is interesting because both FTSJ1 in human, and Trm7 in yeast are capable to methylate the two positions 32 and 34 in their tRNA substrates.

7. RiboMeth Seq

To support the results from the MALDI-TOF analysis, as well as to get a wider overview on the function of CG7009 and CG5220 on tRNA modifications, we performed a RiboMethSeq analysis (Marchand et al., 2016) in collaboration with Pr. Yuri Motorin and Dr. Virginie Marchand. The method is based on the combination of alkaline RNA fragmentation with high-throughput sequencing, allowing simultaneous detection of multiple modifications sites from small amounts of biological samples. The phosphodiester bond at the 3′-end of Nm-containing nucleotide is more resistant to nucleophilic cleavage under alkaline conditions, than nucleotides containing a 2′-OH at their 3′-terminus.

It was shown that, when combining alkaline hydrolysis with random fragmentation of RNA, the RNA fragments starting or ending at the +1 nucleotide to a Nm modified residue were excluded from the sequencing libraries (Birkedal et al., 2015). Thus, by analyzing the "gaps" of the mapped reads ends profiles, the position adjacent of Nm residues can be deduced. In their work, Marchand and colleagues report a method for transcriptome-wide Nm detection based on alkaline hydrolysis of randomly fragmented RNA samples, adapted for an Illumina high-throughput sequencing platform (Marchand et al., 2016). Figure 39 provides an overview of the method.

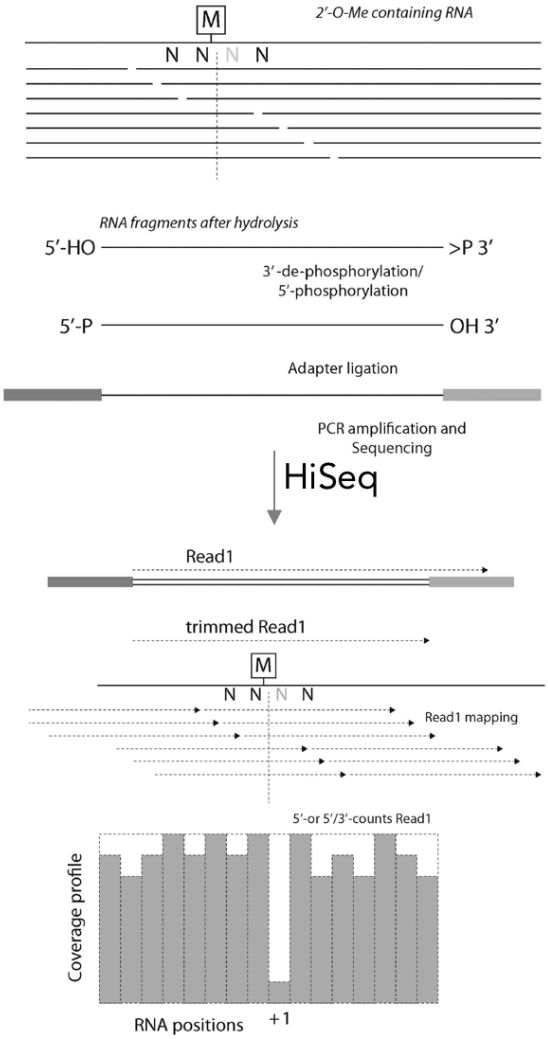

Figure 39. General overview of the RiboMethSeq protocol. RNAs containing 2′-O-Methylation residues (Nm) are randomly fragmented, 5′- and 3′-ends are repaired and adapters are ligated to both extremities. After amplification and barcoding, amplicons are subjected to Illumina sequencing. HiSeq sequencing is generally performed in a single-read mode. Orientations and mapping of reads are indicated. The 2′-O-Me (Nm) residues protect the 3′-adjacent phosphodiester bond from cleavage, generating a typical gap in 5′-/3′-ends coverage profile. The + 1 position indicates the +1 positions regarding the 2′-O-Me residues (0 position correspond to the 2′-O-Me residue). Figure adapted from (Marchand et al., 2016).

The results present on Figure 40 are very recent. After preparation of the desired RNAs in our lab, Pr. Motorin and his co-workers have generated the presented results following the RiboMethSeq protocol. With their help, we analyzed the data in the very last week of my PhD manuscript writing. For this reason, I will only present several major conclusions issued from this experiment, as we have not yet had the time to do a deeper analysis of the data.

In the Illumina-based RiboMethSeq method (Marchand et al., 2016), the presence or absence of Nm is represented as 5′-/3′-ends coverage reads profiles (Figure 39). The 5′-/3′-ends coverage profiles are determined by simultaneous 5′- and 3′-ends reads counting. As the 2′-O-Me (Nm) residues protect the 3'-adjacent phosphodiester bond from cleavage, a gap in the 5′-/3′-ends coverage profile means that the beginning and the ending of reads at a certain nucleotide position is under-represented in comparison to the frequency of reads beginning and ending at adjacent nucleotide positions.

Figure 40 shows the analysis of Nm presence at position 34 for $tRNA^{Phe}$ (Gm_{34}). The reasoning is the following: when Gm_{34} is present (*i.e.*, in the control, Figure 40A), the reads beginning and ending at the 3′ adjacent position 35 will be under-represented, in regards to their environment. This will make the reads beginning and ending at Gm_{34} to be over-represented in comparison to position 35. The under-represented reads coverage at position 35 is thus giving a gap for this position on the 5′-/3′-ends coverage reads profiles.

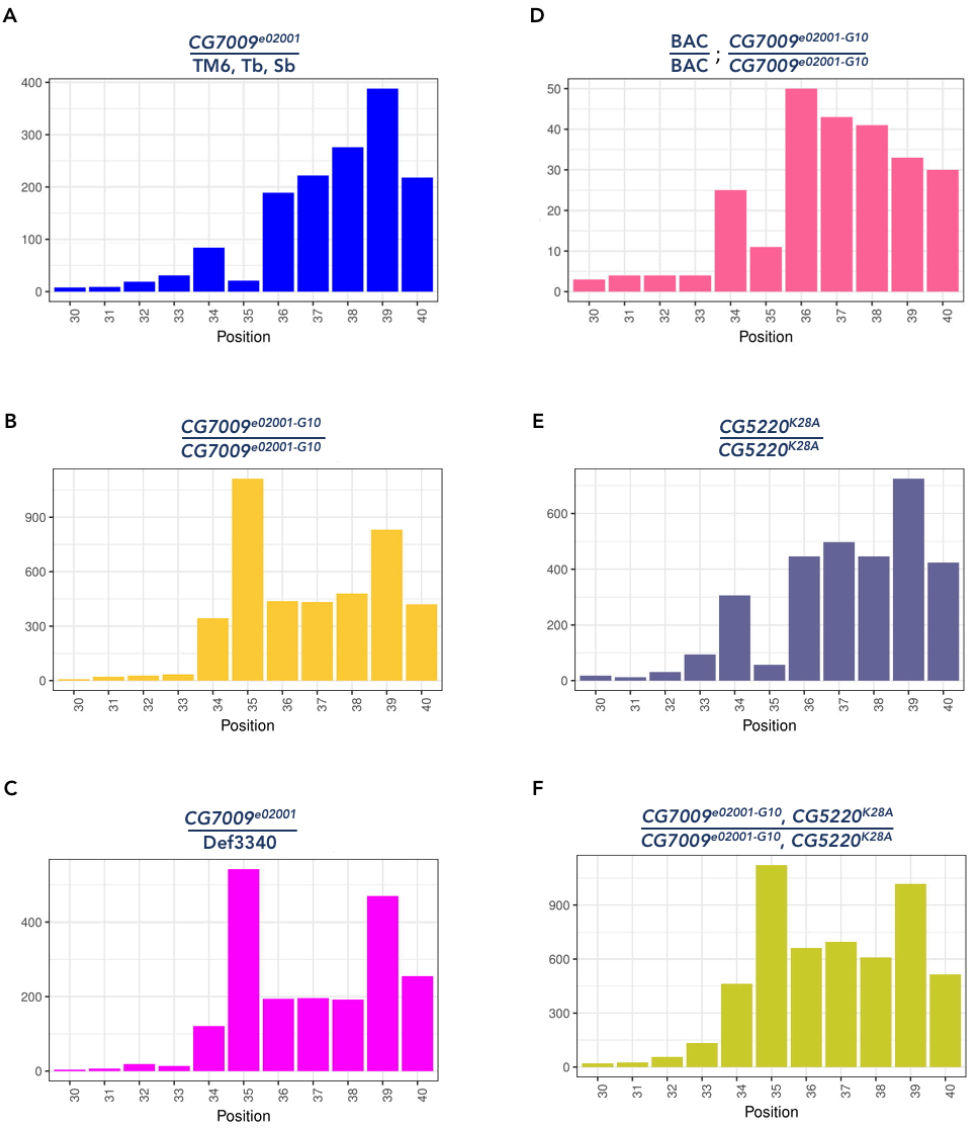

Figure 40. 2'-O-Methylation analysis using RiboMethSeq on tRNA^{Phe} of Drosophila. RiboMethSeq protocol on total RNAs, extracted from whole flies was performed on flies with the indicated genotypes. **A to F:** Fragmentation profiles for regions of *Drosophila* tRNA^{Phe} are shown in different genotypes, indicated on the top of each panels. Here the +1 analyzed position is position 35. The position of interest is position 34. If Nm is present on position 34, a protection will be seen at position 35. If no Nm is present at position 34, no protection will be seen on 35. The ordinates on the graphs represent the total number of reads. Random RNA fragmentation was performed with bicarbonate buffer (OH−) and sequencing on HiSeq-1000 in single read of 50 nt.

Here, when centering the analysis on position 35, "Gaps" are detected in the heterozygous control flies (Figure 40A), in *CG5220* mutants (Figure 40E) and for the genetic rescue of *CG7009* (Figure 40D). The presence of Gm_{34} in *CG5220* mutants confirms the results previously obtained with the MALDI-TOF experiment, indicating that CG5220 is not involved in the catalysis of Gm_{34} on tRNA^{Phe} of *Drosophila*. On the contrary, in both *CG7009* homozygous (Figure 40B) and trans-heterozygous mutants (Figure 40C), as

well as in the *CG7009*, *CG5220* double mutant (Figure 40F) the 5′-/3′-ends coverage profiles of reads beginning and ending at position 35 (position +1 to the expected Gm_{34}) are over-represented in comparison to the adjacent reads 5′-/3′-ends coverage positions, demonstrating that in the absence of CG7009, G_{34} position is not 2′O-methylated, because the "gap" in number of reads beginning or ending at its 3′- adjacent position +1 (35) disappear.

This result is a strong second indication that CG7009 is the 2′-O-methylase of G_{34} in $tRNA^{Phe}$ in *Drosophila*. Importantly, mutants for CG7009 rescued by the BAC transgene seem to restore the Gm_{34} formation (Figure 40D), but a more detailed analysis will be required to confirm this important control.

The same RiboMethSeq read profiles analysis was made for position 33 of $tRNA^{Phe}$ (+1 position to the expected Cm_{32}, data not shown). The results clearly indicated that CG5520, and not CG7009, is responsible for Cm methylation at position 32 in $tRNA^{Phe}$. Importantly this new result confirmed our MALDI-TOF experiments as well as the $tRNA^{Phe}$ Cm_{32} methylation reported by MODOMICs.

Our preliminary analysis was also performed on $tRNA^{Trp}$, an expected target of CG7009 and CG5220. In flies, 8 genes encode $tRNA^{Trp}$ and all of them have a C at position 32. The results clearly indicated that CG5520 is responsible for Cm methylation at position 32 on $tRNA^{Trp}$ as expected (data not shown). However, for the Cm_{34} status of $tRNA^{Trp\text{-}CCA}$, the result is not clearly interpretable. This preliminary analysis seems to indicate that neither CG5220, nor CG7009 are catalyzing Cm_{34} $tRNA^{Trp\text{-}CCA}$.

Very unexpectedly, in regards to the known targets of TRM7 family members throughout evolution (Phe, Leu-TAA and Trp), our analysis showed that CG5220 is also able to methylate $tRNA^{Gln}$ at position 32 (not shown). In *Drosophila* 8 genes are encoding for the tRNA isoacceptor $tRNA^{Gln\text{-}CTG}$ and 4 genes are encoding $tRNA^{Gln\text{-}TTG}$ (ongoing analysis, data not shown). All of those genes encode a C at position 32 of the corresponding tRNAs. Thus, the results of the RiboMethSeq indicate that CG5220 is most probably methylating C_{32} on both tRNA isoacceptors.

Finally, the same unexpected observation was also made for $tRNA^{Glu}$ (ongoing analysis, data not shown). There are 19 genes encoding $tRNA^{Glu}$ with two different anti-codons: $tRNA^{Glu\text{-}TTC}$ and $tRNA^{Glu\text{-}CTC}$. All of them have a C at position 32. CG5220 is also able to methylate $tRNA^{Glu}$ at position 32. Importantly for the validity of our analysis, the modifications composition of $tRNA^{Glu\text{-}TTC}$ was already determined by mass spectrometry analysis, as reported by the MODOMICs consortium. The reported sequence of $tRNA^{Glu\text{-}TTC}$ revealed that, similarly to our analysis, C_{32} of $tRNA^{Glu\text{-}TTC}$ is 2′-O-methylated. This is a crucial

information, as position 32 on this specific $tRNA^{Glu}$ isoacceptor, as well as on $tRNA^{Gln}$ isoacceptors, are, to our knowledge, not yet reported to be catalyzed by a member of the conserved methyltransferase family TRM7.

Altogether, the analysis on the function of CG5220 in flies seems to reveal a specificity of this protein for ribonucleotides C at position 32, inside the AC loop on several tRNAs.

CG7009 is to date involved in G methylation inside the AC loop, at the wobble position 34. Our study on the methylation activity of CG5220 and CG7009 in flies tRNAs is, to our knowledge, the first one that identified and characterized in details the molecular function of those two 2'-O RNA methyltransferases.

Chapter III: Involvement of the *CG7009* and *CG5220* genes in the sncRNA pathways

1. Involvement in the miRNA pathway

A. Confirmation of the genomic screen: CG7009 is an actor of the Ago-2 miRNA pathway

Our team had discovered the *CG7009* gene prior to my arrival in a genome wide screen searching for factors involved in the miRNA and/or the Ago-2-mediated small RNA gene silencing (Figure 41). As a reminder (see Introduction), in flies, a subset of miRNAs presenting higher duplex complementarity is loaded into Ago-2, the siRNA pathway effector protein (Figure 41B).

The sensor used for this screen, presented in the Introduction, was automiG (Figure 41A). This biosensor is expressing the GFP mRNA and two miRNAs (miG1 and miG2) perfectly complementary to the GFP mRNA (Figure 41A and (Carre et al., 2013)). AutomiG was used in an RNAi screen in *Drosophila* S2R+ cells in order to uncover novel factors of the Ago-2-dependent miRNA pathway (Figure 41C). AutomiG is a specific biosensor for the miRNA-Ago-2 dependent pathway, as the silencing of the GFP requires miRNA biogenesis factors like Drosha and Dicer1 (Dcr-1), as well as Ago-2, and not Ago-1 for the GFP repression (Figure 41D).

In the WB on Figure 41D, Ftz and luciferase were used as negative controls and Ago-2, Dcr-1 and Drosha as positive controls, for their function in the self-silencing of automiG (Carre et al., 2013). It is important to remember, that when the self-silencing automiG sensor detects a defect in the miRNA Ago-2-dependent silencing pathway, the GFP signal increases.

The WB on Figure 41D showed that CG7009 loss-of-function leads to the automiG's auto-repression of the GFP, and thus uncovered, for the first time, CG7009 as a novel factor of the Ago-2-dependent miRNA silencing pathway.

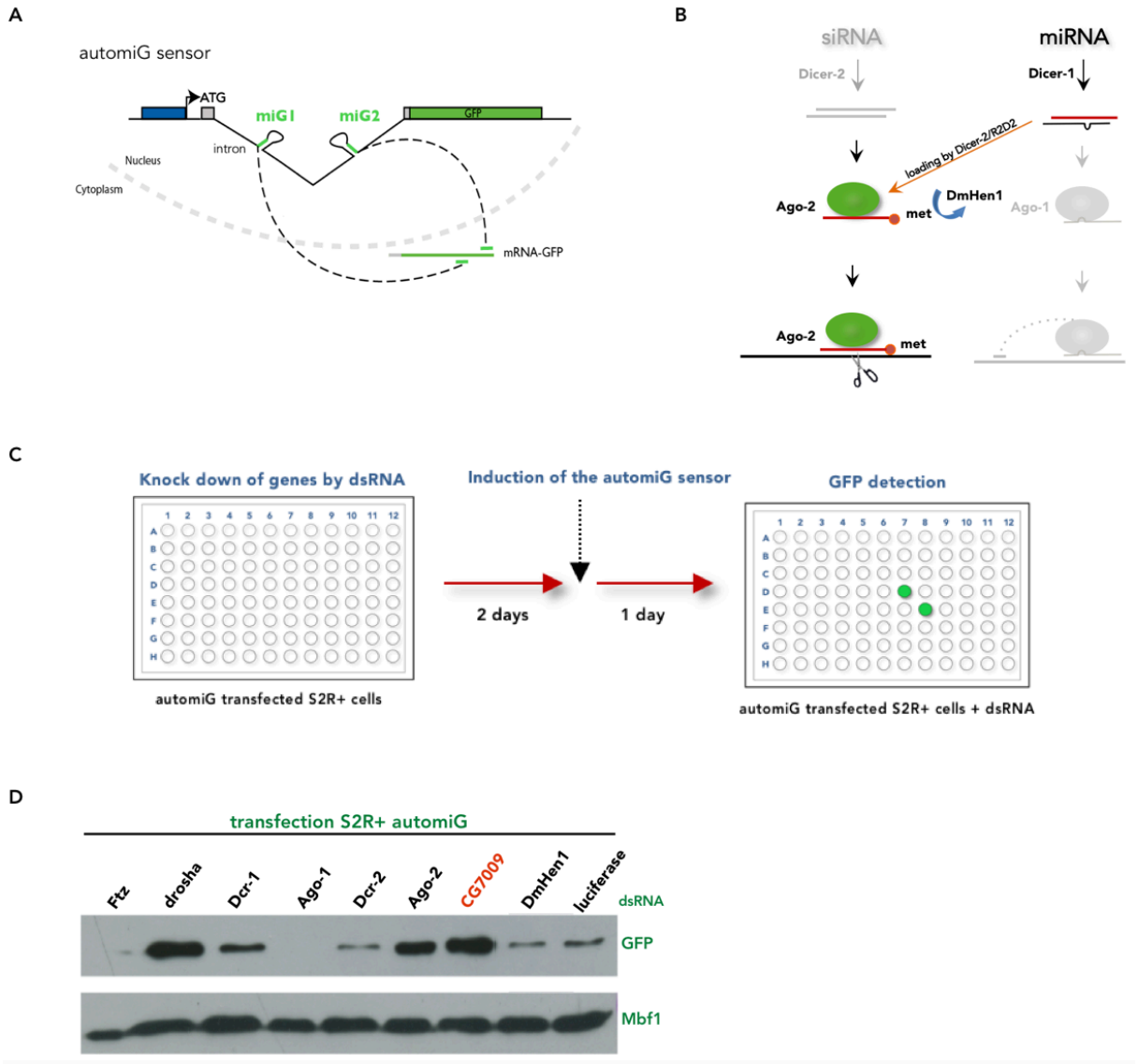

Figure 41. *CG7009* is involved in miRNA-Ago-2 silencing. A. Schematic representation of the automiG system. The two miRNAs programmed against the GFP and the GFP CDS are expressed under the control of the same copper inducible promoter (pMT). **B.** Schematic representation of the overlap between the si- and the miRNA pathways in flies. Structural features of the miRNA duplexes trigger their loading in the Ago-2 protein, where they get 2'-O-methylated at the 3'-end by DmHen1 in difference to Ago-1 loaded miRNAs. Dcr-2/R2D2 is only required for the loading of miRNAs into Ago-2 and not for their biogenesis. Once loaded into Ago-2, the miRNAs silence their targets by endonuclease cleavage, accomplished by their partner Ago-2 **C.** Schematic representation of the RNAi screen in S2 cells. Cells, stably transfected with the automiG sensor, were inactivated for target genes expression by transfection with dsRNAs, corresponding to 94% of the annotated protein-coding genes of *Drosophila*. After 2 days of incubation, in order to establish target gene repression, the automiG sensor expression was induced at 600 µM $CuSO_4$. 24 to 48 hour later the GFP expression was analyzed by microscopy as a readout for the sensor de-repression. **D.** Analysis of the automiG derepression by anti-GFP WB on S2R+ cells stably transfected with the automiG sensor and inactivated by RNAi for the indicated genes. Ftz and luciferase are negative controls. automiG silencing requires Drosha, Dcr-1, Ago-2, CG7009 as well as Dcr-2 for the loading. Mbf1 was used as a loading control. GFP: green fluorescent protein; CDS: coding DNA sequence; Ftz: *fushi tarazu*, WB – Western blot.

Reproducing the result shown Figure 41 was a high priority for our project. For this purpose, *Drosophila* S2 cells were transiently co-transfected with a plasmid carrying the automiG sensor as well as with dsRNA

precursors of siRNAs complementary to the transcripts of *CG7009*, Ago-2 (positive control) and *fushi tarazu*, Ftz (negative control). Three dsRNAs producing siRNAs targeting different parts of the *CG7009* mRNA were used: CG7009#1, CG7009#2 and CG7009#3 (Figure 42A and Methods). The expression of GFP levels, as readout for the extent of the automiG-silencing, was determined using a microscope (data not shown) and *via* Western Blot. γ-Tubulin levels were used as a loading control (Figure 42B and Methods).

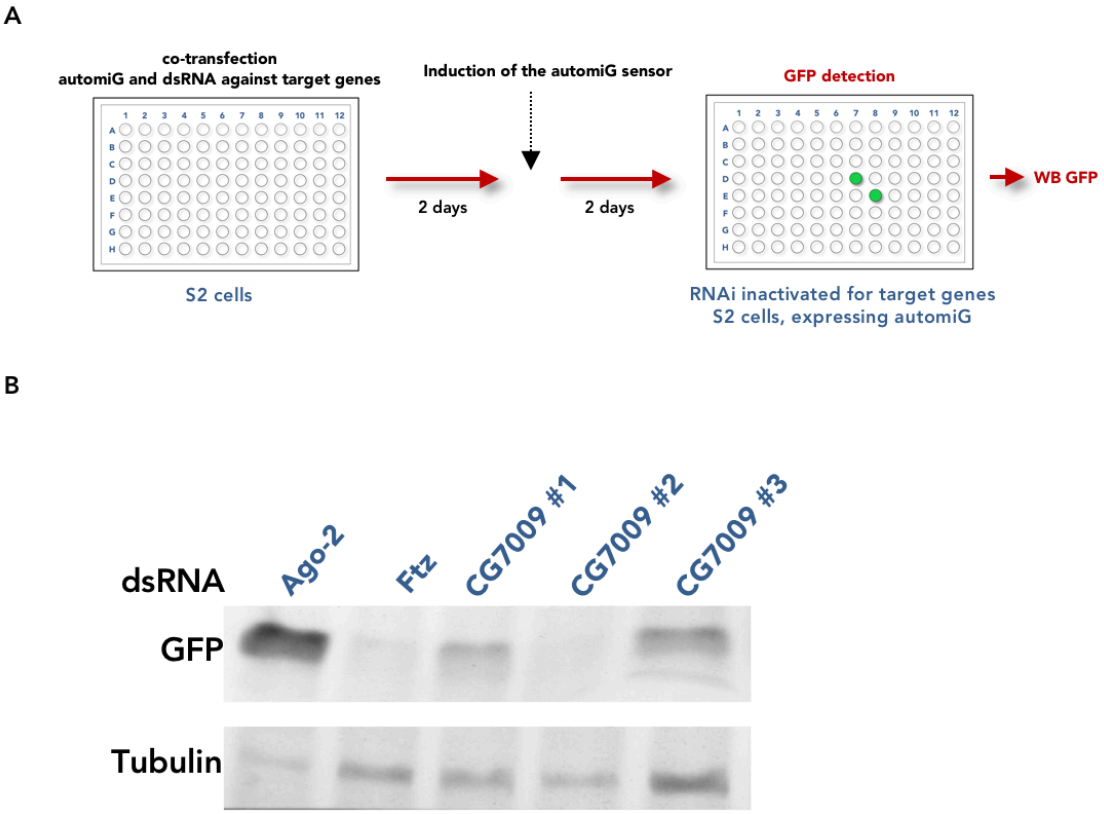

Figure 42. CG7009 is involved in the miRNA-Ago-2 silencing. A. Transfection scheme. **B.** GFP WB. Shown are the dsRNAs used for knocking down the corresponding associated genes, as well as the utilized antibodies. Ftz, negative control, Ago-2 positive control. GFP: Green Fluorescent Protein; Tub: γ-Tubulin.

Analysis of the Western Blot in Figure 42B shows that inactivation of *CG7009* with dsRNAs CG7009#1 and #3 leads to an increase of GFP expression, consistent with a de-repression of the miRNA-dependent automiG biosensor in a *CG7009* knockdown background. I was thus able to reproduce and confirm the involvement of CG7009 in the Ago-2-dependent miRNA silencing using a different transfection scheme and in S2 cell lines of *Drosophila*.

B. CG7009 is involved in the Ago-2-dependent miRNA pathway in living flies

In order to investigate the effects of CG7009 and CG5220 on Ago-2-miRNA silencing in living flies we took advantage of a home-made transgenic reporter system: automiW, a transgene reporter of the Ago-2-dependent miRNA pathway in adult flies (Besnard-Guerin et al., 2015). This construct is derived from the automiG sensor (Carre et al., 2013). The automiW construct contains a UAS promoter driving the expression of the GFP protein together with two artificial miRNAs miW-1 and miW-2 embedded in the Rpl17 intron. The sequences of the mature miW-1 and miW-2 miRNAs are perfectly complementary to sequences of the *white* gene marker in the construct, whose product is involved in red pigment deposition in adult eyes. miW-1 and miW-2 are able to repress expression of the *white* gene, resulting in white colored eyes (Figure 43A and B) (Besnard-Guerin et al., 2015).

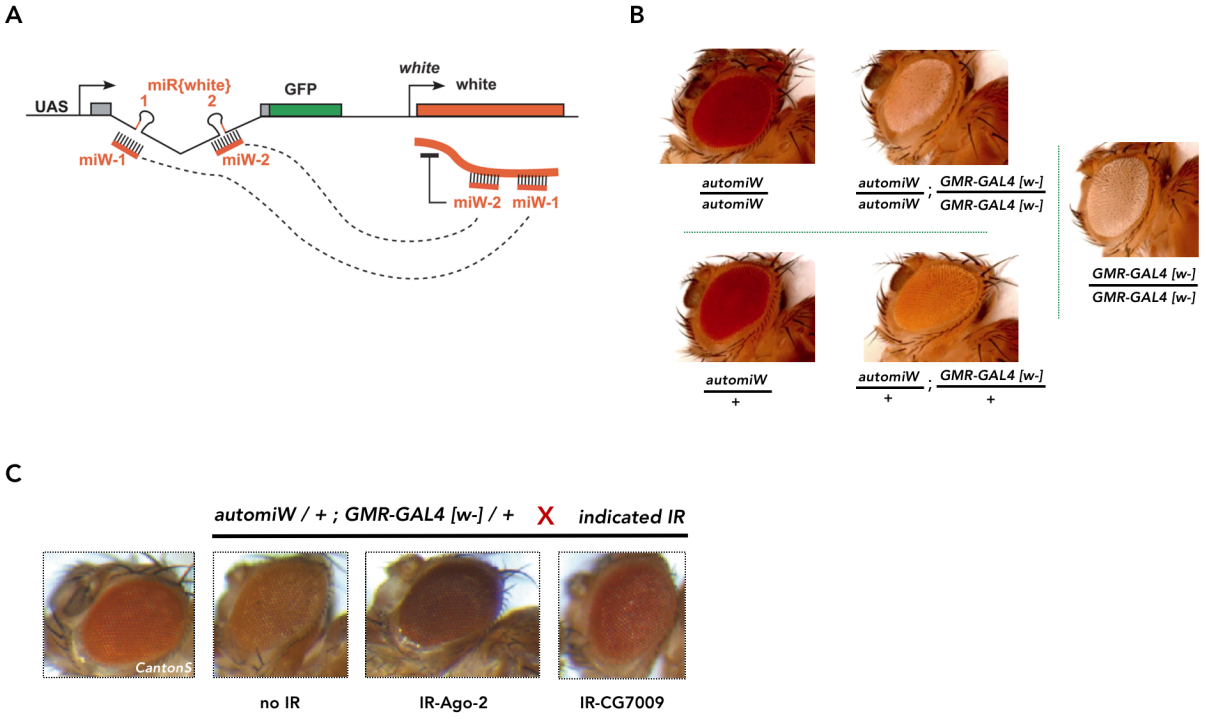

Figure 43. CG7009 is involved in the Ago-2-dependent miRNA pathway in flies. A. Schematic representation of the automiW transgene (Besnard-Guerin et al., 2015). The automiW system allows the expression of the *white* gene and two miRNAs targeting *white* and acting through Ago-2 (not shown). **B**. Analysis of the automiW silencing in adult fly eyes. The automiW transgene is expressed in eyes of *Drosophila* by the presence of a GMR-Gal4 transgene. When the transgene is not expressed (*left*, upper and lower panels), the eye color is red. Expression of the transgene in the eye in the presence of a GMR-Gal4 driver (*middle*, upper and lower panels) leads to loss of eye color pigmentation. Two doses of the automiW transgene silence the *white* gene more efficiently (*middle*, upper vs lower panels). Flies, homozygous for the eye-specific GMR-Gal4 driver do not bear functional copies of the *white* gene (*right* panel). **C.** The automiW transgene was specifically expressed in eyes of *Drosophila* by the presence of a GMR-Gal4 transgene. The silencing by the automiW sensor was analyzed in flies of the *CantonS* strain, having very strong red eye coloration (panel 1). automiW expression was induced in the eyes of *CantonS* flies as a negative control (panel 2, no IR) and upon RNAi knockdown by inverted repeats transgenes (IR) of Ago-2, as a positive control (panel 3) and CG7009 (panel 4).

Experiments performed in our lab demonstrated that inverted-repeat (IR) triggered RNAi knock-downs of Drosha or Ago-2 strongly inhibit the silencing of the *white* gene in flies expressing the automiW transgene in the eyes. These data thus characterized the automiW system as a biosensor for Ago-2-miRNA mediated silencing of the *white* gene by the two artificial miRNAs miW-1 and miW-2 (Besnard-Guerin et al., 2015). A *Drosophila* transgenic line expressing an automiW construct and a GMR promoter-dependent GAL4 driver allows the expression of the miW miRNAs together with the *white* gene copy contained in the automiW transgene only in the differentiated eye, as GMR is an eye specific promoter expressed during this eye developmental stage. Flies of a *CantonS [w+]* genomic background, that are homozygous or heterozygous for the automiW transgene and do not bear a GAL4 driver, do not express the automiW and have a dark red color of the eyes (Figure 43B, *left*, upper and lower panels). In contrast, flies homozygous for the eye-specific driver GMR-GAL4 [w-] having no functional copies of the mini-*white* gene and have white colored eyes (Figure 43B, *right* panel). When the automiW transgene is expressed in the differentiated eyes of adult flies under the control of the GMR-GAL4 [w-] driver (Figure 43B, *middle*, upper and lower panels) the eye color turns white or orange. This indicates that the two artificial miRNAs, miW-1 and miW-2, are capable to silence *in vivo* the expression of the mini-*white* marker carried by the automiW transgene, as well as the endogenous copy of *white* of the *CantonS [w+]* flies. Moreover, the silencing is dose-dependent as the eye loss of pigmentation is stronger when two doses of the automiW transgene are expressed in flies with the genotypes *w+; automiW [w+] / automiW [w+] ; GMR-GAL4 [w-] / GMR-GAL4 [w-]* (Figure 43B, *middle*, upper panel) when compared to flies heterozygous for the transgene of genotype *w+; automiW [w+] / + ; GMR-GAL4 [w-] /* + (Figure 43B, *middle*, lower panel).

During my PhD we analyzed the functions of *CG7009* and *CG5220* in Ago-2-miRNA silencing in adult flies by using the described automiW sensor silencing system. As described above, expression of the automiW transgene in the eyes of *CantonS* flies leads to decrease in the intensity of the eye pigmentation (Figure 43C, compare panel 1 and 2). The simultaneous expression of automiW and an IR against Ago-2 as positive control (Figure 43C, panel 3) or *CG7009* (Figure 43C, panel 4) leads to derepression of the *white* gene leading to darker eye coloration. This indicated that, similarly to what we observed in cells, *CG7009* is implicated in the Ago-2 miRNA pathway in living flies.

We are currently analyzing a small RNA sequencing experiment, which was performed on heterozygous (controls), *CG7009* mutants and *CG7009, CG5220* double mutants. We sequenced both adult somatic tissues (heads) and a germinal tissue (ovaries) for all the indicated genotypes in at least two biological replicates for the indicated conditions. Several important preliminary observations have been validated. Briefly, the *CG7009* single mutation, as well as the *CG7009, CG5220* double mutation, don't seem to be linked to a global down-regulation of the three sncRNAs populations (data not shown, ongoing analysis). However, when studying the different categories of RNA classes, there seem to be a ≈10 % decrease in the global miRNA populations in the ovaries of *CG7009* mutants (data not shown, ongoing analysis).

When analyzing the differential expression of miRNAs in more detail (using the DESeq2 tool in Galaxy), my analysis suggests that several miRNAs are deregulated in ovaries as well as in heads for both mutants and the double mutant compared to the control. I am currently trying to validate these deregulated miRNAs candidates by RT-qPCR on small RNAs and by Northern blot.

Interestingly, the Northern blot below performed on miG1 and Bantam on S2 *Drosophila* Cells expressing the automiG inducible sensor suggests that, miG1 (in difference to bantam) miRNAs, accumulation is affected in a *CG7009* KD condition in S2 cells (Figure 44). However, it is hard to conclude from this Northern blot whether *CG7009* KD affects the stability of miG1 or if the of the automiG sensor expression was affected in the condition *CG7009* KD.

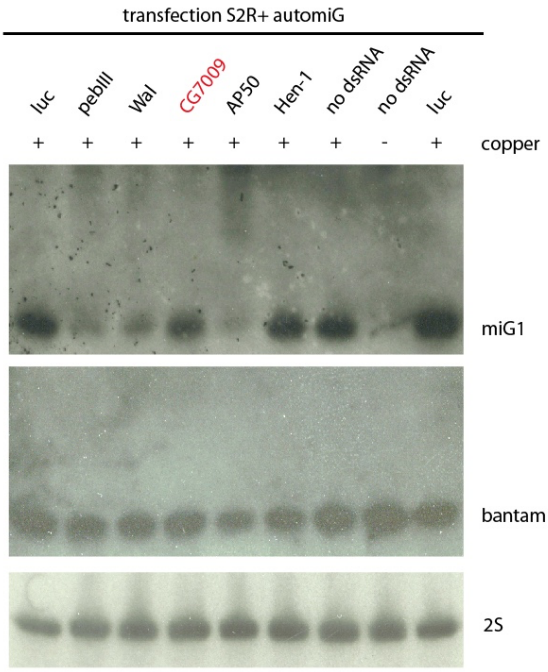

Figure 44. CG7009 is involved in miRNA stability or expression. Northern blot experiment performed in the indicated knockdown cells (the name of the KD gene is indicated on the top) with or without induction of the automiG sensor (+/- copper) in *Drosophila* S2R+ cells. The same membrane was probed 3 times with the different indicated probes (miG1, bantam and 2S rRNA). Probes sequences and methods are available in (Carre et al., 2013).

2. CG7009 is involved in the siRNA pathway

Since Ago-2 is mainly known as the mediator of the siRNA pathway, the involvement of *CG7009* in this silencing pathway was assayed. In a first assay, performed before my arrival in the laboratory, the effect of CG7009 on the siRNA pathway was determined by a dual luciferase assay in S2R+ cells (Figure 45). Cells were transfected by the transfection scheme illustrated on Figure 45C with a dsRNA against Firefly luciferase as well as with two plasmids expressing firefly and renilla luciferases CDS. Renilla was used for transfection efficiency normalization. The cells were simultaneously inactivated for Ago-2 or Dcr-2 (positive controls), Ftz (noted on the figure as control), as negative control, and also for the miRNA pathway factors Ago-1, drosha and Dcr-1 by co-transfection with a corresponding dsRNA (Figure 45D).

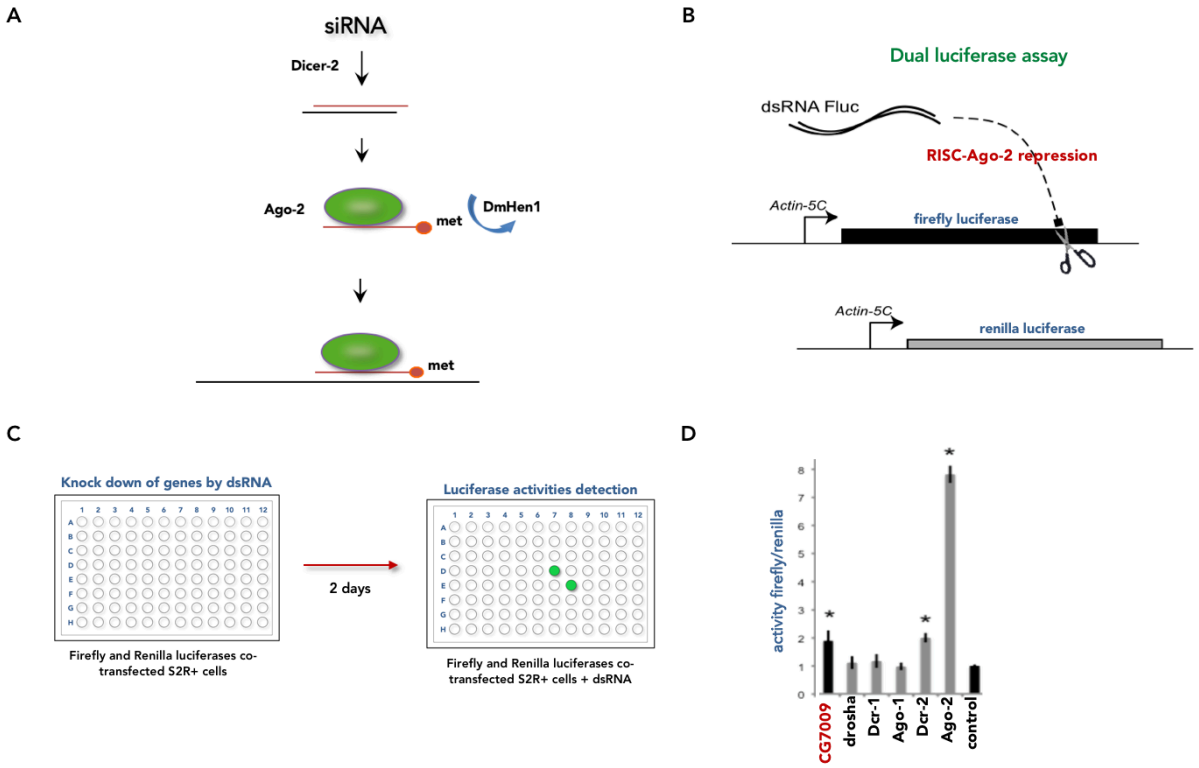

Figure 45. Involvement of CG7009 in the siRNA pathway. A. Schematic representation of the siRNA pathway. The key proteins inactivated as positive controls are shown. **B.** Illustration of the dual luciferase assay. The expression of the coding sequence of firefly luciferase was made dependent of the siRNA pathway by a co-transfection with a dsRNA against firefly in S2R+ cells. The cells were equally transfected with a plasmid allowing the expression of Renilla luciferase, as transfection control. **C.** Illustration of the transfection scheme used to test CG7009's involvement in the siRNA pathway. The activities of firefly and Renilla were measured 48h after the transfection. **D.** Averages values of the ratio of firefly / renilla activities from three independent experiments. The average values of the firefly / renilla activities ratios of a dsRNA inactivation of the GFP (negative control) was set at 1. Statistical analysis was performed using the Student test; *: *p-value* < 0.05.

Figure 45D demonstrated that Ago-2 inactivation strongly inhibits the repression of the firefly luciferase by dsRNA. The same was observed, although to a lesser effect, for the Dcr-2 RNAi-inactivated condition. Interestingly, inactivating *CG7009* led to mild but significant repression of the siRNA-mediated silencing of the firefly luciferase, involving the gene in the siRNA pathway.

At the beginning of my project, I had to reproduce this experiment. A similar transfection scheme to the one used for testing the gene's implication in the miRNA-Ago-2 pathway was used in S2 cells (Figure 46B). The difference here is that the GFP expression was made dependent on the siRNA pathway only, rather than the miRNA-Ago-2 pathway where GFP expression depends on both miRNA biogenesis and Ago-2 repression.

For this purpose, I took advantage of another construct derived from the automiG reporter and termed automiG-Δ1-Δ2 (referred here as automiG00) (Carre et al., 2013). In this construction the two artificial anti-GFP miRNAs have been deleted from the automiG sequence, resulting in a plasmid expressing the GFP CDS under the control of automiG's copper-inducible promoter, pMT (Figure 46A). The GFP transcript can thus be targeted by transfection with anti-GFP siRNAs, similarly to the firefly experiments described above. Accordingly, in a non-functional siRNA pathway background (Ago-2 *knock-down* for example), the anti-GFP siRNAs derived from the dsRNA dicing will not be able to cleave the GFP mRNA, inducing an increase of the GFP signal (Figure 46A).

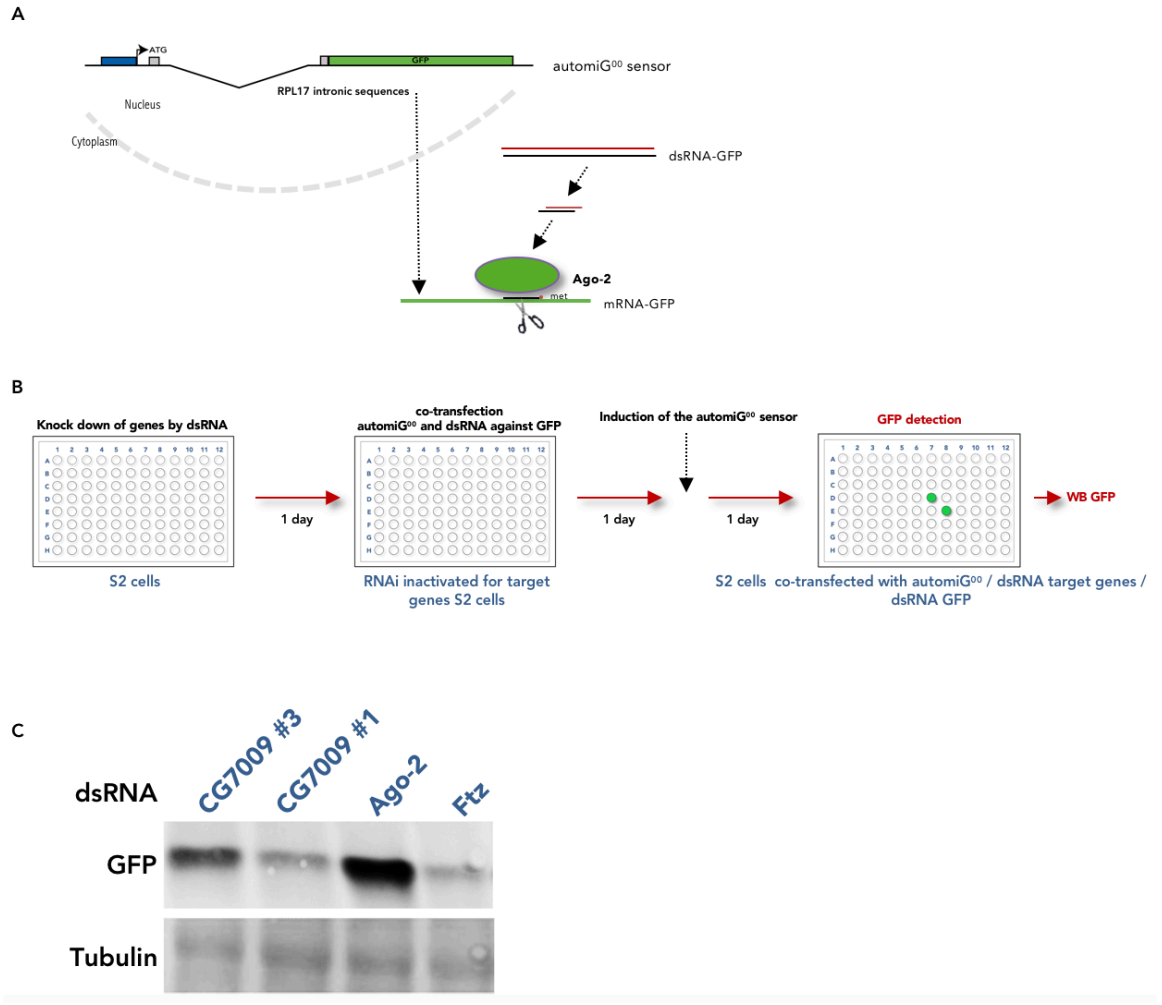

Figure 46. Involvement of CG7009 in the siRNA pathway. A. Illustration of the experimental design used to test the implication of CG7009 in the siRNA pathway: the GFP expression by the automiG^{00} sensor depends on the siRNA silencing, as it is co-transfected with a dsRNA producing siRNA against the GFP. **B.** Illustration of the transfection scheme used to test CG7009's involvement in the siRNA pathway. **C.** Western Blot analysis of the GFP expression. dsRNAs used for knocking down the corresponding associated genes as well as the utilized antibodies are shown. Ftz, negative control, Ago-2 positive control. GFP: Green Fluorescent Protein; Tub: γ-Tubulin.

The result of the Western Blot (Figure 46C) demonstrates that *CG7009* is implicated in the siRNA pathway, since we can see that inactivation of this gene induces an increase in GFP expression levels, in contrast with the *Ftz* control. The CG7009#3 dsRNA condition seems to induce the strongest effect on the RNAi response targeted against GFP. The result obtained with Ago-2, the positive control, further validates our conclusion. I was thus able to confirm the preliminary result from the dual-luciferase assay, using a different system and in a different cell line.

A. Reduced viral resistance to DCV injection

In order to investigate if CG7009 and CG5220 have a biological function in the siRNA pathway, we decided to determine the resistance of the two corresponding mutants to *Drosophila C* virus (DCV) infection. First, we tried to analyze the survival of the flies after intra-thoracic injection, either with DCV particles, or with buffer, as control. However, the flies were dying so fast after the DCV injection that the results were hardly interpretable (data not shown). We decided to analyze the antiviral defense in the genotypes of interest by determining, after injection, the DCV load by RT-qPCR (Figure 47).

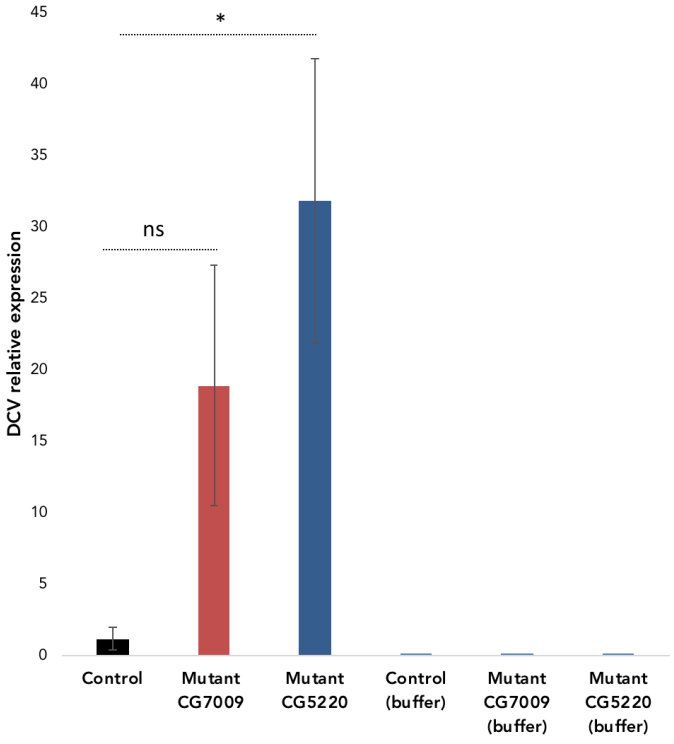

Figure 47. Effect of *CG7009* and *CG5220* mutations on the resistance to DCV. RT-qPCR analysis on the DCV viral load in flies injected with a 50x diluted stock of DCV at 10^8 PFU / ml (plaque-forming units / ml) or buffer as control (Methods). Two replicates of 5 flies from the following genotypes were used: $CG7009^{e02001}$ / TM6, Tb, Sb, as control (black bar); $CG7009^{e02001\text{-}G10}$ / $CG7009^{e02001\text{-}G10}$, as mutant for *CG7009*; and $CG5220^{K28A}$ / $CG5220^{K28A}$, as mutants for *CG5220*. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. The primers used for the RT-qPCR are DCV_FW and DCV_Rev (see Methods). Rp49 was used for normalization (primers SD22 and SD23, (Methods).

Figure 47 represents the viral load of injected with DCV $CG7009^{e02001}$ / TM6, Tb, Sb flies as control (black bar), $CG7009^{e02001\text{-}G10}$ homozygous mutants (red bar), and $CG5220^{K28A}$ homozygous mutants (blue bar, Methods). The last three bars on Figure 47 represent the DCV load of flies with the same genotypes, injected with buffer as control. It turned out that even with a very sensitive technique as RT-qPCR, it is difficult to obtain enough robust data in order for it to be challenged statistically.

The DCV load in the control heterozygotes was always very variable between the replicates of the experiment, even if very low concentrations (up to 50 times dilution of the DCV initial stock 10^8 PFU/ml)

of the injected virus were tested. This complicated the possibility to draw a clear conclusion on the experimental result.

It is possible that, due to experimental imperfection, the flies are too sensitive to small differences in the amount of injected virus. However, in a same replicate, the tendency is that there is more DCV load in the two mutants compared to the control. Sometimes more DCV load was detected in the *CG7009* mutant, and sometimes more was detected with the *CG5220* mutants, but they were always having a more important viral load than the heterozygous control flies.

This observation suggests a role of the two genes *CG7009* and *CG5220* in the anti-viral defense against DCV, viral defense well known to involve the siRNA pathway in flies. Additional experiments are required to solidify those results further, *e.g.* using more files per replicate or dilute the DCV stock even more.

3. Involvement in the somatic piRNA pathway

A. Test of the genes positively scoring in the automiG screen with a *gypsy::LacZ sensor*

To further analyze the candidates identified by our genomic screen with the automiG sensor, we took advantage of another reporter construct available in our laboratory, allowing to identify actors involved in the somatic piRNA pathway (*gypsy::lacZ* reporter)(Sarot et al., 2004).

This construct expresses the *lacZ* gene under control of the *gypsy* retrotransposon long terminal repeat (LTR) and produces a transcript containing 700nt of the *gypsy* sequence fused to the *lacZ* gene (Figure 49A). This region is naturally targeted by a high number of isolated piRNAs, which can thus guide efficient silencing of the 700nt of *gypsy* and consequently of the *LacZ* fused gene in flies having a functional somatic piRNA pathway. In contrast, a knockdown of key somatic piRNA pathway components, which will lead to a strong decrease in the levels of *gypsy*-derived piRNAs, triggers reporter de-silencing. Impairment of the silencing of the *gypsy::LacZ sensor* by the somatic piRNAs is detected by the appearance of *lacZ* expression. A readout for *LacZ* expression is the blue color produced from the action of the LacZ-encoded **β**-Galactosidase on the substrate X-gal: the higher the intensity of the ovariole's blue coloration, the stronger the de-repression of the somatic piRNA pathway (Sarot et al., 2004).

Using this sensor we discovered that 6 of the 17 candidates, which were already implicated in the miRNA-Ago-2 repression pathway, scored positively with the *gypsy::LacZ* sensor and thus seem to also

be involved in the somatic primary piRNA pathway. One of the genes that scored in both of the described screens is *CG7009* (Figure 48 and Figure 49B).

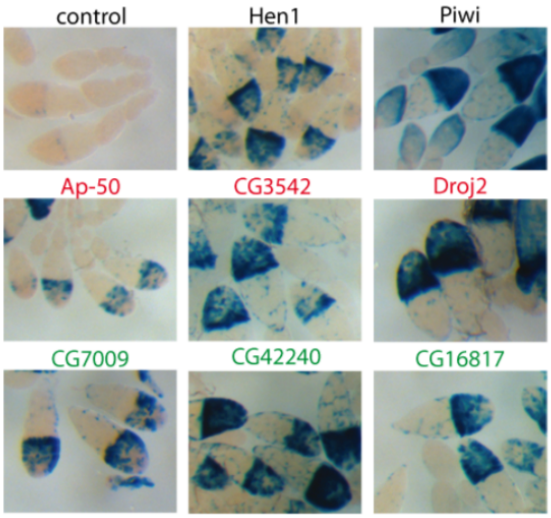

Figure 48. Silencing of a *gypsy::lacZ* sensor in follicular cells of wild-type ovaries (control) or in ovaries with RNAi knockdown of the indicated gene. Hen1 and Piwi are known positive controls. Three (in red) of the 6 genes identified in our RNAi screen have been independently identified in other screens for genes involved in piRNA silencing (Czech et al., 2013; Handler et al., 2013).

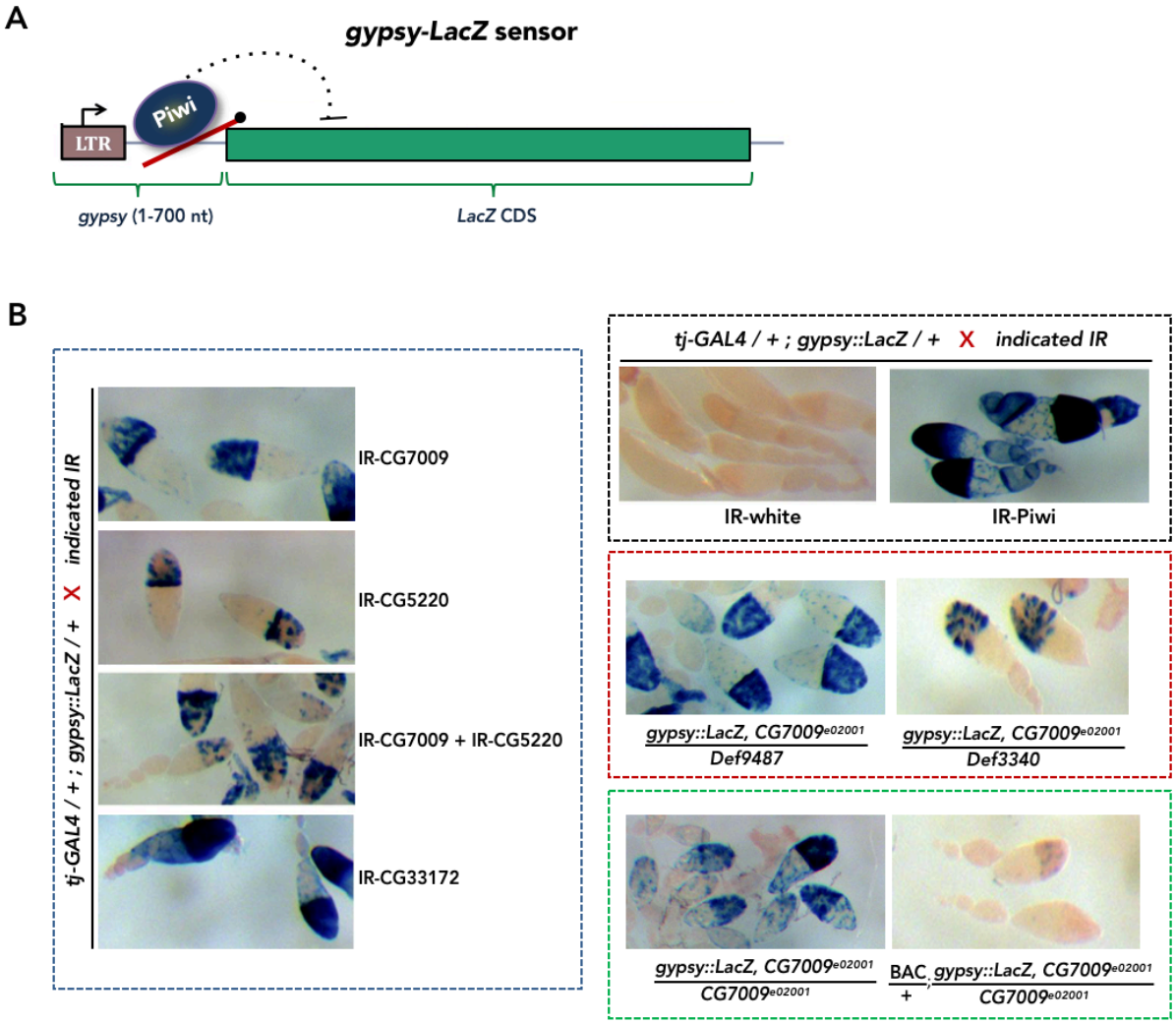

Figure 49. CG7009 has a function in the somatic piRNA pathway. A. Schematic representation of the *gypsy*-LacZ reporter showing the portion of the *gypsy* element inserted in the reporter construct. **B.** (***black panel***) X-Gal staining of ovarioles, as readout for the effects of the knockdown by RNAi using inverted repeats (IR) of Piwi, as positive control, *white*, as negative control, or *CG7009* in flies expressing the *gypsy::lacZ* reporter; (***blue panel***) X-Gal staining of ovarioles, as readout for the effects of the knockdown by RNAi using inverted repeats (IR) of CG7009, CG5220, CG7009 and CG5220 double inactivation, and CG33172; (***red panel***) X-Gal staining of ovarioles, as readout for the effects of the trans-heterozygous mutations of CG7009: $CG7009^{e02001}$/Def9487 (*left*) and $CG7009^{e02001}$/Def3340 (*right*) in flies expressing the *gypsy::LacZ* transgene; (***green panel***) X-Gal staining of ovarioles, as readout for the effects of the homozygous mutations of CG7009: $CG7009^{e02001}$/ $CG7009^{e02001}$ (*left*) and the genetic rescue of the reporter derepression by the line BAC/+; $CG7009^{e02001}$/ $CG7009^{e02001}$ in flies expressing the *gypsy::LacZ* transgene.

X-gal staining was performed on ovaries dissected from 4-5 days old females with the indicated RNAi inactivation (*blue* and *black panels*) or mutant genotypes (*green* and *red panels*). RNAi inactivation of the *white* gene, by using inverted-repeats (IR), served as a negative control; RNAi inactivation of Piwi, was used as a positive control (Figure 49B, *black panel*). RNAi inactivation of *CG7009*, *CG5220*, or

simultaneous inactivation of both induced the sensor de-repression. This result implicated both CG7009 and CG5220 in the somatic piRNA pathway.

Interestingly, RNAi inactivation of *CG33172*, the predicted partner of CG7009 and putative ortholog of the yeast and human proteins Trm734 and WDR6 (respectively), triggered even stronger de-silencing of the *gypsy::LacZ* sensor. As a reminder, we were able to detect and *in vitro* interaction between CG7009 and CG33172 recombinant proteins produced in bacteria (Figure 32A). It is interesting to mention that a synonym name for the yeast Trm734 is RTT10 (*Regulator of Ty1 transposition protein 10*). In *S. cerevisiae*, RTT10 was implicated in the control of *Ty1* transposition (Nyswaner et al., 2008). In a RTT10 mutant, *Ty1* was shown to be up regulated (30-fold increase). Trm7 Nm methylase, the yeast ortholog of CG7009, also scored in the same study as a *Ty1* regulation gene.

The derepression of the *gypsy::LacZ* sensor was also confirmed with two hetero-allelic null mutants expressing the sensor: $CG7009^{e02001}$, *gypsy::LacZ* / Def9487 and $CG7009^{e02001}$, *gypsy::LacZ* / Def3340 (Figure 49B, *red panel*), as well as with the homozygous mutant $CG7009^{e02001}$, *gypsy::LacZ* / $CG7009^{e02001}$, (Figure 49B, *green panel*), bearing the second hit mutation. However, the fact that this phenotype was retained between the homozygous and trans-heterozygous *CG7009* mutants indicated that the reporter de-silencing is due to *CG7009* loss-of-function and not to the second hit mutation.

Moreover, we were able to rescue the reporter de-repression with a transgenic line "BAC" in the homozygous mutant for *CG7009* (BAC/+ ; $CG7009^{e02001}$, *gypsy::LacZ* / $CG7009^{e02001}$). However, it is important to mention that the BAC rescue is to be taken with caution, as the result is not always as clear as the depicted in the picture of Figure 49B. As discussed for the survival experiments, we have observed that the insertion of the "BAC" renders the flies "sick", and probably does not restitute all of the *CG7009* functions properly, despite its capacity to rescue the methylation activity on $tRNA^{Phe}$ (Figure 40D).

B. RT-qPCR analysis confirms the role of CG7009 in *gypsy* silencing in flies expressing *gypsy*::LacZ

In an attempt to validate the result of the *gypsy::LacZ* staining in ovarioles, I performed RT-qPCR on total RNA extracted from ovaries of 4 days old $CG7009^{e02001\text{-}G10}$, *gypsy::LacZ / +* females, used as control and on *CG7009* mutants expressing the gypsy-LacZ sensor: $CG7009^{e02001\text{-}G10}$, *gypsy::LacZ* / $CG7009^{e02001\text{-}G10}$ (Figure 50). For this experiment I used primers annealing to a region in the first 700 nt of the *gypsy*

sequence (Figure 50A), and thus giving information on *gypsy* expression from both the *gypsy* sensor transgene and the potential endogenous copies of *gypsy* retroelements. I also used primers specific to the *LacZ* CDS region (Figure 50B), providing information on the expression and thus repression level of the *gypsy::LacZ* sensor. A third set of primers, annealing to the *Pol* gene of the retrotransposon (between 2.6 and 2.8 kb of gypsy CDS), and thus amplifying a sequence of *gypsy* not present in the sensor, gave me information on the expression level of only the endogenous *gypsy* transcripts (Figure 50C). As a control for the specificity of the primers annealing to *LacZ* I performed RT-qPCR on flies not expressing the *gypsy::LacZ* transgene (not shown), and indeed confirmed that no amplification is detected for *LacZ* in this experiment.

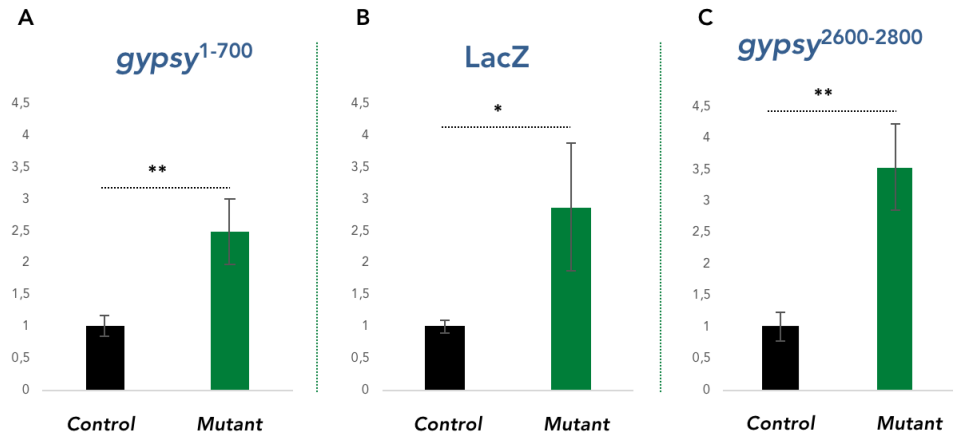

Figure 50. *gypsy* expression in *CG7009* mutants, expressing the *gypsy::LacZ* sensor. Determination by RT-qPCR of the relative expression level of *gypsy* (**A** and **C**) and LacZ (**B**) in ovaries from 4 days old flies with genotype $CG7009^{e02001\text{-}G10}$, *gypsy::LacZ* / $CG7009^{e02001\text{-}G10}$, denoted as "Mutant" (*green bars*), compared to *gypsy* expression in ovaries from 4 days old heterozygous females $CG7009^{e02001\text{-}G10}$, *gypsy::LacZ* / +, denoted as "Control" (*black bars*). Average values of three independent biological replicates are shown. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. The sequences of the primers used for the RT-qPCR, as follows: primers MA087 and MA088 in **A.**, LacZ_2-FW and LacZ_2-Rev in **B.** and primers MA037 and MA038 in **C.** are available in the Methods section. Rp49 was used for normalization (primers SD22 and SD23, Methods).

The result on Figure 50B validates the effects observed for *CG7009* loss-of function, as previously detected with the ovarioles X-gal staining. As expected, the *gypsy* sequence contained in the *gypsy::LacZ* transgene as well as on the potential endogenous copies of *gypsy* ($gypsy^{1\text{-}700}$) also showed up-regulation in *CG7009* mutants (Figure 50A). The last bar shows that in flies expressing the sensor in a *CG7009* mutant background, the endogenous copies of *gypsy* are transcriptionally up-regulated (Figure 50C, $gypsy^{2600\text{-}}$

[2800] primers). In the three tested conditions I detected significant up-regulation of *gypsy* mRNA levels when the *gypsy*-sensor transgene is present in the tested flies.

An *in-silico* analysis of the primers used showed that they do not hybridize, in the conditions of this experiment, with the *flamenco* locus transcript, which contains some *gypsy* anti-sense sequences that can potentially be amplified by this RT-qPCR. Thus, the difference in *gypsy* expression levels between the control heterozygous individuals and the homozygous mutant flies detected in this RT-qPCRs is most probably a readout for differences in the transcript levels of *gypsy* between *CG7009* mutants and heterozygous individuals expressing this transgene. This result solidifies the previously detected involvement of CG7009 in *gypsy* expression regulation.

I would also like to mention that we tested for a role of *CG7009* in the piRNA pathway operating in germ cells. We did this by using Burdock-lacZ, another genetic sensor specific for the germinal piRNA pathway (Handler et al., 2013). We couldn't find clear, reproducible effects with neither *CG7009* nor with *CG5220* mutants (Figure 51). However, this result should be challenged in the future with the new CRISPR alleles now available in our laboratory.

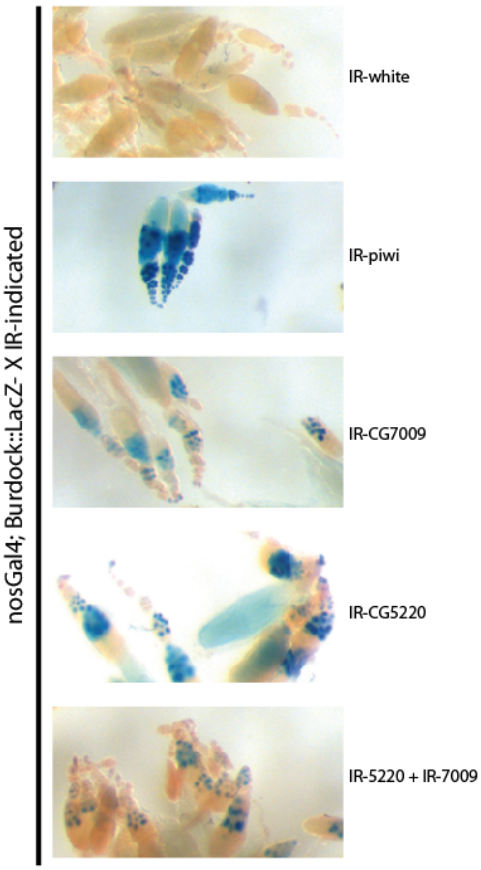

Figure 51. Burdock piRNA sensor in germ cells. X-Gal staining of ovarioles, as readout for the effects of the knockdown by RNAi using inverted repeats (IR) of Piwi, as positive control, *white*, as negative control, or *CG7009*, *CG5220* or both CG7009 + CG5220 in flies expressing the *Burdock*::lacZ reporter (Handler et al., 2013).

4. *CG7009* and *CG5220* mutants not expressing *gypsy*::LacZ transgene have opposite effects on the endogenous *gypsy* expression in the ovaries

A very unexpected result was obtained when I amplified *gypsy* mRNA in flies not expressing the *gypsy::LacZ* sensor by RT-qPCR (Figure 52A and B). Both of the tested set of primers, allowing amplification of regions 1-700 (Figure 52A) and 2600-2800 (Figure 52B) of the retrotransposon sequence, showed significant down-regulation of endogenous *gypsy* in the ovaries of $CG7009^{e02001\text{-}G10}$ / $CG7009^{e02001\text{-}G10}$ females (*red bars*) and $CG5220^{K28A}$ / $CG5220^{K28A}$ females (*green bars*) compared to *gypsy* expression in $CG7009^{e02001}$ / TM6, Tb, Sb heterozygous flies (*black bars*).

Thus, in striking contrast to the sensor de-repression and the RT-qPCR results seen in presence of the *gypsy::LacZ*, endogenous *gypsy* gene seemed to be down-regulated by *CG7009* lack-of-function when the *gypsy::LacZ* sensor is absent.

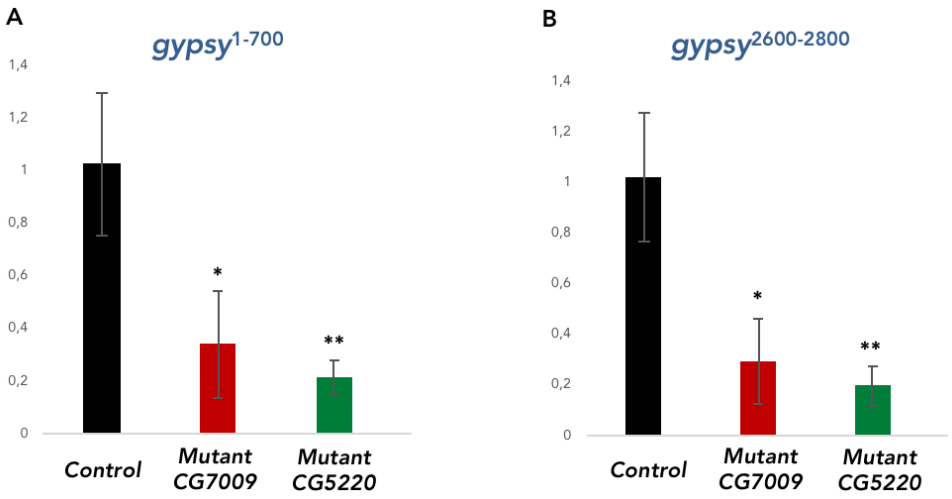

Figure 52. Endogenous *gypsy* expression in *CG7009* and *CG5220* mutants, in the absence of the *gypsy::LacZ* sensor. Determination by RT-qPCR of the relative expression levels of *gypsy* transcript regions 1-700 nt (**A**) and 2600-2800 nt (**B**) in ovaries from 4 days old flies with genotype $CG7009^{e02001\text{-}G10}$ / $CG7009^{e02001\text{-}G10}$, denoted as "Mutant *CG7009*" (red graphs), and in flies $CG5220^{K28A}$ / $CG5220^{K28A}$, denoted as "Mutant CG5220" (green graphs), both compared to *gypsy* expression in ovaries from 4 days old heterozygous females $CG7009^{e02001}$ / TM6, Tb, Sb, denoted as "Control" (black graphs). Average values of three independent biological replicates are shown. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. The sequences of the primers used for the RT-qPCR, as follows: primers MA087 and MA088 in **A.** and primers MA037 and MA038 in **C.** are available in the Methods section. Rp49 (primers SD22 and SD23) was used for normalization.

This result has been reproduced multiple times, including with trans-heterozygous mutants *CG7009*^{e02001}/ Def3340 (Figure 55) and *CG7009*^{e02001}/ Def9487 (data not shown). We further searched to explain this discrepancy between the expression of *gypsy* in *CG7009* mutant in the presence and absence of the *gypsy::LacZ* transgene.

A. The spliced form of *gypsy* is not increased in *CG7009* mutants

It has been shown that when the gypsy retrovirus is mobilized it gets spliced, in order to express its envelope protein *env* and to infect the germline from the follicular cells (Pelisson et al., 1994). The primers that were used in our RT-qPCR for amplification of the *gypsy* regions 1-700 or 2600-2800 cannot give a product from the spliced form of *gypsy* (Figure 53).

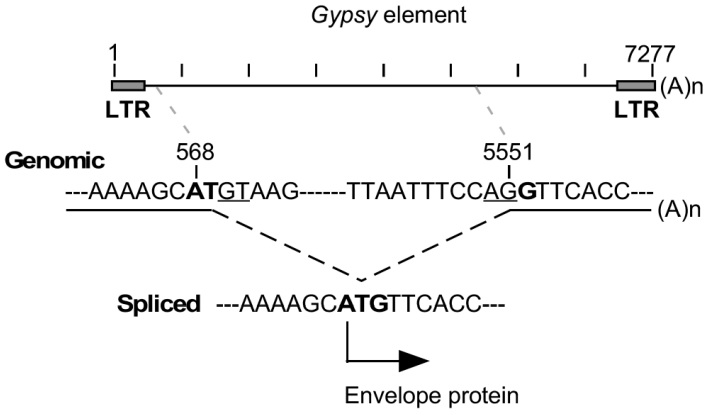

Figure 53. Schematic representation of the splicing of *gypsy*. Image from (Genenncher et al., 2018). The reverse primer MA088 used to amplify *gypsy* sequence 1-700 nt hybridize exactly at nucleotides 639-658 and can thus not give amplification on the spliced from of *gypsy*.

It is possible that an increase of the spliced form of *gypsy* spliced, accompanied by a decrease of its un-spliced form, is occurring in *CG7009* and *CG5220* mutants and not in the *CG7009*^{e02001} / TM6, Tb, Sb control heterozygotes. If this is the case, by amplifying the *gypsy* un-spliced form (*gypsy*$^{1-700}$ or *gypsy*$^{2600-2800}$ primers) we would obtain a decrease in mutants and not in control, as seen in the RT-qPCR result on Figure 52A and B.

Therefore, I performed an RT-qPCR on *CG7009* and *CG5220* mutant females with specific primers for the spliced form of *gypsy* (Genenncher et al., 2018). Even if the expression levels were very variable between the replicates (data not shown), the tendency was that similarly to the other *gypsy* transcripts, the spliced transcript is down-regulated in both *CG7009* and *CG5220* mutants when compared to control in the absence of *gypsy::LacZ* sensor.

B. CG7009- and CG5220-mediated *gypsy* down-regulation is independent of the heat-shock response

The study of Genenncher and colleagues (Genenncher et al., 2018), to which my team was associated, showed that the tRNA methylase Dnmt2 affects *gypsy* transcription in a more pronounced way during the heat-shock response. NSun2, another tRNA methylase, was shown to mediate its effect on *gypsy* transcription in a heat-shock-independent fashion (Genenncher et al., 2018). However, it is important to mention that lack-of-function of both of these tRNA methylases was linked to *gypsy* transcripts up-regulation and not down-regulation, as seen for CG7009 in the ovaries (Figure 52A and B) and in whole females (data not shown) flies that do not carry a *gypsy*-sensor construct.

In an attempt to understand the effect of *CG7009* mutation on *gypsy*, I tested the transcription of *gypsy* (region 2600-2800) in heat-shocked 4 days old females by RT-qPCR (Figure 54).

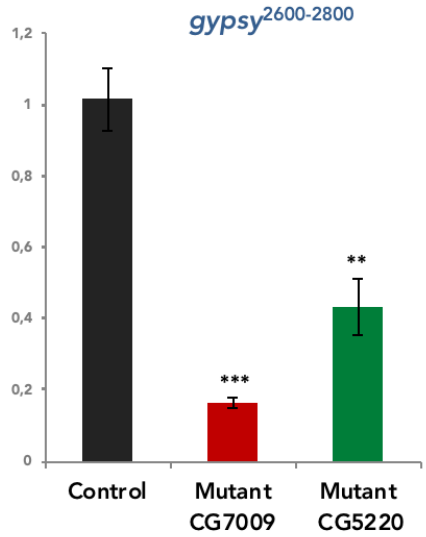

Figure 54. The heat-shock response does not affect expression of *gypsy* in *CG7009* and *CG5220* mutants. Determination by RT-qPCR of the relative expression levels of *gypsy* transcript regions 2600-2800 nt in 4 days old females subjected to 1 hour heat shock in water bath at 37°C, followed by 5 hours of recovery at 25°C. The flies have the following genotypes: $CG7009^{e02001\text{-}G10}$ / $CG7009^{e02001\text{-}G10}$, denoted as "Mutant *CG7009*" (red graph); $CG5220^{K28A}$ / $CG5220^{K28A}$, denoted as "Mutant CG5220" (green graph). Both mutants were compared for *gypsy* expression with heterozygous females $CG7009^{e02001}$ / TM6, Tb, Sb, denoted as "Control" (black graph). Average values of three independent biological replicates are shown. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. The sequences of the primers used for the RT-qPCR are primers MA037 and MA038 (Methods). Rp49 (primers SD22 and SD23, Methods), was used for normalization.

Figure 54 shows *gypsy* transcript levels of females with genotypes *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ and *CG5220*^{K28A} / *CG5220*^{K28A} compared to control flies with genotype *CG7009*^{e02001} / TM6, Tb, Sb. The result shows that endogenous *gypsy* is down-regulated in the *CG5220* or *CG7009* mutant backgrounds during the heat-shock response, as seen before in a non heat-shock condition. Thus, it seems that, in contrast to the m^5C methylase Dnmt2, CG7009 and CG5220 function independently of the heat-shock response regarding *gypsy* control.

C. The effect of CG7009 on *gypsy* down-regulation is independent of ageing

It has been reported that during normal ageing TEs, including *gypsy*, get de-repressed naturally (Li et al., 2013a). Moreover, this study reported that TE activation in ageing flies is linked to neuronal decline. In regards to the function of FTSJ1, the human CG7009 ortholog, in the development of neuropathology and to our results involving CG7009 in the sncRNA pathways regulating TEs, I performed another investigation on *gypsy* expression by RT-qPCR in ageing flies of 4, 14 and 28 days, the idea being that maybe CG7009's function on *gypsy* up-regulation appears later in the development of the fly. This study was conducted with flies *CG7009*^{e02001}*/Def3340*, as at this time the homozygous line CG7009 (G10) was not yet established. For the experiment I analyzed *gypsy* expression separately in ovaries, heads (from females) and whole males. In the ovaries, *gypsy* expression from region 2600-2800 remained similarly repressed in 4, 14 and 28 days old mutants for *CG7009* compared to the heterozygous balanced individuals (data not shown).

This result indicated that the observed repression of endogenous *gypsy*, detected in the ovaries of *CG7009* mutants, is independent of the flies' age.

D. *Gypsy* is up-regulated in *CG7009*, *CG5220* double mutants

Interestingly RT-qPCR analysis showed that *gypsy* expression is increased in the ovaries of *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} / CG7009$^{e02001\text{-}G10}$, *CG5220*^{K28A} (Rec1) homozygous double mutants (Figure 55, orange bar) in comparison to control heterozygous flies (black bar).

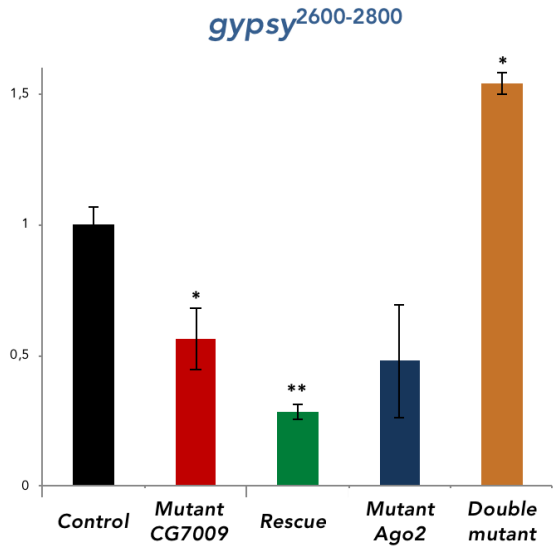

Figure 55. *gypsy* is up-regulated in the ovaries of *CG7009, CG5220* double mutants. $CG7009^{e02001\text{-}G10}$, ^{CG5220K28A} mutants. Determination by RT-qPCR of the relative expression level of *gypsy* transcript region 2600-2800 nt in 4 days old females with the following genotypes: $CG7009^{e02001}$ / TM6, Tb, Sb, denoted as "Control" (black bar); $CG7009^{e02001}$ / Def3340, denoted as "Mutant *CG7009*" (red bar); BAC/BAC; $CG7009^{e02001}$ / Def3340, denoted as "Rescue" (green bar); $Ago\text{-}2^{414}$ /$Ago\text{-}2^{414}$, denoted as "Mutant Ago2" (blue bar); and $CG7009^{e02001\text{-}G10}$, $^{CG5220K28A\text{-}Rec1}$ / $CG7009^{e02001\text{-}G10}$, $^{CG5220K28A\text{-}Rec1}$, denoted as "Double Mutant", (orange bar). Average values of three independent biological replicates are shown. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. The primers used for the RT-qPCR are MA037 and MA038 (see Methods). Rp49 (primers SD22 and SD23, see Methods), was used for normalization.

E. A putative role of the siRNA pathway in the repression of a *gypsy::LacZ* sensor

In the ovaries of CG7009 mutant flies the siRNA pathway is expected to be compromised. It was thus possible that the repression of *gypsy* transcription in the ovaries of flies having a *gypsy::LacZ* sensor can be influenced by the impaired siRNA background in this flies. If this was to be true a siRNA pathway component was expected to also affect the *gypsy::LacZ* sensor regulation. In order to test this we analyzed the effect of an siRNA pathway component, Dcr-2 on the silencing of a *gypsy::LacZ* reporter (Figure 56).

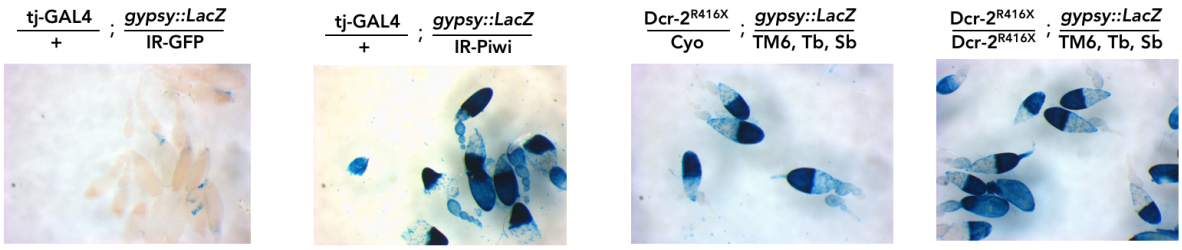

Figure 56. Dcr-2 is involved in *gypsy::LacZ* silencing. X-Gal staining of ovarioles, as readout for the effects on *gypsy::LacZ* silencing of the knockdown by RNAi using inverted repeats (IR) of GFP, as negative control, (first panel) and Piwi, as positive control (second panel). X-Gal staining of ovarioles, as readout for the effects on *gypsy::LacZ* silencing of the heterozygous (third panel) and homozygous (fourth panel) mutations of Dicer-2 (*Dcr-2*^{R416X}) in flies expressing the *gypsy::lacZ* reporter. GFP: green fluorescent protein

Figure 56 indicates that Dcr-2 is involved in the repression of the *gypsy::LacZ* sensor (Figure 56, last two panels). Even one dose less of Dcr-2 seems to be sufficient to trigger the sensor de-repression (Figure 56, third panel). An RNAi inactivation of GFP was used as a negative control (Figure 56, first panel) and an RNAi inactivation of Piwi, served as positive control of the experiment (Figure 56, second panel).

We were surprised that if there is an implication of Dcr-2 in the regulation of *gypsy::LacZ* silencing it was not characterized before, because several genome-wide screens unraveled multiple factors involved in the piRNA pathway (Czech et al., 2013; Handler et al., 2013). However, those were RNAi based screens, and our results were obtained with mutants for Dcr-2 and not with RNAi KD inactivation which depends on Dcr-2. Our experience with RNAi-mediated inactivation of Dcr-2 (Carre et al., 2013) shows that we have to perform the tests very soon after the expression of a dsRNA against Dcr-2 (a well-known processive enzyme) in order to see the result of the silencing, as RNAi-mediated silencing is strongly dependent on Dcr-2 activity. It is possible that an effect of Dcr-2 on the *gypsy::LacZ* sensor was missed by those screens.

The result on Figure 56 has to be taken with caution as it has been tested only with one mutant for Dcr-2 line. If it gets confirmed using other mutant combination it opens the possibility that there is indeed an effect of the siRNA pathway enzyme Dcr-2 on the regulation of a *gypsy::LacZ*-sensor.

F. CG7009 induces de-repression of endogenous *gypsy* in heads

In their study, Li and colleagues, linked neuronal decline to the up-regulation of *gypsy* (Li et al., 2013a) and *CG7009* is the ortholog of the neuro-pathology linked gene FTSJ1. Thus, I also tested if *CG7009* has an impact on *gypsy* expression in the heads of the flies. This tissue, different than the gonadal somatic cells, provides a supplementary advantage: an analysis in the head allows to determine CG7009's function in the soma, unbiased by peculiarities specific to germinal tissues.

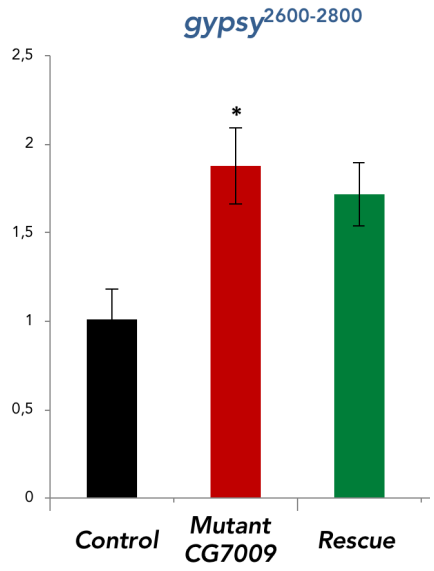

Figure 57. *gypsy* is up-regulated in the heads of *CG7009* mutants. Determination by RT-qPCR of the relative expression level of *gypsy* transcript region 2600-2800 nt in 4 days old females with the following genotypes: $CG7009^{e02001}$ / TM6, Tb, Sb, denoted as "Control" (black bar); $CG7009^{e02001}$ / Def3340, denoted as "Mutant *CG7009*" (red bar); BAC/BAC; $CG7009^{e02001}$ / Def3340, denoted as "Rescue" (green bar); and $Ago\text{-}2^{414}$ /$Ago\text{-}2^{414}$, denoted as "Mutant Ago2" (orange bar). Average values of two independent biological replicates are shown. *p-values* were determined by Student's test; *: *p-value* < 0.05; **: *p-value* < 0.01; ***: *p-value* < 0.001; ns: *non-significant p-value* > 0.05. The primers used for the RT-qPCR are MA037 and MA038 (Methods). Rp49 (primers SD22 and SD23, Methods), was used for normalization.

I performed an RT-qPCR with primers amplifying the internal *gypsy* sequence (2600-2800) in heads of 4 days old female $CG7009^{e02001}$ / Def3340 trans-heterozygous mutants (Figure 57, red bar) and $CG7009^{e02001}$/TM6, Tb, Sb heterozygotes as control (Figure 57, black bar). *Gypsy* transcripts are significantly upregulated in the *CG7009* mutant heads when compared to the control (compare red and black bars). The BAC line is uncapable to restore this up-regulation of *gypsy* (Figure 57, green bar).

In conclusion, we showed that CG7009 and CG5220 seem to promote *gypsy* expression in the ovary when the *gypsy::LacZ* sensor is not present. As both genes are involved in the siRNA pathway, this could

be due to the fact that in an impaired background for the siRNA pathway, *gypsy* transcripts are better down-regulated in the ovary. This hypothesis is also supported by the observed decrease of *gypsy* transcripts levels in Ago-2 mutants, as seen by RT-qPCR (Figure 55).

In the ovaries of *CG7009*, *CG5220* double mutants endogenous *gypsy* transcripts are de-repressed. This suggests that CG7009 and CG5220 might have a redundant function with regards to *gypsy* regulation. Moreover, the effects of CG7009 on *gypsy* transcript regulation seem to be tissue specific, as in the heads the protein seems to be required for *gypsy* repression, contrary to its effect in the ovarian follicular cells.

We were able to involve CG7009, CG5220, CG33172 and Dcr-2 in the regulation of the *gypsy-LacZ* sensor. This suggests a link of the siRNA pathway to the regulation of the *gypsy*-sensor, whose expression is considered as dependent only on the somatic piRNA pathway. Many validated piRNA pathway components were discovered by the use of the *gypsy*-sensor transgene, and we do not question its capacity to unravel novel piRNA pathway factors. The result implicating Dcr-2 in the sensor regulation however has to be reproduced in order to state unambiguously its validity.

Chapter IV: CG7009 is involved in tRNA fragmentation in flies

1. Differential tRFs accumulation in *CG7009* mutants from tRNA^{Phe}

During my PhD and during our collaboration with Dr. Matthias Schaefer, we both performed several Northern blots (NB) analyses of tRNA^{Phe} profiles in mutants for *CG7009* and *CG5220*. On my part, I also performed tRNA profiles analyses by NB on tRNA^{Trp}, as later seen target of CG5220 (see RiboMethSeq chapter) and on tRNA^{Leu}, which by now remains predicted targets of CG7009 and / or CG5220. I decided first to present one NB performed by our collaborator Dr. Schaefer (Figure 58), as some of the general conclusions on this NB are coherent with my own results.

Figure 58A, represents a NB on total RNAs extracted from 4 days old males of the indicated genotypes. w^{1118} was used as control genotype. Three different mutants for *CG7009* were analyzed: the cleaned allele *CG7009* (G10), and the two trans-heterozygous mutants *CG7009*^{e02001} / Def3340 or Def9487. *CG5220* catalytically dead mutant (K28A) and the double mutant *CG7009 (G10), CG5220*^{K28A} Rec1 were also analyzed.

The same genotypes are represented on Figure 58B, with the difference that the flies were subjected to 1 hour heat-shock at 37°C, followed by 3 hours of recovery at 25°C before they were sacrificed. 5S rRNA served as loading control in both experiments.

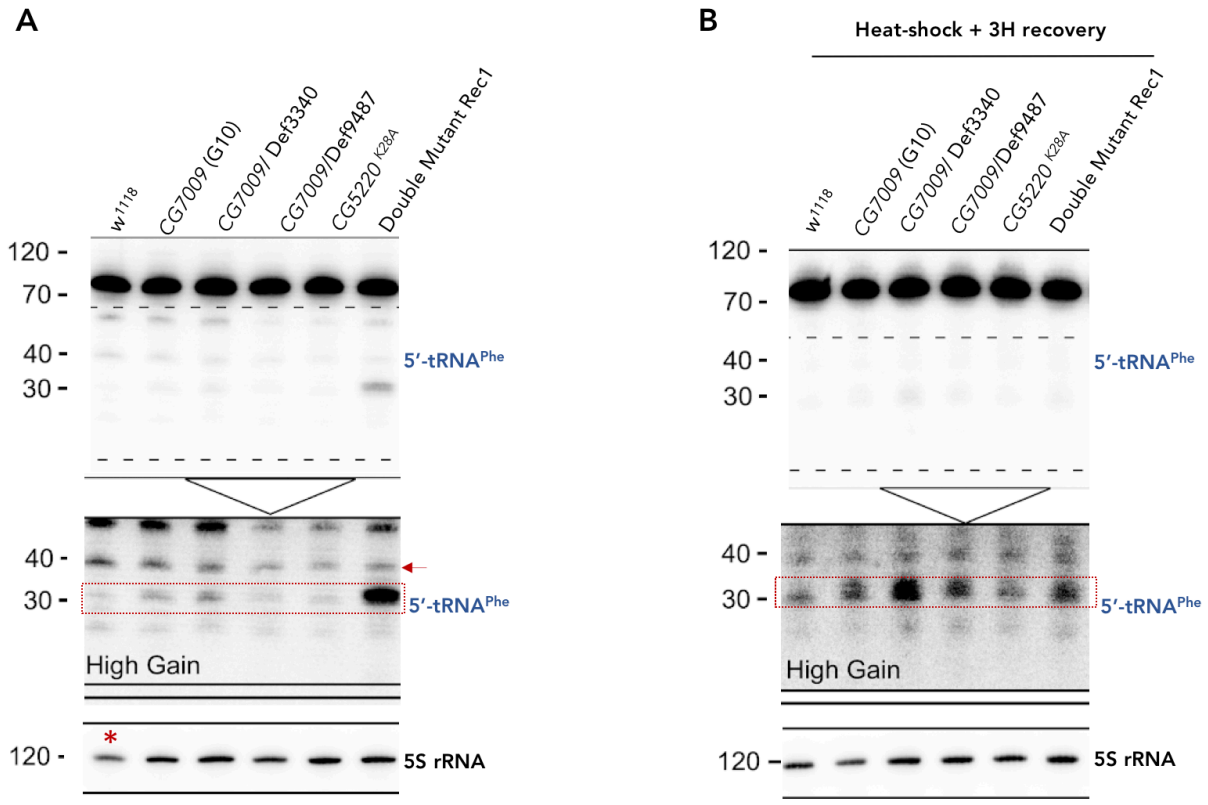

Figure 58. Northern blot characterization of 5'-tRNA^{Phe}-derived tRFs. Northern blot on the indicated genotypes was performed on control (**A**) or heat-shocked flies (1 hours at 37°C on water bath), collected after 3 additional hours of recovery at 25°C (**B**) using a 5'-tRNA^{Phe} specific probe, upper panels (Methods) and 5S rRNA probe as loading and transfer control, lower panel. Mature tRNA^{Phe} size is ~70nt, (*upper panels*, **A** and **B**). 5'-tRNA^{Phe}-derived tRNA fragments (5'-tRFs^{Phe}) were detected at ~35nt (*red doted squares*, **A** and **B**) and at ~45nt (*red arrow*, **B**). nt: size in nucleotides.

On the upper panel of the NB presented in Figure 58, we can see the full-length tRNA^{Phe} detected with a 5'-tRNA^{Phe}-specific probe. A first observation is that mutation of the indicated genes seems not to affect, at least drastically, the expression of full-length tRNA^{Phe}. This suggests that if the effects we characterized in our study on CG7009 and CG5220 are due to translational problems, they will be probably not caused by a severe decrease in the mature tRNA^{Phe} levels.

On the NB on Figure 58A, a first observation is the appearance of a band migrating slightly higher than 30 nt in the double mutants *CG7009 (G10), CG5220* (Figure 58A, red doted square). The appearance of this band of 30-35 nt, detected strongly in the double mutants is linked to a decrease of a band migrating at around 40 nt (Figure 58A, compare lanes 1 and 6, red arrow). It seems that, at least in the homozygous CG7009 G10 mutants and trans-heterozygous mutants *CG7009^{e02001}* / Def3340 (Figure 58A, lanes 2 and 3), the band of 30-35 (red doted square), well seen in the double mutants, is increased in comparison to control. However, it is hard to clearly conclude especially because the loaded material in w^{1118} (Figure 58A, red star) was lower than in the different mutants combinations on the blot.

However, the increase of the band of 30-35 nt becomes more obvious on the NB on Figure 58B, (red doted square) performed on heat-shocked flies. On Figure 58B we clearly detect more 5'-tRNA^{Phe}-derived fragments (5'-tRF^{Phe}) at 30-35 nt in the mutants for *CG7009* and in the double mutants *CG7009, CG5220*, compared to *CG5220* mutants and flies or the strain w^{1118}. On the contrast in the NB on heat-shocked flies is harder to deduce if there is a decrease of the 5'-tRF^{Phe} of approximately 40 nt.

The profiles of 5'-tRNA^{Phe} derived fragments are very complex, but several important conclusions from these NBs agree with my previous observations.

(i) In *CG7009* mutants and in the double mutants *CG7009*, *CG5220* Rec1 we detect a differential accumulation of tRNA^{Phe} 5'-derived fragments (5'-tRF^{Phe}).

(ii) The accumulation of 5'-tRF^{Phe} is not accompanied by decrease in the levels of the corresponding mature tRNAs.

(iii) A stabilization of a distinct 5'-tRF^{Phe} of 35 nt is seen in *CG7009* mutants and double mutants *CG7009*, *CG5220* Rec1.

(iv) This 35 nt 5'-tRF^{Phe} stabilization seems to be accompanied by a decrease of a 5'-tRF^{Phe} of 40-45 nt (suggesting that the 40-45 nt band is a precursor of 35 nt).

My experiments showed an effect of CG5220 only once during the reproduction of the experiment with 4 different independent biological replicates, each from 4 days old females (data not shown). In the rest of the cases the profiles of tRNA^{Phe} fragmentation in *CG5220* mutants were the same as the profiles of heterozygous flies $CG7009^{e02001}$ / TM6, Tb, Sb, which I used as control. This indicated that some variability is present in the accumulation of 5'-tRF^{Phe} fragments. I wasn't able to make a clear conclusion on the implication of CG5220 in tRNA^{Phe} fragmentation, but the data strongly suggests that CG5220 is not as implicated in differential accumulation of 5'-tRFs^{Phe} as CG7009. On the contrary, several reproducible important observations were made for the role of CG7009 on the fragmentation of tRNA^{Phe}.

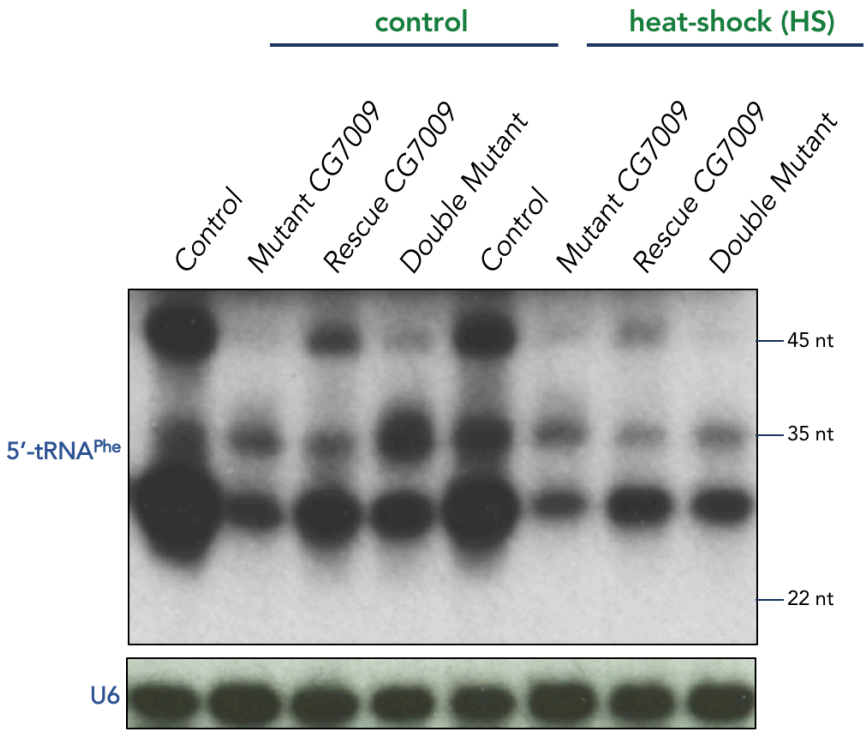

Figure 59. Northern blot characterization of 5′-tRNA^{Phe}-derived tRFs. Northern blot was performed on 10 µg of total RNAs extracted from 4 days old females with the following genotypes: trans-heterozygotes Control flies (*CG7009^{e02001}* / TM6, Tb, Sb), Mutants for *CG7009* (*CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$*), Rescue *CG7009* (BAC / BAC ; *CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$*) and Double Mutants (*CG7009$^{e0200\text{-}G101}$*, *CG5220^{K28A}* / *CG7009$^{e02001\text{-}G10}$*, *CG5220^{K28A}*). The NB was performed on control (*lanes 1 to 4*) and on heat-shocked flies (*lanes 5 to 8*). For the HS condition, flies were incubated 1 hour at 37°C on water bath, followed by 5 hours of recovery at 25°C. The NB was performed using a 5′-tRNA^{Phe} specific probe, and U6 snRNA probe as loading and transfer control (Methods). The mature tRNA^{Phe} is not presented. 5′-tRNA^{Phe}-derived tRNA fragments (5′-tRFs^{Phe}) are detected at ~35nt and at ~45nt. nt: size in nucleotides.

The result on Figure 59 represent a NB on total RNAs extracted from females with the indicated genotypes hybridized with the same specific 5′-end of tRNA^{Phe} probe as in Figure 58 and with a U6 probe as loading control (Methods). The full-length tRNA^{Phe} is not represented.

In the heterozygous individuals *CG7009^{e02001}* / TM6, Tb, Sb, used as control genotype I detected a massive accumulation of diverse 5′-tRFs^{Phe} (Figure 59, lane 1). A 5′-tRF^{Phe} of 45 nt disappears in *CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$* homozygous flies, and a distinct fragment of approximately 35 nt is stabilized (Figure 59, lane 2).

This phenomenon seems to be partially restored by the rescue line BAC / BAC ; *CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$* (Figure 59, lane 3). The disappearance of the 5′-tRF^{Phe} of 45 nt is also detected in double

mutants *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} / *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A}, and the accumulation of the 35 nt fragment is even more pronounced than in the single mutants for *CG7009* (Figure 59, lane 4).

On the right part of the membrane of this NB (Figure 59, lanes 5 to 8) are loaded RNAs from females of the same age, which were subjected to heat-shock. The tendencies concluded for the non-heat-shocked flies with the different genotypes are similar for the flies subjected to heat-shock. On this stage of the project we have focused our attention on the condition non-heat-shocked flies.

2. Differential tRFs accumulation in *CG7009* mutants from tRNA^{Leu}

As we do not yet have any link between CG7009 or CG5220 with the third described substrate of Trm7, tRNA$^{Leu\text{-}TAA}$ (Pintard et al., 2002), I decided to join my NB result analyzing the profiles of fragments derived from the 3'-end of tRNA^{Leu} of *Drosophila* (Figure 60).

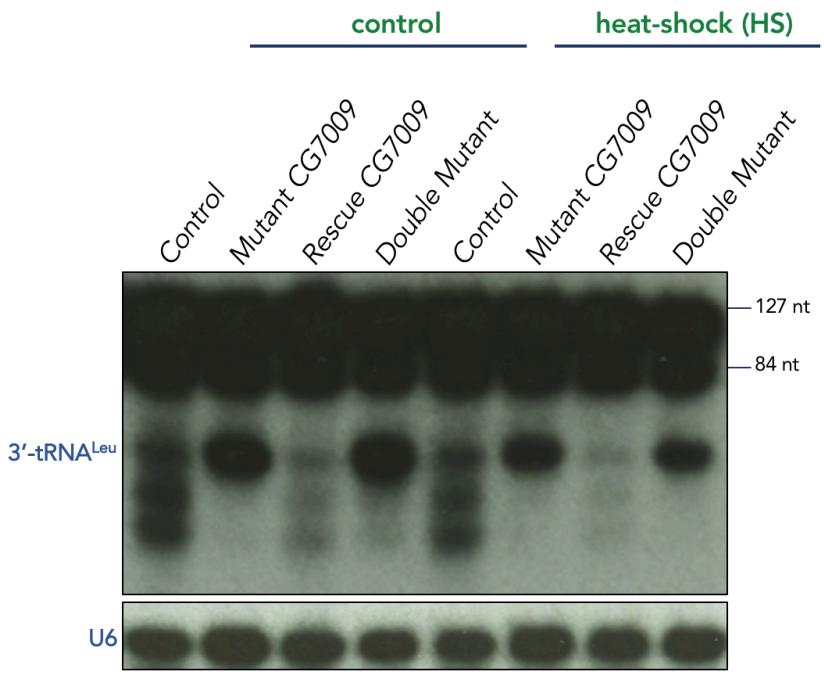

Figure 60. Northern blot characterization of 3'-tRNA^{Leu}-derived tRFs. Northern blot was performed on 10 µg of total RNAs extracted from 4 days old females with the following genotypes: trans-heterozygotes Control flies (*CG7009*^{e02001} / TM6, Tb, Sb), Mutants for *CG7009* (*CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$), Rescue *CG7009* (BAC /BAC ; *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$) and Double Mutants (*CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} / *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A}). The NB was performed on control (*lanes 1 to 4*) and on heat-shocked flies (*lanes 5 to 8*). For the HS condition, flies were incubated 1 hour at 37°C on water bath, followed by 5 hours of recovery at 25°C. The NB was performed using a 3'-tRNA^{Leu} specific probe, and U6 snRNA probe as loading and transfer control (Methods). Mature tRNA^{Leu} (87 nt) and pre-tRNA^{Leu} (127 nt) are detected. The NB represents the same membrane as in Figure 59, re-probed for 3'-tRNA^{Leu}.

The result on Figure 60 represent a NB on total RNAs extracted from females with the indicated genotypes hybridized with a probe designed against the 3'-end of $tRNA^{Leu}$ and an anti-U6 probe as loading control. As the 3'-end of $tRNA^{Leu}$ is conserved between the different $tRNA^{Leu}$ isoacceptors not only $tRNA^{Leu\text{-}TAA}$ is detected on the NB. This explains the two bands, detected for the full-length $tRNA^{Leu}$ (Figure 60, 127 and 84 nt), as the intron-containing pre-$tRNA^{LeuCAA}$ (127 nt) is also detected.

In the heterozygous individuals *CG7009*^{e02001} / TM6, Tb, Sb, used as control and in the rescue line BAC / BAC ; *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ (Figure 60, lane 3) a smear with no distinct fragment is visible (Figure 60, lane 1). In *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ homozygous flies and in double mutants *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} / *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A}, we observe a distinct fragment accumulation, not seen in the heterozygous flies nor in the rescue line.

The next part of the membrane (Figure 60, lanes 5 to 8) is the corresponding flies this time subjected to heat-shock. The tendencies concluded for the non-heat-shocked flies with the different genotypes are similar for the flies subjected to heat-shock.

This $tRNA^{Leu}$ profile of fragmentation is even harder to understand as 2 different isoacceptors of $tRNA^{Leu}$ are revealed. However, this is clear that, the observed profile is changing between the WT and the *CG7009* mutant conditions.

3. Dcr-2 is involved in the fragmentation of $tRNA^{Phe}$ in *CG7009* mutants

It is known that Dicer proteins are involved in tRFs generation (Cole et al., 2009; Martinez et al., 2017). We decided thus to see if the fragments accumulating in *CG7009* mutants are dependent on Dcr-2 activity. For this, I analyzed by NB the 5'-$tRFs^{Phe}$ in *Dcr-2, CG7009* double mutants.

In the NB result on Figure 61, I compared 5'-$tRNA^{Phe}$ fragmentation profiles in heterozygous females *CG7009*$^{e02001\text{-}G10}$ / TM6, Tb, Sb as control (lane 1), *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ homozygous mutants (lane 2) and double mutants Dcr-$2^{L811fsX}$ / Dcr-$2^{L811fsX}$; *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ (lane 3) to challenge the loss of Dcr-2 in combination with a *CG7009* mutant background on tRFs profiles.

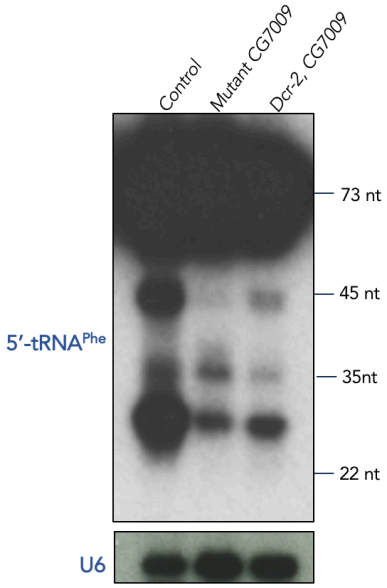

Figure 61. Northern blot characterization of 5′-tRNA^{Phe}-derived tRFs. Northern blot was performed on 10 µg of total RNAs extracted from 4 days old females with the following genotypes: trans-heterozygotes Control flies (*CG7009^{e02001}* / TM6, Tb, Sb), Mutants for *CG7009* (*CG7009$^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$*), and Double Mutants (*Dcr-2^{L811fsX}* / *Dcr-2^{L811fsX}* ; *CG7009 $^{e02001\text{-}G10}$* / *CG7009$^{e02001\text{-}G10}$*). The NB was performed using a 5′-tRNA^{Phe} specific probe, and U6 snRNA probe as loading and transfer control (Methods). The mature tRNA^{Phe} is at 73 nt. 5′-tRNA^{Phe}-derived tRNA fragments (5′-tRFs^{Phe}) are detected at ~35nt and at ~45nt. nt: size in nucleotides.

On Figure 61 we can see a strong accumulation of 45 nt fragments in the heterozygotes as seen (lane 1). In *CG7009* mutants the fragment of 45 nt is decreasing and a band corresponding to 35 nt looks stabilized (lane 2). Surprisingly, the simultaneous mutation of *dcr-2* and *CG7009* partially restores the wild-type fragment's profile (lane 3).

The reappearance of the fragment of 45 nt as well as the decrease of the band at 35 nt indicates that Dcr-2 endonuclease activity can be responsible for the cleavage of this 45 nt fragment in *CG7009* mutants and for the consecutive accumulation of the smaller fragment of 35 nt.

It will be very interesting to see if the accumulation of the 35 nt or the decrease of the 45 nt fragment is causative for the deregulation of the sncRNA pathways in *CG7009* mutants. It is possible that in the absence of CG7009, its tRNAs substrates are de-stabilized and probably fragmented by Dcr-2. If Dcr-2 gets titrated for processing the tRFs observed in *CG7009* mutants, maybe it becomes enough "unavailable" to process long dsRNAs to siRNAs. We named this hypothesis the "saturation theory". The result on Figure 61, is the first genetic link between CG7009 and Dcr-2, which favors the saturation theory of Dcr-2.

CONCLUSION AND PERSPECTIVES

Methylation of sncRNAs

RNA 2'-O-methylation, a post-transcriptional modification added to the 3'-terminal nucleotides of sncRNAs directly impacts their stability (Kamminga et al., 2010; Li et al., 2005; Lim et al., 2015b). In flies piRNAs and Ago-2-loaded small RNAs: mi- and siRNA are methylated by DmHen1/Pimet. Dm-Hen1/Pimet, as well as its orthologs in other species, are the only known methylases which function in the sncRNA pathways. In regards to the important biological processes mediated by the sncRNA pathways, such as regulation of gene expression, antiviral defense and preservation of the genomic integrity *via* the repression of transposons, it is not surprising that their deregulations have been associated with multiple phenotypes, like increased susceptibility to viral infection or infertility. Hen-1 proteins mediated methylation is conserved throughout the evolution, suggesting its importance. It is thus surprising that loss-of-function on DmHen-1 in *Drosophila* is not associated with the aforementioned "classical" phenotypes, characteristic of a loss-of-function of the major actors of the sncRNA pathways. In flies *DmHen*-1 mutations are not linked with a major decrease of the sncRNA populations. The loss of the protein has been associated with shortening of piRNAs, which in turn depends on the age of the flies (Wang et al., 2016). However, DmHen-1 mutant flies have a good fitness and are fertile. In contrast, in zebrafish and mice, loss of function of DmHen-1 orthologs (hen1 and HENMT1) has been associated with female and male sterility, respectively (Kamminga et al., 2010; Lim et al., 2015a, b). It is thus puzzling why DmHen-1, which maintains its function to methylate sncRNAs in *Drosophila*, is not associated to any destabilization of its target small RNA populations, or to phenotypes such as fertility problems. An interesting observation comes from a pioneer study, which described the activity of DmHen-1 by *in vitro* analysis of its encoded protein catalytic activity (Horwich et al., 2007). The authors showed that unknown factor(s) contained in flies extracts strongly enhances DmHen-1 recombinant protein enzymatic activity *in vitro*. This opened the possibility that in flies, (a) protein(s) redundant to DmHen-1 might compensate its loss-of-function.

Methylation of tRNA and tRFs

RNA methylation is also important for the stability of tRNAs. Transfer RNAs are the most heavily modified RNAs throughout species. These modifications are not only required to maintain their structure, they also allow to avoid frame shifting during translation, cleavage of tRNAs by endonucleases and contribute to the diversity of nucleotides. tRNA m^5C methylases, like NSUN2 and Dnmt2 have also been shown to be important for the stability of tRNAs, as their loss is linked to differential tRNA-derived fragments (tRFs) accumulation (Blanco et al., 2014; Schaefer et al., 2010). Both proteins are also linked to through unknown exact mechanism to the sncRNA pathways (Durdevic et al., 2013b; Genenncher et al., 2018). Thus, there seems to be a correlation between the role of RNA methylation for the stability of tRNAs and the sncRNAs. But an interesting observation arises: when tRNAs lose their methylation, tRFs are produced. Research on tRFs in recent years showed that these molecules can function as sncRNA-like molecules, by binding to partners of canonical sncRNAs, *i.e.* Dicer or Argonaute proteins, by regulating similar targets, *i.e.* TEs, or by regulating similar biological processes and being involved in the emergence of common pathologies, *i.e.* cancer, neuronal disorders and viral resistance.

CG7009 and CG5220

During my PhD I characterized the molecular functions of CG7009 and CG5220, two novel tRNA 2'-O-methyltransferases, which have conserved orthologs from yeast to human, Trm7 and FTSJ1, respectively. I was able to link both CG7009 and CG5220 to the regulation of the sncRNA pathways. Additionally, CG7009 is also associated to differential tRFs accumulation. My PhD studies thus unraveled two novel aspects of these processes. On one hand we show that tRNA methylases catalysing different post-transcriptional modification than m^5C, like 2'-O-methylation, seem also involved in the regulation of the sncRNA pathways efficiency. We also show a first report of a 2'-O methyltransferase distinct from DmHen-1 and which impacts sncRNAs. On the other hand, our results suggest that there is also a link between 2'-O-methylation and differential accumulation of tRFs.

Involvement in the sncRNA pathways

Historically we started working on CG7009 because of its implication in the sncRNA pathways. Before my arrival in the laboratory experiments had shown that this gene is involved in the siRNA and the Ago-2-dependent miRNA pathways in cells. During my PhD, we were able to validate these results, as well as to obtain new results that CG7009 is also involved in the somatic piRNA pathway and in the Ago-2-miRNA silencing in adult flies. During my project I discovered that CG7009 has a potential paralog, CG5220. Further experiments allowed us to show that, similarly to CG7009, CG5220 is also involved in Ago-2 miRNA silencing and in the somatic piRNA pathways in flies.

Mutants Characterization

Discovering the function of CG7009 and CG5220 was a long process, which led to several supplementary developments and observations. Indeed, in the process of their characterization multiple tools were constructed and a molecular characterization of several publicly available mutants was performed. I precisely characterized the insertion site of a transposon, PBac{RB}, present in the *CG7009*^{e02001} mutant allele and confirmed that its sequence in the BL18008 fly stock was identical to the one of the reference genome. I determined the expression levels of the gene's transcript and concluded that *CG7009*^{e02001} is a null allele of *CG7009*, as no mRNA is detectable in the homozygous mutant. I also established a *Drosophila* rescue line (BAC) that carries the wild type allele of *CG7009* surrounded by 21kb of its genomic region inserted in the second chromosome. However, due to some discussed complications with this line, we are currently establishing a UAS-CG7009 rescue line, as well as a UAS-CG5220 rescue line, in order to reproduce some crucial experiments as well as to unambiguously validate the effects of the proteins of interest on the determined processes.

Importantly, with the help of Dr. Schaefer, we established three new mutant alleles for the *CG5220* gene using the CRISPR/Cas9-mediated mutagenesis: (i) one allele bearing a 2-nucleotide deletion and leading to a frameshift mutation with a premature stop codon, potentially leading to a truncated protein from this allele. Further functional assays will determine whether this allele is null for CG5220 function. (ii) *CG5220*^{K28A}, a catalytically dead allele, bearing a substitution in the active center of CG5220. The use of this mutant allele made the characterization of the molecular function of CG5220 possible. (iii) *CG5220*^{A26P}, an allele bearing the A26P substitution. This particular substitution is interesting because a corresponding

substitution in FTSJ1, the human ortholog of CG5220 and CG7009, is sufficient to provoke a neuronal disorder, non-syndromic X-linked intellectual disability (NSXLID) in its carriers (Guy et al., 2015).

Interestingly in humans, this allele is linked to loss of only one of the two positions methylated by FTSJ1 on its substrate $tRNA^{Phe}$: the wobble position of the anticodon, position $G_{34,}$ while the other positions, methylated on C_{32} by FTSJ1, remained intact. This suggested a higher importance of this position methylation for ID pathogenesis in human. While during my PhD we managed to construct the corresponding allele for the gene CG5220, we are currently establishing the corresponding alleles for *CG7009*, *i.e.* $CG7009^{A26P}$ and $CG7009^{K28A}$. We are interested to investigate if, similarly to FTSJ1 in humans, there will be molecular consequences of this A to P substitutions in flies on the targets of the CG7009 and CG5220 proteins (which were identified in this study).

Molecular Function

During my PhD we discovered an important difference when comparing the modification of C_{32} and G_{34} of $tRNA^{Phe}$ between flies and human (as well as yeast). Indeed, while in yeast and human Trm7 and FTSJ1 (respectively) catalyze Cm_{32} and Gm_{34} formation on $tRNA^{Phe}$ (Guy et al., 2015; Pintard et al., 2002), I was able to show that *Drosophila* evolved two paralogs CG7009 and CG5220, which diverged and specialized in the modification of distinct nucleotides in $tRNA^{Phe}$: CG7009 is responsible for Nm deposition at the wobble position 34 of the anti-codon loop of $tRNA^{Phe}$, while CG5220 is involved in deposition of Nm at position 32. These interesting findings were confirmed in several different experiments: CG7009's function on Gm_{34} was shown by a MALDI TOF analysis preceded by RNase T1 digestion, by a RiboMethSeq analysis and by a LC-MS analysis; CG5220s function on the methylation of C_{32} of $tRNA^{Phe}$ Nm was shown by RNase A digestion coupled to MALDI TOF and by RiboMethSeq.

Analysis of the distribution of Nm on tRNAs in flies by RiboMeth seq further indicated that, apart C_{32} in $tRNA^{Phe}$, CG5220 is also methylating C_{32} in $tRNA^{Trp}$, a conserved target of Trm7 and FTSJ1 (Guy et al., 2015; Pintard et al., 2002), as well as methylating C_{32} in $tRNA^{Gln}$ and in $tRNA^{Glu}$ isoacceptors (data not shown). The latter two tRNAs haven't been characterized before as targets of the conserved Trm7 methyltransferase sub-family. Taken together, these results support a substrate and position specificity of CG5220 in *Drosophila* for tRNAs having a cytosine at position 32.

In the future, it will be interesting to gain more information on what factors influence the different specificities of these two proteins in flies. One could wonder for example if it is a position specificity, meaning that CG7009 will preferentially methylate position 34 and CG5220 will "prefer" position 32? Conversely, their differing specificity could be base-related, *i.e.* CG7009 has more affinity for G, while CG5220 would have more affinity for C? Ultimately, it could be a combination of both hypotheses. Further investigation of the effects of *CG7009* and *CG5220* mutations on another of their predicted targets can give hints to answer these questions.

On the molecular level $tRNA^{Trp}$ is very interesting, because it has a C_{32} and a C_{34}. As RNase A cuts after a C or U if not Cm or Um, it is possible to use the same approach as for $tRNA^{Phe}$ – RNaseA digestion coupled to MALDI TOF. By analyzing $tRNA^{Trp}$ modifications in *CG7009* and *CG5220* single mutants, as well as in the *CG7009, CG5220* double mutant, we would be able to deduce more details on whether CG7009 is specific for position 34 or for the base G, and similarly for CG5220 and the position 32 and C. If CG5220 methylates both C_{32} and C_{34} in $tRNA^{Trp}$ this divergence between CG7009 and CG5220 would be based on base specificity; on the contrary if CG7009 methylates C_{34} and CG5220 C_{32} in $tRNA^{Trp}$, the specificity difference between the two proteins will be more probably based on position. However, the preliminary analysis of the RiboMethSeq on the modification of position 34 of $tRNA^{Trp}$, which is also a C, suggests that, in contrast with Trm7 and FTSJ1, neither CG7009 nor CG5220 are responsible for catalysis of this modification.

A fascinating question is if the functional divergence of CG7009 and CG5220 in flies is linked to their implication in the sncRNA pathways. Indeed, as discussed in the Introduction, proteins involved in the sncRNA pathways, such as Ago-2 and Dcr-2, are amongst the most rapidly evolving genes (Obbard et al., 2006). Another example is Moonshiner, the paralog of the subunit of TFIIA, specifically involved in the transcriptional initiation from dual-strand piRNA clusters (Andersen et al., 2017). It is thus possible that, during evolution, the selection pressure to respond to rapidly evolving viruses or retroelements has selected a duplication and divergence of those two paralogs of Trm7 and FTSJ1 in flies.

<u>tRFs accumulation</u>

During my PhD, we also established and characterized two double mutants for *CG7009, CG5220* by using two new alleles of *CG5220* ($CG5220^{K28A}$ and $CG5220^{A26P}$). The experiments described in this

manuscript were performed with the double mutants *CG7009(G10), CG5220^{K28A} (Rec1)*, which bear the catalytically dead for *CG5220* allele *CG5220^{K28A}*. A Small RNA-Seq was performed on RNAs extracted from the ovaries and the heads of these double mutants, as well as from single *CG7009* mutants and control flies. Interestingly, in a preliminary analysis of this experiment, I was able to detect differentially expressed tRFs in *CG7009* single mutants, as well as in *CG7009, CG5220* double mutants. I also detected differentially expressed tRFs from tRNA^{Glu} and tRNA^{Gln}, the two novel unexpected targets of CG5220, before having the RiboMethSeq results. It will be interesting to investigate if there is a correlation between the affected modifications on the different tRNAs, as detected by the RiboMethSeq, and the differentially expressed tRFs observed in the small RNA sequencing. If this is the case, such analysis will solidify our discovery that Nm RNA modification on tRNAs is involved in the accumulation of differential patterns of tRFs.

I analyzed the differential level of tRFs from the three predicted targets of CG7009 and CG5220: tRNAs Phe, Trp and Leu-TAA, by Northern blot. This allowed me to conclude that differential tRFs, derived from the 5'-end of tRNA^{Phe} and from the 3'-end of tRNA^{Leu}, accumulate in a CG7009-dependent manner. Interestingly, I also found that the production of 5'-tRNA^{Phe}-derived fragment of 35 nucleotides in *CG7009* mutants is linked to Dcr-2, the key protein of the siRNA pathway. The result from this NB not only provided the first proof of a genetic link between CG7009 and Dcr-2, but also favored one of the currently most studied theories in my lab: the "saturation theory".

<u>Saturation theory</u>

Indeed, we are very interested to understand if the tRFs observed in *CG7009* mutants are causative for the deregulations of the sncRNA pathways that we described. It is possible that the tRFs "saturate" the proteins of the sncRNA pathways, notably Argonaute proteins, as they are common for the three pathways, making them unavailable for their classical substrates, the small RNAs. To test if CG7009 has a function in the Ago-2 and in the Piwi mediated silencing, we are currently immunoprecipitating these two proteins and are investigating their bound small RNAs. A recent result of Dilyana Dimitrova, a second year PhD student in the lab, indicated that 5'-tRFs from tRNA^{Phe} are loaded into Ago-2 in flies. We next aim to compare if in *CG7009* and *CG5220* mutant backgrounds this detected tRFs accumulate and if this is linked to a decrease in the levels of classical sncRNAs. As a simple experiment to determine if the tRFs are

triggering the deregulation of the sncRNA pathways, we can transfect cells expressing the automiG sensor with those specific tRFs from $tRNA^{Phe}$ and analyze if the sensor becomes de-repressed.

We are currently also challenging the saturation theory in flies by overexpressing Piwi and Ago-2 in a *CG7009* mutant background and in the presence of the *gypsy-LacZ* sensor and a version of the automiG adapted for expression of the GFP in adult flies. If for example overexpression of Piwi is sufficient to abolish the effect of the *CG7009* mutation on *gypsy*-LacZ, this will strongly suggest that Piwi is saturated in a *CG7009* mutant background.

The "saturation theory" could explain, at least partly, the effects of CG7009 loss-of-function on the siRNA and Ago-2-dependent miRNA pathways. It might be difficult to argue that one specific fragment will be capable to titrate the small RNA pathways proteins, *i.e.* Dcr-2 as suggested by the Northern blot result in Figure 61, strongly enough abolish the pathways in order to obtain the observed effects with the automiG and automiW genetic sensors. But, as we saw, *CG7009* mutant does not impact the fragmentation profile only of 5'-$tRNA^{Phe}$. *CG7009* mutation also induces accumulation of a fragment derived from $tRNA^{Leu}$. This strengthens our discovery, that tRNA 2'-O-methylation, similarly to m^5C tRNA methylation, is associated with changes in tRNA fragmentation in flies. It will be also interesting to analyze if Dcr-2 function is linked to the accumulation of this fragment.

However, the preliminary analysis a small RNA-Seq result suggests that the mechanisms through which CG7009 impacts the sncRNA pathways are more complicated than the saturation theory. As the analysis of this experiment is still ongoing I will not show the data but will state several hypotheses based on the preliminary observations. First, there seems to be no general decrease of sncRNAs populations linked to CG7009 loss-of-function, although the sncRNAs are expected to be destabilized if they are not bound to their interacting Argonaute partner, wich argues against a general saturation theory. A second observation is that an analysis of the differentially expressed miRNAs, as well as small RNAs matching TE sequences (thus endo-siRNAs and piRNAs), suggests that individual small RNAs are not only down-regulated, as it would be expected in the case of a "simple" saturation of the Argonautes by the tRFs. Indeed, many of the sncRNAs in this preliminary analysis seem to be up-regulated in *CG7009* mutants and double mutants *CG7009*, *CG5220*. I found this observation fascinating and tried to understand how some small RNAs, *i.e.* piRNAs and endo-siRNAs can be up-regulated in *CG7009* mutants and double mutants *CG7009*, *CG5220* while others are down-regulated. I thought about two possibilities, remaining as open questions: (i) Is there a tendency that all the up-regulated piRNAs are always derived from the same

clusters, while all the down-regulated piRNAs would derive from different clusters? It will be interesting to analyze if there is such a correlation between the expression tendency and the cluster of origin. This could suggest the existence of a regulation network between the clusters, in which, when one specific piRNA-producing cluster is downregulated, another cluster will be up-regulated. (ii) Is it possible that CG7009 impacts the populations of maternally deposited piRNAs in the next generation? This could lead to redistribution of heterochromatin and provide a potential explanation for such a diverse differential expression pattern of the small RNAs mapping to TE sequences. The future validation of my small RNA-Seq analysis by Dr. Christophe Antoniewski will shed more light on the global role of *CG7009* and *CG7009, CG5220* double mutants on the sncRNAs biogenesis and identities.

Hypothesis to explain CG7009 and CG5220 effects on gypsy transcription in flies' ovaries

The results linked to the analysis of the transcriptional expression of endogenous *gypsy* showed repression of *gypsy* transcript in the ovaries of *CG7009* and *CG5220* mutants compared to the control flies. In the same time *CG7009* mutants were shown to de-repress the *gypsy::LacZ sensor* as well as endogenous *gypsy* transcription in flies having a *gypsy-LacZ* sensor. My experiments suggested that the presence of the *gypsy::LacZ* transgene could play a role in this phenomenon of inverse effect of *CG7009* mutation on *gypsy* transcription in the ovaries of flies having or not this transgene. Two hypotheses outlined below could explain the effect of CG7009 on *gypsy* transcription in the two different cases:

Hypothesis 1:

In the Drosophila germline, TE repression by the piRNA pathway is more efficient when the siRNA pathway is compromised. CG7009 and CG5220 functions are rendering gypsy repression more efficient, either by abolishing the siRNA pathway or through another unknown mechanism.

As discussed in the Introduction, there is a cooperation between the si- and the piRNA pathway in *Drosophila* somatic tissues for the regulation of several TEs. CG7009 loss-of-function impairs the siRNA pathway, as we were able to show in Drosophila S2 cells. Is it then possible that in the ovaries of *CG7009* mutant flies (not having the *gypsy::LacZ* sensor) the piRNA pathway more efficiently represses *gypsy* in this siRNA pathway impaired background, which could explain our RT-qPCR results showing a down-regulation on endogenous *gypsy* (Figure 52)? This hypothesis is also supported by the observed down-regulation of *gypsy* transcription in ovaries of *ago-2* mutants, as seen of Figure 55.

Thus, *CG7009* and *CG5220* mutations in the ovaries is linked to down-regulation of the transcript levels of *gypsy*. This indicates that CG7009 and CG5220 are 2 tRNA 2'-O-methylases whose normal function seems to promote *gypsy* de-repression in the ovaries of *Drosophila*.

Hypothesis 2:

gypsy::LacZ triggers silencing of endogenous gypsy sequences via an siRNA pathway dependent mechanism. CG7009 disrupts this sensor-initiated silencing.

How to explain the up-regulation of endogenous *gypsy* and of the *gypsy::LacZ* sensor in *CG7009* mutants expressing the *gypsy::LacZ* sensor? A clue comes from the study on the initiation of the transcription from dual-strand piRNA clusters by Andersen and colleagues (Andersen et al., 2017). The authors showed that, when they delete the promoter of a piRNA cluster adjacent gene locus *Pld*, the piRNA cluster seems to extend into the promoter-less *Pld* locus. Moreover, the authors showed that deleting the promoter of the *Pld* locus is associated with a sharp increase of siRNAs, corresponding to the *Pld* sequence and only a slight increase of piRNA sized small RNAs mapping to this genomic region. However, a sharp increase in piRNAs corresponding to *Pld* sequences was observed when the authors simultaneously deleted the promoter and introduced an ectopically expressed cDNA of *Pld*. If we look at the *Pld* cDNA as a "target" of those *Pld*-locus-derived piRNAs, their result indicated that the presence of a target can influence the level of piRNAs produced against that target (here the *Pld* sequences).

A similar effect of the homologous target, acting in *trans*, in a phenomenon of temperature-induced epigenetic conversion of a non-piRNA producing cluster to piRNA producing cluster was recently described by Casier *et al*. (https://www.biorxiv.org/content/early/2018/07/03/360743).

Let's consider the *gypsy::LacZ* sensor as a "target", which similarly to the cDNA of *Pld* induces the production of more piRNAs and/or endo-siRNAs involved in its own silencing, as well as of homologous sequences. We can name this phenomenon *target-induced silencing*. Then the de-repression of *gypsy::LacZ* and endogenous *gypsy* sequences in *CG7009* mutants might be due to a role of *CG7009* in abolishing this *target-induced silencing*.

However, how can we explain that apart from the sequence 1-700, which is present on the *gypsy*::LacZ sensor, other *gypsy* sequences (*i.e.* 2600-2800), not contained in this transgene, are also overexpressed in *CG7009* mutants expressing the sensor?

A study on the genetic interaction between a novel *gypsy* element inserted in a euchromatic site and a *flamenco* permissive allele, suggested that *gypsy* elements might cause post-transcriptional silencing by the endo-siRNA pathway of *flamenco* (Guida et al., 2016). Indeed, as parts of the *flamenco* locus contains *gypsy* antisense sequences, potential dsRNA parts can form with *gypsy* TE transcripts and be processed by Dcr-2. Interestingly, the authors showed that not only *gypsy*, but also other transposons regulated by the *flamenco* locus were affected by the novel insertion of *gypsy* (Guida et al., 2016).

If the *gypsy::LacZ sensor* is "seen" by the defense mechanisms of flies as a novel *gypsy* insertions, not only the 1-700 gypsy sequence contained in the sensor, but also other sequences regulated by *flamenco* could be affected (*i.e. gypsy* region 2600-2800). This could explain the up-regulation of *gypsy* sequences in *CG7009* mutants expressing the gypsy::lacZ transgene (Figure 50C).

I thus hypothesized that *gypsy::LacZ* is triggering its own repression by *target-induced* or *novel-insertion* induced post-transcriptional gene silencing and *CG7009* mutation is abolishing this silencing through a not known yet mechanism.

The opposite effect of CG7009 and CG5220 on *gypsy* expression in the presence and the absence of the *gypsy*-sensor indicated that, for certain genes, a role in the de-repression of the *gypsy* transposon can be misleading when determined with a *gypsy-LacZ* sensor. This suggested that the sensor presence is linked to a more complicated repression system and that the sensor presence itself is not neutral in regards to *gypsy* transcripts expression. The investigation of the opposite effects of *CG7009* mutants expressing a *gypsy*-sensor on *gypsy* expression led to the proposition of a siRNA pathway-based *gypsy*-sensor induced silencing mechanism, affecting the repression of *gypsy* transposon in the ovaries. This initiated us to test for a role of Dcr-2 in the repression of a *gypsy*-sensor. Although the positive effect of Dcr-2 on the derepression of a *gypsy-LacZ* sensor is very preliminary and should be reproduced with other genetic combinations, its potential validity raises interesting questions on how many of the piRNA factors identified with this sensor function uniquely in the piRNA pathway and how many of them can be actually involved in such a putative siRNA-piRNA cooperation network? We are currently establishing an Ago2, *gypsy*::LacZ strain by recombination, in order to test the effect of another major siRNA pathway component on this piRNA pathway sensor.

It will be equally interesting to see if CG7009 also affects the expression of other *flamenco* transcripts, like ZAM or Idefix. As a ZAM-LacZ and Idefix-LacZ reporters are also available (reviewed in (Goriaux et al.,

2014b)) we are able to investigate if, similarly to *gypsy*, ZAM or Idefix will be repressed by *CG7009* and up-regulated only in the presence of the reporter.

Interacting partners

Another result obtained during my PhD was the co-IP of the recombinant FLAG-CG33172 and GST-CG7009 proteins co-expressed in bacteria. We also showed that CG33172 is involved in the repression of *gypsy-LacZ* sensor, similarly to CG5220 and CG7009. The fact that the potential yeast ortholog of CG33172 has been linked to Ty1 regulation, further suggests that if CG33172 is not the partner of CG7009 and CG5220, at least these proteins are involved in similar processes. It will be interesting to validate this result in flies extracts. For this purpose, we are currently editing the flies' genomes in order to obtain transgenic *Drosophila* lines expressing tagged versions of CG7009 and CG5220. These lines can be very useful in IP-experiments to identify interacting proteins by subsequent mass spectrometry. This tagged line will be also extremely valuable to determine the sub-cellular localization of CG7009 and CG5220, which can shed more light into the pathways and mechanisms into which these proteins are involved.

Crystallography

During my PhD I also constructed two full-length N-terminally tagged recombinant CG7009 proteins, FLAG_CG7009 and GST_CG7009 and determined the optimal purification conditions for FLAG_CG7009. I managed to produce both of these proteins. These proteins will be used in a collaboration with Dr. Julien Henri in order to try to resolve the crystal structure of CG7009. It will be interesting to also produce CG5220, as well as its human ortholog FTSJ1. The structure of the three proteins might provide further information on the structural basis underlying the diversification of those three proteins in the evolution.

Functional conservation between flies and human

The characterization of CG7009 and CG5220 performed during my PhD potentially links sncRNAs and tRFs biology to human physiology and pathophysiology *via* their high orthology with the human FTSJ1 protein.

Patients carrying mutation in the FTSJ1 methylase suffer NSXLID. In our laboratory we initiated a work on LCLs bearing a mutation in FTSJ1 derived from such patients. We are currently trying to determine if, similarly to what I see in flies, there is an accumulation of tRFs in these cell lines. If so, this tRFs accumulation can be very useful for future functional diagnosis of the disease. We also want to investigate if the sncRNA pathways are impaired, similarly to what we observe in flies. Such finding can lead to better understanding of the etiology of the disease.

A future project of the lab is also to perform small RNA-Seq in FTSJ1-deficient cells. This work could lead to the discovery of a novel role of sncRNAs and/or tRFs in the development of NSXLID, a pathology whose development is so far only attributed to tRNA methylation defects that could trigger translation failure. Our results concerning the effects produced in CG7009 and CG5220 mutants do not support a general translation failure as an underlying reason and alternatively suggest a revisiting of the mechanisms by which FTSJ1 is causing NSXLID.

I am currently writing an article on the functional characterization of CG7009 and CG5220 Nm methylases and their involvement in the sncRNA pathways in flies and tRFs biology. We were recently invited by Genes to submit our work in a special issue on "RNA Modifications" as well as by Non-coding RNA Investigation journal to submit an article on "functions of small RNA silencing in the fruit fly". Considering the originality and impact of our research in the emerging field of epitranscriptomics, we are considering the submission of a second article at longer term on the human aspect of this project to which I was also collaborating during my PhD.

MATERIAL and METHODS

Drosophila stocks

#	Lab Stock ID#	Category	Genotype	Notes
1	w1118	mutant allele	w1118	FlyBase ID FBal0018186
2	CG7009e02001	mutant allele	w1118; CG7009e02001 (mini-white)	Bloomington stock 18008
3	Def9487	deficiency for CG7009	w1118; Df(3R)ED10845, P{3'.RS5+3.3'}ED10845 / TM6C, cu1 Sb1	Bloomington stock 9487
4	Def3340	deficiency for CG7009	Df(3R)e-R1, Ki1/TM3, Sb1 Ser1	Bloomington stock 3340
5	CG7009e02001-G10	mutant allele	w1118; CG7009e02001 (mini-white)	This work, cleaned by backcross stock #2
6	BAC	rescue allele for CG7009	w1118; CH322 177K12 (Pacman)	This work, FlyBase cloneFBlc0000784
7	CG5220 K28A	mutant allele	w1118; CG5220 248.5.2 / TM3, Ser	This work
8	CG5220 A26P	mutant allele	w1118; CG5220 247.21.1 / TM3, Ser	This work
9	CG7009e02001-G10, CG5220 K28A Rec1	double mutant allele	w1118; CG7009e02001-G10 (mini w), CG5220 248.5.2 / TM6, Tb, Sb	This work
10	CG7009e02001-G10, CG5220 K28A Rec4	double mutant allele	w1118; CG7009e02001-G10 (mini w), CG5220 248.5.2 / TM6, Tb, Sb	This work
11	CG7009e02001-G10, CG5220 A26P RecJ	double mutant allele	w1118; CG7009e02001-G10 (mini w), CG5220 247.21.1 / TM6, Tb, Sb	This work
12	Piwi2	mutant allele	P{ry11}piwi2	FBal0060945
13	Dcr-2 R416X	mutant allele	w1118; dcr2R416X/Cyo	FBal0156732
14	Dcr-2L811fsX	mutant allele	yd2 w1118 P{ey-FLP.N}2; Dcr-2L811fsX; P{Dcr-2E1371K.t7.2}3	Bloomington stock 32064
15	Ago2 414	mutant allele	w1118; Ago2414 /TM6cTbSb	FBal0162855
16	nos-GAL4	GAL4 driver	w1118 ;; w+,nos>Gal4	Bloomington stock GAL4 lines
17	DA-GAL4	GAL4 driver	w1118 ;; p[w, da-GAL4]	Bloomington stock GAL4 lines
18	GMR-GAL4	GAL4 driver	P{GMR-GAL4.w-}	FBti0072862
19	tj-GAL4	GAL4 driver	P{tj-GAL4.U}	FBtp0089190
20	IR-white	RNAi white	y1 v1; P{TRiP.HMS00017}attP2	Bloomington stock: 33623
21	IR-Piwi	RNAi Piwi	w1118; UAS-IR(Piwi) CG 6122	VDRC N° stock: 22235 (GD)
22	IR-Ago2	RNAi Ago2	w1118; UAS-RNAi(Ago2)	VDRC N° stock: 49473 (GD)
23	IR-Ago2	RNAi Ago2	w1118; ; UAS-RNAi(Ago2)	VDRC N° stock: 100356 (KK)
24	IR-CG7009	RNAi CG7009	w1118; ; UAS-RNAi (CG7009)	VDRC N° stock: 27789 (GD)
25	IR-CG5220	RNAi CG5220	w1118; ; UAS-RNAi (CG5220)	VDRC N° stock: 34972 (GD)
26	IR-CG5220	RNAi CG5220	w1118; ; UAS-RNAi (CG5220)	VDRC N° stock: 108672 (KK)
27	IR-CG33172	RNAi CG33172	w1118; P{KK102903}VIE-260B	VDRC N° stock: 100006 KK
28	automiW	genetic sensor	UAS-automiW (w+)	(Besnard-Guerin et al., 2015)
29	gypsy lac Z	genetic sensor	R ; tjgal4 / Cyo ; Gypsy lacZ / Tb,Sb	(Sarot et al., 2004)
30	burdock LacZ	genetic sensor	UAS-Dicer2 (X) / hs-hid (Y) ; NGT-Gal4 ; nos-Burdock sensor (Burdock: GFP.LacZ 3'UTR Burdock)	(Handler et al., 2013)
31	CC2	double balanced line	w*; T(2;3)ap[Xa] / CyO, P{ [w*]=Act-GFP}CC2; TM6C, Sb1 Tb1	home-made

Primers and Probes

primer	sequence	experiment
SD22 (RP49_FW)	GACGCTTCAAGGGACAGTATCTG	RT-qPCR
SD23 RP49_Rev)	AAACGCGGTTCTGCATGAG	
MA-037 (Gypsy (2772) _FW)	CCAGGTCGGGCTGTTATAGG	
MA-038 (Gypsy (2663) _Rev)	GAACCGGTGTACTCAAGAGC	
MA-087 (Gypsy (493) _FW and Gypsy spliced_FW)	CAACAATCTGAACCCACCAATCT	
MA-088 Gypsy (658) _Rev	ATTACGTGGCCAGATACACCA	
MA-090 (Gypsy spliced_Rev)	TATGAACATCATGAGGGTGAACA	
CG7009_qPCR2_FW	GAGTTTTGTCTGCCCGATGG	
CG7009_qPCR2_Rev	ACTTGGCTCGTTTTCTGCAG	
CG5220_qPCR2_FW	GATTAACCCTGCTCGCGATG	
CG5220_qPCR2_Rev	TCCAGGGGATAAGATGCGTC	
DCV_FW	TCATCGGTATGCACATTGCT	
DCV_Rev	CGCATAACCATGCTCTTCTG	
LacZ_2_FW	ACTATCCCGACCGCCTTACT	
LacZ_2_Rev	GTGGGCCATAATTCAATTCG	
CG7009_Hind_ATG	GCAAAGCTTATGGGCAGGACTTCGAAGGATA	
CG7009_EcoRI_ATG	AAGAATTCATGGGCAGGACTTCGAAGGATA	
CG7009_NotI_Stop	GCAGCGGCCGCTTACGTTACACAGGCACCTAACT	Cloning of recombinant proteins
CG5220_Hind_ATG	GCAAAGCTTATGGGGAAAACATCAAAGGA	
CG33172_hind_ATG	CAACTGGCAAAGCTTATGGTTTTGATTTCTGACGC	
CG33172_EcoRI_ATG	AAGAATTCATGGTTTTGATTTCTGACGC	
CG33172_NotI_Stop	ACTGGCAGCGGCCGCTTAAAGTATATTACTTATGCTCATAGTCTGC	
pAWF_CG7009_FW	CACCATGGGCAGGACTTCGAAGGATA	
pAWF_CG7009_Rev	CGTTACACAGGCACCTAACT	
F_U tRNA-Phe or FU	GCCGAAATAGCTCAGTTGGG	RTLP
Rev tRNA-Phe or R	TGCCGAAACCCGGGATTGAA	
T7_Ftz_FW	GAATTGTAATACGACTCACTATAGGGCTGGCAAAGTCGCCATTCT	dsRNA synthesis
T7_Ftz_Rev	GAATTGTAATACGACTC-ACTATAGGGCCAACATGTATCACCCCCA	
T7_GFP_FW	GAATTGTAATACGACTCACTATAGGGCTTACTTGTACAGCTCGTC	
T7_GFP_Rev	GAATTGTAATACGACTCACTATAGGGCATGGTGAGCAAGGGCGAG	
CG7009-dTOPO FW	CACCATGGGCAGGACTTCGAAGGAT	genotyping
CG7009-dTOPO Rev	TTACGTTACACAGGCACCTAACTTC	
CG7009-middle FW	TCCACTGGAATGCACGACTT	
CG7009-middle Rev	AAGTCGTGCATTCCAGTGGA	
pB-3SEQ	CGATAAAACACATGCGTCAATT	
pB-5SEQ	CGCGCTATTTAGAAAGAGAGA G	
VIE0197: 5220 mutant screening FW	GATATATCGATAGGCTGGCC	
VIE0198: 5220 mutant screening Rev	CAGGTATCGTAGAGTTTCCG	
VIE0259: 5220 mutant screening Rev [A to P]	ACGTGGAGCAACTTGAATGG	
VIE0260: 5220 mutant screening Rev [K to A]	TCGTCCACGTGGAGCAACGC	
tRNA Phe GAA 5' probe (MA_075)	GCTCTCCCAACTGAGCTATTTCGGC	Northern blot
tRNA Leu TAA 3' probe (MA_103)	TGCCAGGTGTGGGGTTCGAACCCA	
as-U6 (primer 5379 CA lab primers)	CGATTTTGCGTGTCATCCTTGC	

Genomic DNA extraction

A routine procedure in the laboratory for extraction consists in freezing the flies at least briefly at -20°C followed by homogenization in *Squishing buffer* (10 mM Tris HCl pH 8, 2.1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K). After incubation of the lysates for 30 min at 37°C Proteinase K is heat inactivated for 10 min at 95°C. The samples are then briefly centrifugated and the gDNA contained in the supernatant is recovered.

***Drosophila* total RNA extraction**

Whole flies were grinded manually with a pestle and total RNA was extracted with TRI-Reagent (Sigma) and Chloroform (ThermoFischer), following the manufacturer's instructions. After recovery of the aqueous phase 2/3 volumes of Isopropanol are added for RNA precipitation, the reactions are incubated 10 min on RT or 1 hour on ice and centrifugated for at least 30 min at maximum speed. The RNA pellets are then washed with 70% ethanol, air dried and dissolved in RNase free water. Quantification is performed on *NanoDrop1000* spectrophotometer.

RT-qPCR

Whole flies, dissected ovaries or heads were crushed with a pestle in TRI-Reagent (Sigma Aldrich). Proteins and cell debris were precipitated by addition of 2/3 Volumes of Chloroform (Thermo Fisher) and the RNA was precipitated using Isopropanol, by following the described RNA extraction protocol. After DNase digestion using the TURBO DNA-free™ Kit (Ambion), the total RNAs were used in a reverse transcription reaction with Random Primers (Promega) and Super Script® II Reverse Transcriptase (Invitrogen). The cDNA that was obtained after this reaction was then used to perform a RT-qPCR on a CFX96 Touch™ Real-Time PCR Detection System (Life Sciences) using targets-specific primers, as well as Rp49 primers used for normalization. The analysis was performed using double delta Ct and basic Student test for statistical analysis.

tRNA^{Phe} *in vitro* transcription

DNA template containing tRNA$^{Phe(GAA)}$ was obtained from PCR on genomic DNA, using a chimeric forward primer containing the T7 promoter sequence at their 5'-end. 500 ng of the templates were using

the MEGAscript® T7 Kit (Ambion) following the manufacturer's recommendations. After digestion of the DNA template with Turbo DNase kit (Ambion) for 30 min at 37°C, *in vitro* transcribed $tRNA^{Phe(GAA)}$ was extracted with TRI-Reagent (Sigma Aldrich), followed by isopropanol precipitation. The RNA pellet was then washed with 70% ethanol, air dried and dissolved in RNase free water. Quantification was performed on *NanoDrop1000* spectrophotometer and the purity and size of the *in vitro* transcribed RNA was controlled by agarose gel electrophoresis in 1X Tris-Acetate EDTA (TAE) buffer and ethidium bromide (EtBr) staining. Migration products were visualized under UV light box. RNA was stored at -20°C. *In vitro* transcribed tRNAs were used in methylation assays and in RTL-P experiments.

Genotyping of *CG7009*^{e02001} / *CG7009*^{e02001}mutants

Genotyping of the strain mutant for *CG7009* was performed by PCR (Polymerase Chain Reaction) and Sanger sequencing. Genomic DNA (gDNA) from 5 homo- or heterozygous individuals for the *CG7009* mutant allele (denoted by *CG7009*^{e02001}) was extracted according to the protocol *5 Fly Drosophila Genomic Prep for iPCR in 96-well Format* by R. Buchholz, W. Miyazaki and N. Dompe. The prepared gDNA was then amplified by PCR using a pair of convergent primers hybridizing at the 5'- and the 3'-ends of the gene *CG7009* entitled CG7009-dTOPO FW and CG7009-dTOPO Rev respectively. The reaction was carried out on a 1:10 dilution of the prepared gDNA with the Q5 High-Fidelity DNA Polymerase (BioLabs), according to the manufacturer's instruction and using the following program: initial denaturation at 98°C for 30s; 30 cycles of amplification (98°C for 10s, annealing for 30s at 63°C and elongation at 72°C for 4min); final elongation at 72°C for 2 min.

PCR products were added with 1/6 volume of Gel loading dye 6X (50mM EDTA, 0.2% SDS, 50% glycerol, 0.05% w/v bromophenol blue) for a final concentration 1X and were separated by electrophoretic migration in 1% Agarose gel dissolved in 1X TAE buffer (Tris-Acetate electrophoresis buffer: 40mM Tris-OH, 20mM Acetic Acid, pH 7.8) added with 1/1000 of ethidium bromide (EtBr) at 0.5 mg/mL. 8 µL of the molecular weight marker 1kb Ladder (Thermo Scientific) were also loaded on the gel.

The insertion site was deduced by classical sequencing (*GATC Biotech*) of the gDNA of the heterozygous *CG7009*^{e02001} mutant, using two divergent oligonucleotides, pB-5SEQ and pB-3SEQ hybridizing at the 5' and the 3'end of the inserted piggyBac transposon respectively.

Transgenic *Drosophila* strain carrying WT genomic *CG7009* locus

Verification and amplification an attB-P[acman] clone *CH322-117K12* (*BACPAC Resources*®) was received as transformed EPI300™ (TransforMax™) electrocompetent *E. coli* living stock in agar tube (stab culture). After spreading in agar plate with Chloramphenicol (12.5 µg/ml) and grow over night of the stab culture at 37°C, 4mL pre-culture started from a single clone was grown overnight in Luria-Bertani (LB) broth, supplemented with Chloramphenicol (12.5 µg/ml) at 37°C. In the morning, 250 mL culture in LB medium supplemented with the appropriate amounts of chloramphenicol was started. Cell were grown at 37°C under vigorous agitation. At an OD_{600} 0.4 bacteria were induced by adding Arabinose at 0.01% to induce multi copy production of the plasmid. The bacteria were incubated for 5 hours at 37°C under agitation. DNA of the BAC was prepared using *Large-Construct Kit* (QIAGEN®), following the manufacturer's instructions. Validation of the presence of *CG7009* gene in the clone *CH322-117K12* was performed by sequencing using CG7009-dTOPO FW, CG7009-dTOPO Rev, CG7009-middle FW and CG7009-middle Rev oligonucleotides. The prepared BAC was sent for targeted attP/ attB transformation in y^1w^{67c23}; P{CaryP}attP40 landing line of *Drosophila* (*BestGene Inc.*) by embryos injection. The "BAC transgenic line" of *CG7009* was established using attP/ attB recombination system and appropriate transformation protocol elaborated by the *BestGene Inc* society.

Construction of sgRNA vector for fly transgenesis

The sgRNA, corresponding to PAM sequences "A-P" and "K-A", was subcloned into the bicistronic *Drosophila* CRISPR/ Cas9 vector pDCC6, expressing sgRNA cassette under the control of the Drosophila U6:96Ab (U6-2) promoter as well as an hsp70Bb promoter driving Cas9 expression. One gRNA, was expressed from vector pDCC6 as chimeric gRNA. In short, two DNA oligos were bought, *in vitro* phosphorylated, annealed and cloned as dsRNA into the pDCC6 vector cut with BbsI (Oligo 1: 5'-CTTCGAGCAACTTGAAGGCACTCC-3' and Oligo 2: 5'-AAACGGAGTGCCTTCAAGTTGCTC-3'). The chimeric gRNA anneals to the genomic DNA of the CG5220 locus. ssODN: donor template K to A mutation in CG5220 was 136 nt (5'-TTCATATATTTATTTACAATGGGGAAAACATCAAAGGACAAAAGAGATATCTATTACCGACAAGCCAAAGACGAAGGCTGGAGGGCGAGGAGTGCCTTCGCGTTGCTCCACGTGGACGAAGCCTACGGAATTCTAA-3'); ssODN: donor template A to P mutation in CG5220: 136 nt (5'-TTCATATATTTATTTACAATGGGGAAAACATCAAAGGACAAAAGAGATATCTATTACCGACAAGCCAA

AGACGAAGGCTGGAGGGCGAGGAGTCCATTCAAGTTGCTCCACGTGGACGAAGCCTACGGAATTCT
AA-3'). Mutagenesis was performed as described in (Gokcezade et al., 2014).

Allele-specific PCR

The primer sequences for screening for indels the flies from the F1 cross-out are VIE0197 (genomic primer forward in *CG5220*, used with VIE0198 yields 438 bp amplicon that includes the targeted region). VIE0198: (genomic primer reverse in CG5220, used with VIE0197 yields 438 bp amplicon that includes the targeted region). These primers should be used to sequence the region.

In order to look for the specific mutations the following primers were used: VIE0197 as forward and VIE0259 as reverse primer to detect *CG5220* edited with A to P (yields a 279 bp PCR product), fly stock denomination: #247.21.1 or A26P; and VIE0197 as forward and VIE0260 as reverse primers to detect CG5220 edited with K to A (yields a 284 bp PCR product), fly stock denomination: #248.5.2 or K28A. VIE0259 and VIE0260 reverse primers are only annealing to the modified loci in CG5220 and should yield no PCR product if the locus is wild type.

Identification of *CG5220* mutant alleles in *CG7009, CG5220* double mutant recombinants

The PCR was carried out gDNA from $CG5220^{247.21.1}$ mutants with specific primers VIE0197 and VIE0259 and on $CG5220^{248.5.2}$ mutants with the corresponding specific primers VIE0197 and VIE0260. VIE0197 forward primer (Eurogentec) and reverse primers VIE0260 and VIE0259 were used to amplify *CG5220* containing nucleotides substitutions in CG5220 performed by CRISPR/Cas9 (M. Schaefer). PCR was carried out with *DreamTaq* Polymerase Master Mix (*ThermoFischer*), using the following program: annealing at 51°C (30 sec), elongation 72°C (30 sec), final extension of 2 min at 72°C. The PCR products were migrated on 1.4% agarose electrophoresis, stained with ethidium bromide. Upon amplification detection, gDNA from both genotypes was subjected to PCR with the high fidelity Q5 DNA Polymerase (BioLabs) with *VIE0197* and *VIE0198* primers and the amplification products were columns purified (*NucleoSpin® Gel and PCR Clean-up, Macherey-Nagel*) and subjected to Sanger sequencing (GATC Biotech) The obtained sequences correspond to fragments covering 438 bp of the *CG5220* gene. Sequences analysis was carried with *4Peaks* and *ApE*.

Injection of DCV virus in adult flies / injection test with different concentration

Flies with the following genotypes were subjected to intra-thoracic injection with the *Drosophila C* virus (DCV): *CG7009*^{e02001} / + or *CG7009*^{e02001} / TM6, Tb, Sb (controls); *CG7009*^{e02001}/ Def3340, *CG7009*^{e02001}/ Def9487, *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ (*CG7009* mutants); BAC / Cyo; *CG7009*^{e02001}/ Def9487, BAC / +; *CG7009*^{e02001}/ Def3340, BAC / BAC; *CG7009*^{e02001}/ Def3340; BAC / BAC; *CG7009*$^{e02001\text{-}G10}$/ *CG7009*$^{e02001\text{-}G10}$ (*CG7009* rescue lines); *CG5220*^{K28A}/ *CG5220*^{K28A} (*CG5220* catalytic mutant); *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A}/ *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} Rec1 (*CG7009, CG5220* double mutant), *CG5220*^{K28A}/ *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} Rec7 (*CG7009, CG5220* double mutant), Dnmt2^{99} (positive control). In the experiment setup 2 to 4 days old flies were divided in tubes of 10 (5 males + 5 females) and 20 flies from each genotype were injected with DCV while 20 other flies were injected, as controls, with adult Injection Buffer, containing 10 mM Tris pH 6.3 and 1mM $MgCl_2$. Each fly was injected with 4.6 nL of DCV concentration of $2x10^6$ PFU/ml (9.2 PFU/injection). Intra-thoracic injections were made using the Drummond Automatic Nanoliter Injector "NANOJECT II". After injection, flies were kept at 25°C. Three to four days after the injection, 3 injected flies from each genotypes and conditions were frozen at -20°C. Those flies were then crushed with a pestle in TRI-Reagent (Sigma Aldrich) and total RNA was extracted as described above. After DNase digestion using the TURBO DNA-free™ Kit (Ambion), the total RNAs were used in a reverse transcription reaction with Random Primers (Promega) and Super Script® II Reverse Transcriptase (Invitrogen). The cDNA that was obtained after this reaction was then used to perform a Quantitative-PCR using primers DCV_FW and DCV_Rev (Primers and Probes Table above). Rp49 was used for normalization.

Survival experiments

Lifespan tests were carries out in each experiment on 10 females, aged between 1 and 2 days. During the different experiments the following genotypes have been used in different combinations: *CG7009*^{e02001} / + or *CG7009*^{e02001} / TM6, Tb, Sb (controls); *CG7009*^{e02001}/ Def3340, *CG7009*^{e02001}/ Def9487, *CG7009*$^{e02001\text{-}G10}$ / *CG7009*$^{e02001\text{-}G10}$ (*CG7009* mutants); BAC / Cyo or + or BAC; *CG7009*^{e02001}/ Def9487, BAC / Cyo or + or BAC; *CG7009*^{e02001}/ Def3340, BAC / BAC; *CG7009*^{e02001}/ Def3340, BAC / BAC; *CG7009*$^{e02001\text{-}G10}$/ *CG7009*$^{e02001\text{-}G10}$ (*CG7009* rescue lines); *CG5220*^{K28A}/ *CG5220*^{K28A} (*CG5220* catalytic mutant); *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A}/ *CG7009*$^{e02001\text{-}G10}$, *CG5220*^{K28A} Rec1 (*CG7009, CG5220* double mutant), The flies were kept at 25°C, and flipped every 2 to 3 days in new tubes.

RNA Interference in S2 cells

Double stranded RNAs were synthesized by *in vitro* transcription (MEGAscript® T7 Kit, Ambion) of PCR products amplified from w^{1118} genomic DNA using primers flanked by T7 promoters. Sequences of amplicon templates for dsRNA production are available at the *Drosophila* RNAi Screening Center (http://www.flyrnai.org/cgi-bin/RNAi_gene_lookup_public.pl) and identified as follows: Ago-2 DRSC10847, CG7009#1 DRSC16171, CG7009#2 DRSC39197, CG7009#3 DRSC39198. PCR products for T7 transcription of *fushi tarazu* (Ftz) or GFP dsRNAs were amplified using primers: T7_Ftz_FW and T7_Ftz_Rev (*fushi tarazu*); T7_GFP_FW and T7_GFP_Rev (Primers and Probes Table above).

Standard S2 cells transfection scheme

100 µL of S2 cells at a concentration of 10^6 cells/ml are resuspended in Schneider's *Drosophila* medium (GIBCO-Invitrogen) and plated in 96-well plates. Cells are transfected with dsRNA or co-transfected with dsRNA and sensor (see sub-chapters) using Effecten transfection reagent (Qiagen) following the manufacturer's instructions. After transfection, cells were left to rest 30 min and supplemented with 50 µL Schneider's *Drosophila* medium (GIBCO-Invitrogen), containing 10% heat-inactivated fetal calf serum, 100 U/ml penicillin and 100 mg/ml streptomycin. Cells were grown at 23°C without CO_2. After 24 to 48 hours (see sub-chapters), the copper-inducible metallothionein promoter was induced by 600 µM CuSO4 and GFP fluorescence appearance was followed using an inverted epifluorescence basic microscope in a dark room.

Ago-2-mediated miRNA pathway involvement in cells

Cells were co-transfected with 0.1 µg of AutomiG-vector and 0.32 µg of dsRNA targeting either Ago-2, either CG7009 (#1, #2, or #3), or Ftz. 48 hours later the automiG promoter was induced by adding 600uM final concentration of copper (more details in (Carre et al., 2013).

siRNA pathway involvement

Cells were initially transfected with 0.32 µg of dsRNA targeting either Ago-2, either CG7009 (#1, #2, or #3), or Ftz. 24 hours later a second co-transfection was performed with 0.14 µg of dsRNA against GFP

and 0.1 µg of AutomiG-Δ1-Δ2vector (Carre et al., 2013). 30 min later 50 µL of serum- and antibiotics-supplemented medium was added as described above, and the experiment followed the described standard transfection scheme.

GFP Western Blot on automiG transfected S2 cells

Expression of the GFP was analyzed by western-blot using anti-GFP (Roche®) and anti-γ-Tubulin (Sigma®) as loading control. 72 hours following transfection, the culture medium was carefully removed and 80µL of Sample Buffer Laemmli 2X (Sigma®) was added in each well. *Drosophila* S2 cells are weak and this treatment is sufficient for cracking the cells. The samples were boiled in the Laemmli buffer and 18 µL were loaded on 4-20% Mini-PROTEAN®TGX™ Gel, 12 well (Biorad®). After transfer (liquid) onto nitrocellulose membrane, membranes were blocked in 5% milk, 1X PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4), 0.1% Tween, and incubated overnight with anti-GFP (1:2000; Roche®) or anti-γ-Tubulin (1:2000; Sigma®) antibodies diluted in the blocking solution under agitation. Appropriate secondary antibody (1:5000) coupled to alkaline phosphatase (Biorad®) was added and incubated for 2 hours at room temperature under agitation. Detection was performed using the BCIP/NBT reagents (Thermo Scientific®).

Cloning of CG7009 in pAWF Gateway plasmid

A first PCR was performed to amplify the CDS of CG7009 from the home made pGEX4T1_CG7009 vector using the following designed for Gateway cloning primers pAWF_CG7009_FW and pAWF_CG7009_Rev. The obtained PCR product was then cloned into a pENTR/D-TOPO vector following the manufacturer's instructions of pENTR™/D-TOPO® Cloning Kit (Invitrogen). The obtained pENTR_CG7009 vector was then transformed in *E. coli* ccdB resistant strain, allowing the amplification of the pENTR_CG7009 plasmid. The construction was checked by PCR on colonies, then the plasmids were in purified using NucleoSpin® Plasmid kit (Macherey-Nagel), and verified by Sanger Sequencing (GATC Biotech), using the CG7009_middle Rev primer. Once verified the plasmid was recombined using LR recombination reaction (https://www.thermofisher.com/fr/fr/home/life-science/cloning/gateway-cloning/protocols.html#lr) into the destination vectors pAWF and pMT-DEST48, allowing the production of FLAG_CG7009 or CG7009_V5_His, respectively in *Drosophila* S2 cell lines. The obtained constructions

were verified by PCR with primers CG7009_middle Rev for pAWF and CG7009_middle FW for pMT-DEST48 and by Sanger sequencing (GATC Biotech).

Expression and co-IP of CG5220_FLAG_HA protein in S2 cells

CG5220 CDS, fused to a C-terminal FLAG-HA tag, cloned into a pMK33-CFH-BD vector and allowing expression of recombinant protein CG5220_FLAG_HA in *Drosophila* S2 cells, was ordered from *Berkeley Drosophila Genome Project*'s collection: expression clone FMO01654 (https://sina.lbl.gov/cgi-bin/labtrack/community_query/cloneReport.pl?clone_id=FMO01654). The plasmid was verified by Sanger sequencing, amplified in bacteria and transfected in S2 cells. CG5220 interacting complexes were co-IP from cells transiently transfected with the vector and used in *in vitro* methylation assays as described.

Recombinant Protein Plasmid constructs

FLAG_CG7009. Amplification of full-length cDNAs of *CG7009* (#SD16956 in FlyBase) was done using standard PCR techniques with VENT polymerase (New England BioLabs-NEB). Amplification products were cloned between the HindIII and NotI restriction sites of the pET-28a (Novagen) vector (modified to contain FLAG peptide) using the primers CG7009_Hind_ATG and CG7009_NotI_Stop (see primers and probes).

GST_CG7009. Glutathione *S*-transferase (GST) fusion constructs were generated by PCR amplification of full-length cDNAs of *CG7009* using standard PCR with VENT polymerase (New England BioLabs). Products were cloned between the EcoRI and NotI restriction sites of the pGEX-4T1 (GE Healthcare) vector using primers CG7009_EcoRI_ATG and CG7009_NotI_Stop (see primers and probes).

The GST-DmHen1 and GST-ΔDmHen1 constructs, a kind gift of Mikiko Siomi, were previously described (Saito et al., 2007).

FLAG_CG33172. Amplification of full-length cDNAs of *CG33172* (clone MIP10235 in BDGP) was done using standard PCR techniques using Q5 high fidelity DNA Polymerase (*New England BioLabs*). Amplification products were cloned between the HindIII and NotI restriction sites of the pET-28a (Novagen) vector (modified to contain FLAG peptide) using the primers CG33172_Hind_ATG and CG33172_NotI_Stop (see primers and probes).

Bacterial heat shock transformation protocol

Competent bacteria BL21 Star (DE3, Invitrogen), C41, TOP10 (Invitrogen), HB101 (Promega) were transformed by heat-shock with 100-200 ng of plasmid DNA according to each manufacturer's instructions. After expression on the corresponding antibiotic resistance genes by incubation for 1-1,5 hours at 37°C under agitation of 250 rpm, 1/10 and 9/10 of the transformed bacteria were plated on LB agar plates, supplemented with the corresponding antibiotics.

Methylation assays with CG7009 recombinant protein

In vitro methylation assays were made in 50µl of reaction mixture, using 12 µg of RNA substrates (~0.5 nmol) and 2 µg of recombinant proteins produced in the laboratory. The different RNAs used as substrate were *in vitro* synthesized $tRNA^{Phe}$, and chemically synthesized miRNA and piRNA (Eurogentec). The recombinant proteins FLAG_CG7009, GST_CG7009, FLAG_CG33172, GST_DmHen1, GST_ΔDmHen1, and Dnmt2 (kind gift of M. Schaefer) produced in *E. coli* were used in the different methylation assays. Each protein was incubated with one of the substrates at a time, at 25°C in a buffer containing: 50 mM Tris-HCl (pH 8), 100 mM KCl, 5 mM MgCl2, 2 mM DTT, 5% glycerol, 0.2 µl RNasin (Promega) and 0.25 µCi S-adenosyl-L-[methyl-3H] Methionine (SAM-3H, PerkinElmer). After 3 hours of incubation at 25°C, the RNA was precipitated and extracted using the standard Isopropanol RNA extraction protocol. All the precipitated RNAs was resuspended in 10-13 µL of Gel Loading Buffer II (Invitrogen) and loaded on 15% PAA-Urea TBE precast gels (BioRad) and migrated according to the manufacturer's instructions. The gels were then dried using a vacuum pump and exposed to a photographic film (GE Healthcare).

Immunoprecipitation of CG7009 and CG5220 from S2 cells for methylation assays

Cells expressing CG5220_FLAG_HA, FLAG_CG7009 or CG7009_V5_His were frozen, then homogenized in 50 µl of Lysis Buffer containing: 20 mM Tris pH 7.4-8, 100 mM NaCl, 0.5% NP40, DTT 1mM and PIC 1X. After a 3000 rpm centrifugation at 4°C for 5 mins, the lysate was mixed to Flag-M2 beads (Sigma Aldrich, ANTI-FLAG® M2-Agarose from mouse) and incubated for 2 hours at 4°C. The beads were then washed 3 times with BC500 (a buffer containing 20 mM Tris HCl pH 7.4-8, 500 mM NaCl, 10% Glycerol, 0.2 mM EDTA, 15 mM **β**-mercaptoethanol, 0.1% Triton, 1mM DTT and 1X PIC) and 1 time with the buffer BC100 (containing 20 mM Tris HCl pH 7.4-8, 100 mM NaCl, 10% Glycerol, 0.2 mM EDTA,

15 mM β-mercaptoethanol, 1mM DTT and 1X PIC). FLAG_CG5220 was then eluted by addition of FLAG peptide (Sigma Aldrich) at 400µg/ml final concentration. Each of 4 elutions was incubated 30 minutes on ice before centrifugation at 3000 rpm for 5 mins at 4°C. 4 µl of each immonuprecipitate were used in methylation assays carried out in identical conditions as for the recombinant proteins produced in bacteria.

Production and affinity purification of recombinant fusion proteins

Fusion proteins GST- and FLAG- G7009; GST- DmHen1 and GST- DmHen1 (mutated) and FLAG-CG33172 were expressed in *Escherichia coli* BL21 Star (DE3) (Invitrogen®) or C41 and purified over glutathione-coupled resin (Pharmacia®) / Glutathione Sepharose 4B resin (GE Healthcare) as previously described (Naar et al., 2002) for GST-containing constructs, or on anti-FLAG M2 beads (Sigma) for pET-28a-Flag fusion proteins. Bound peptides were eluted on Bio-spin disposable chromatography columns (Bio-Rad) either with 20 mM GSH (pH 8) or with 400 µg/ml Flag peptide (Sigma®) in BC100 (100mM KCl) buffer for 20 min on ice.

***In vitro* interaction of GST-CG7009 and FLAG-CG33172**

GST (pGEX4T1-CG7009) and FLAG (pET28a-FLAG-CG33172) fusion constructs were generated separately, as described above in *Recombinant Protein Plasmid constructs* section.

In a 2nd approach the proteins were co-expressed in bacteria and the transformants were selected on LB agar supplemented with both corresponding antibiotics (kanamycin for the pET28a-FLAG vector and ampicillin for the pGEX4T1 vector). The recombinant proteins were then purified as described (Naar et al., 2002) and quantified by spectrophotometry at 280 nm using NanoDrop1000. When the two proteins were co-expressed WB was performed immediately after the purification. When the two recombinant proteins were expressed separately, an additional step of interaction was performed after the purification, in which 5 µg of each protein were mixed in a final volume of 200 µl and incubated 90 min on rotating wheel, 25°C. Then, the proteins were incubated with either 20 µl beads Glutathione Sepharose 4B washed in BC100 buffer for GST pull-down or 50 µl anti-FLAG M2 beads for FLAG-IP for 2 hours at 4°C on rotating wheel, following the purification protocol of (Naar et al., 2002).

For WB, samples were boiled in 2X Laemmli buffer (Sigma), migrated on 4-20% Mini-PROTEAN TGX gel (Bio-Rad), and transferred on a PVDF (polyvinylidene fluoride) membrane (Amersham Hybond, GE

Healthcare) for 2 hours (300 mA) at 4°C. Then the membrane was blocked with 5% non-fat dry milk dissolved in 1X TBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH2PO4, 0.1% Tween) for 1 hour at room temperature. After washing with 1X TBST the membranes were incubated with 1:10,000 anti-GST HRP (horseradish peroxidase) conjugate (Amersham GE Healthcare) or 1:10,000 anti-FLAG (Sigma) antibodies for 1 hour at room temperature under agitation. GST was visualized by adding enhanced chemiluminescent (ECL) substrates (ThermoFisher) of the anti-GST HRP. For FLAG detection in the anti-FLAG WB, a secondary 1:10,000 anti-mouse antibody coupled to alkaline phosphatase (AP) (Promega) was incubated for 1 hour at room temperature under agitation. Detection was performed using BCIP (5-bromo-4-chloro-3-indolyl-phosphate) and NBT (nitro-blue-tetrazolium), (ThermoFischer) reagents diluted in AP buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$). Fusion proteins purified by anti-FLAG affinity purification were assayed for purification quality control by anti-FLAG WB and with anti-GST WB to assay the co-immunoprecipitation of GST-7009 with FLAG-CG33172.

RTLP

RT was performed in 20 µl reaction volume on 200 ng total RNA, 10 µM RT primers *Rev tRNA-Phe* (see *Primers and Probes* table), dNTP mix at 1 mM high or 1 µM low concentration and water adjusted to 12 µl. Samples were denatured at 65°C for 5 min and kept on ice. After adding RT buffer, initial annealing with primer *Rev tRNA-Phe* was performed at 42°C for 2 min. Then Superscript II RT (10 U/µl per reaction - ThermoFischer) was then added and the reactions were incubated at 42°C for 50 min, followed by RTase deactivation (70°C, 15 min). PCR was carried out in 20 µl final volume (0.2 mM dNTP and 0.2 µM of each primer: *Primers and Probes* table) with DreamTaq DNA polymerase (1.25 U - ThermoFischer) on 2 µl of the cDNA template, using primers Fu tRNA-Phe and Rev tRNA-Phe (see *Primers and Probes* table). PCR program: initial denaturation 94°C for 4 min; 28 cycles of denaturation (94°C for 30 s), annealing (51°C for 30 s), and extension (72°C for 30 s); final extension 72°C for 5 min. PCR products were migrated on 2.2% agarose gel, stained with EtBr, and visualized under UV light.

MALDI-TOF and LC-MS RNA Extraction

1000 females and 1000 males were collected in lot of 50 and homogenized on a tissue homogenizer (Precellys 24®, Bertin Technology) in 1 mL TRI-reagent (Sigma Aldrich). A total 20 mL of fly lysates, corresponding to 1000 flies, were extracted together in 50 mL falcon tubes. Extraction followed the

standard lab extraction procedure by adjustment of the reagents volumes. RNA was resuspended in RNase free water.

Purification of $^{Phe}tRNA_{\#AA}$

Total RNA preparations were supplemented with LiCl to a final concentration of 0.8 M and incubated overnight at 4°C to precipitate high molecular mass molecules. The precipitate was eliminated by centrifugation and the supernatant was supplemented with two volumes of 100% ethanol and incubated at −20°C for 2 h to precipitate small RNAs. After centrifugation, pelleted small RNAs were washed twice in 70% ethanol and resuspended in 1 ml of RNAse-free water. tRNAs were further purified using the NucleoBond RNA/DNA 400 kit (Macherey-Nagel) following manufacturer's instructions, except that the elution step was performed with 5 ml of 100 mM Tris-acetate (pH 6.3); 15% ethanol and 600 mM KCl. Eluted tRNA were ethanol precipitated and resuspended in 1 ml of RNAse-free water.

Purification of $^{Phe}tRNA_{\#AA}$ was made with a 5' biotinylated complementary oligonucleotide (5'-biotine-TGGTGCCGAAACCCGGGATTGAACCGGGG-3') coupled to Streptavidin Magnesphere Paramagnetic particles (Promega). Annealing of specific tRNA was made in 1 x TMA buffer (Tris-HCl pH 7.5 10 mM, ethylenediaminetetraacetic acid (EDTA) 0.1 mM, tetramethylammonium chloride 0.9 M) by heating the mixture at 95°C for 3 min and cooled to 60°C for 30 min. Paramagnetic particles were then washed three times with 1 x TMA buffer and specific tRNA was recovered by heating the final suspension at 95°C for 3 min. $^{Phe}tRNA_{\#AA}$ was desalted and concentrated four times to 50 µL in Vivacon 500 devices (Sartorius; 2000 MWCO) using 100 mM ammonium acetate (pH 5.3) as a final buffer. *Note*: if used for RiboMethSeq, LiCl has to be washed away because it interferes with ligation of adaptors during library preparation.

MALDI-TOF analysis of digested $^{Phe}tRNA_{\#AA}$

For mass spectrometry analysis, about 50 µg of $^{Phe}tRNA_{\#AA}$ were digested with 100 units of RNAse T1 (Sigma), which cleaves after G residues and generates 3'-phosphate nucleosides, in a final volume of 10 µL at 37°C for 4 h. One microliter of digest was mixed with 9 µL HPA (40 mg/ml in water: acetonitrile 50:50) and 1 µL of the mixture was spotted on the MALDI plate and air-dried ("dried droplet" method). MALDI-TOF MS analyses were performed directly on the digestion products using an UltrafleXtreme spectrometer (Bruker Daltonique, France). Acquisitions were performed in positive ion mode.

Northern blot analysis

For Northern blot analysis of miRNA or tRNA, 10 µg of S2R+ total RNA extracted using Trizol Reagent (Invitrogen) and resuspended in water were resolved in 15% urea-polyacrylamide gel, transferred to Hybond-NX membrane (GE Healthcare) and EDC-crosslinked (Sigma Aldrich). Oligonucleotides were 5′ 32P end-labeled using T4 polynucleotide kinase (Fermentas) and used as probes. Hybridization was performed over-night at 37°C in PerfectHyb Plus (Sigma) hybridization buffer. Probes are detailed in the primer table. The bantam and the miG-1 probes were previously described (Brennecke et al., 2003; Carre et al., 2013). More details on NB procedure are available in (Carre et al., 2013).

RiboMethSeq preparation and analysis

Adapted from: Illumina-based RiboMethSeq approach for mapping of 2′-O-Me residues in RNA. Virginie Marchand; Florence Blanloeil-Oillo; Mark Helm and Yuri Motorin. Nucleic Acids Research, Volume 44, Issue 16, 19 September 2016, Pages e135, https://doi.org/10.1093/nar/gkw547.

RNA fragmentation conditions: RNA (1–250 ng) are subject to alkaline hydrolysis or metal ion-based RNA cleavage (magnesium or zinc ions). RNA hydrolysis is performed in 50 mM bicarbonate buffer pH 9.2 for 4–10 min at 95°C or in 100 mM Tris–HCl buffer, pH 8.0, containing either 0.1 mM ZnCl2 or 2 mM MgCl2 for 2 or 3 min, respectively. The reaction is stopped by ethanol precipitation using 3M Na-OAc, pH 5.2 and glycoblue as a carrier in liquid nitrogen. In addition, the zinc-based RNA cleavage is stopped by addition of 2 µl of 0.5 M EDTA pH 8.0 before ethanol precipitation. After centrifugation, the pellet is washed with 80% ethanol and resuspended in nuclease-free water. The sizes of generated RNA fragments is assessed by capillary electrophoresis using a PicoRNA chip on Bioanalyzer 2100 (Agilent, USA) and are ranging from 30 to 200 nt. End repair: RNA fragments without any gel-purification step are directly 3′-end dephosphorylated using 5 U of Antarctic Phosphatase (NEB, UK) for 30 min at 37°C. After inactivation of the phosphatase for 5 min at 70°C, RNA fragments are phosphorylated at the 5′-end using T4 PNK and 1 mM ATP for 1h at 37°C. End-repaired RNA fragments are then purified using RNeasy MinElute Cleanup kit (QIAGEN, Germany) according to the manufacturer's recommendations except that 675 µl of 96% ethanol were used for RNA binding. Elution is performed in 10 µl of nuclease-free water.

Library preparation: RNA fragments are converted to library using NEBNext® Small RNA Library kit (NEB ref E7330S, UK or equivalent from Illumina, USA) following the manufacturer's instructions. DNA library quality is assessed using a High Sensitivity DNA chip on a Bioanalyzer 2100. Library quantification

is done using a fluorometer (Qubit 2.0 fluorometer, Invitrogen, USA). Sequencing: Libraries are multiplexed and subjected for high-throughput sequencing using an Illumina HiSeq 1000 instrument with a 50 bp single-end read mode (or using an Illumina MiSeq for paired-end read runs). Since clustering of short fragments is very efficient, libraries were loaded at 8 pM concentration per lane.

Bioinformatics pipeline: Initial trimming of adapter sequence was done using Trimmomatic-0.32 or equivalent with the following parameters: LEADING:30 TRAILING:30 SLIDINGWINDOW:4:15 MINLEN:17 AVGQUAL:30. Alignment to the reference mRNA sequence is done using Bowtie2 in End-to-End mode and k = 1. 5'-end counting is done directly on *.sam file using a dedicated Unix script. Simultaneous 5'- and 3'-ends counting is done by bedtools v2.25.0 after conversion to *.bed file. Final analysis is performed by calculation of score MAX for detection of 2'-O-Me residues, and MethScore for their quantification.

Calculation of score MAX and MethScore (not presented in this PhD manuscript): To calculate score MAX, the relative change of end coverage position by position is calculated in 5'-'3' and reverse direction. The relative change is normalized to average values for -6 and +6 nucleotides. The normalized relative change for 5'-'3' and reverse direction are averaged and the maximal value between the average and normalized relative change is retained (score MAX). MethScore is calculated essentially as described in (Birkedal U. Christensen-Dalsgaard M. Krogh N. Sabarinathan R. Gorodkin J. Nielsen H. Profiling of ribose methylations in RNA by high-throughput sequencing Angew. Chem. Int. Ed. Engl. 2015 54 451 455) for ScoreC using the same relative impact of neighboring nucleotides.

REFERENCES

Abe, M., Naqvi, A., Hendriks, G.J., Feltzin, V., Zhu, Y., Grigoriev, A., and Bonini, N.M. (2014). Impact of age-associated increase in 2'-O-methylation of miRNAs on aging and neurodegeneration in Drosophila. Genes Dev *28*, 44-57.

Adams, C.M., and Eischen, C.M. (2014). Inactivation of p53 is insufficient to allow B cells and B-cell lymphomas to survive without Dicer. Cancer Res *74*, 3923-3934.

Agris, P.F. (2004). Decoding the genome: a modified view. Nucleic Acids Res *32*, 223-238.

Akkouche, A., Mugat, B., Barckmann, B., Varela-Chavez, C., Li, B., Raffel, R., Pelisson, A., and Chambeyron, S. (2017). Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. Mol Cell *66*, 411-419 e414.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.

Anand, A., and Kai, T. (2012). The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. EMBO J *31*, 870-882.

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.

Anderson, P., and Ivanov, P. (2014). tRNA fragments in human health and disease. FEBS Lett *588*, 4297-4304.

Angelova, M.T., Dimitrova, D.G., Dinges, N., Lence, T., Worpenberg, L., Carre, C., and Roignant, J.Y. (2018). The Emerging Field of Epitranscriptomics in Neurodevelopmental and Neuronal Disorders. Front Bioeng Biotechnol *6*, 46.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., *et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol *11*, 1017-1027.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Barckmann, B., El-Barouk, M., Pelisson, A., Mugat, B., Li, B., Franckhauser, C., Fiston Lavier, A.S., Mirouze, M., Fablet, M., and Chambeyron, S. (2018). The somatic piRNA pathway controls germline transposition over generations. Nucleic Acids Res *46*, 9524-9536.

Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J.P., *et al.* (2015). Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. Cell Rep *12*, 1205-1216.

Bassett, A.R., Kong, L., and Liu, J.L. (2015). A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. J Genet Genomics *42*, 301-309.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Besnard-Guerin, C., Jacquier, C., Pidoux, J., Deddouche, S., and Antoniewski, C. (2015). The cricket paralysis virus suppressor inhibits microRNA silencing mediated by the Drosophila Argonaute-2 protein. PLoS One *10*, e0120205.

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. Angew Chem Int Ed Engl *54*, 451-455.

Bjork, G.R., Durand, J.M., Hagervall, T.G., Leipuviene, R., Lundgren, H.K., Nilsson, K., Chen, P., Qian, Q., and Urbonavicius, J. (1999). Transfer RNA modification: influence on translational frameshifting and metabolism. FEBS Lett *452*, 47-51.

Blanco, S., Dietmann, S., Flores, J.V., Hussain, S., Kutter, C., Humphreys, P., Lukk, M., Lombard, P., Treps, L., Popis, M., *et al.* (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. EMBO J *33*, 2020-2039.

Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crecy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A., *et al.* (2018). MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res *46*, D303-D307.

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.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25-36.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387-1392.

Buchan, J.R., and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. Mol Cell *36*, 932-941.

Bugl, H., Fauman, E.B., Staker, B.L., Zheng, F., Kushner, S.R., Saper, M.A., Bardwell, J.C., and Jakob, U. (2000). RNA methylation under heat shock control. Mol Cell *6*, 349-360.

Caldas, T., Binet, E., Bouloc, P., Costa, A., Desgres, J., and Richarme, G. (2000). The FtsJ/RrmJ heat shock protein of Escherichia coli is a 23 S ribosomal RNA methyltransferase. J Biol Chem *275*, 16414-16419.

Carre, C., Jacquier, C., Bouge, A.L., de Chaumont, F., Besnard-Guerin, C., Thomassin, H., Pidoux, J., Da Silva, B., Chalatsi, E., Zahra, S., *et al.* (2013). AutomiG, a biosensor to detect alterations in miRNA biogenesis and in small RNA silencing guided by perfect target complementarity. PLoS One *8*, e74296.

Carre, C., and Shiekhattar, R. (2011). Human GTPases associate with RNA polymerase II to mediate its nuclear import. Mol Cell Biol *31*, 3953-3962.

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., and Zamore, P.D. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. Mol Cell *42*, 172-184.

Chatterjee, K., Nostramo, R.T., Wan, Y., and Hopper, A.K. (2018). tRNA dynamics between the nucleus, cytoplasm and mitochondrial surface: Location, location, location. Biochim Biophys Acta *1861*, 373-386.

Chen, Y.A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaya, E., Li, S., Vempati, S., Laver, J.D., Patel, D.J., *et al.* (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. Mol Cell *63*, 97-109.

Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W., Green, P.J., Barton, G.J., and Hutvagner, G. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA *15*, 2147-2160.

Couvillion, M.T., Sachidanandam, R., and Collins, K. (2010). A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. Genes Dev *24*, 2742-2747.

Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development *127*, 503-514.

Cristodero, M., and Polacek, N. (2017). The multifaceted regulatory potential of tRNA-derived fragments. Non-coding RNA Investig *1*, 7.

Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci *41*, 324-337.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., *et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell *50*, 749-761.

de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature *490*, 112-115.

Dennis, C., Brasset, E., Sarkar, A., and Vaury, C. (2016). Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in Drosophila. Nat Commun *7*, 13739.

Desset, S., Meignin, C., Dastugue, B., and Vaury, C. (2003). COM, a heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster. Genetics *164*, 501-509.

Diebel, K.W., Smith, A.L., and van Dyk, L.F. (2010). Mature and functional viral miRNAs transcribed from novel RNA polymerase III promoters. RNA *16*, 170-185.

Ding, S.W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell *130*, 413-426.

Donertas, D., Sienski, G., and Brennecke, J. (2013). Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev *27*, 1693-1705.

Dong, H., Lei, J., Ding, L., Wen, Y., Ju, H., and Zhang, X. (2013). MicroRNA: function, detection, and bioanalysis. Chem Rev *113*, 6207-6233.

Dong, Z.W., Shao, P., Diao, L.T., Zhou, H., Yu, C.H., and Qu, L.H. (2012). RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. Nucleic Acids Res *40*, e157.

Durdevic, Z., Hanna, K., Gold, B., Pollex, T., Cherry, S., Lyko, F., and Schaefer, M. (2013a). Efficient RNA virus control in Drosophila requires the RNA methyltransferase Dnmt2. EMBO Rep *14*, 269-275.

Durdevic, Z., Mobin, M.B., Hanna, K., Lyko, F., and Schaefer, M. (2013b). The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. Cell Rep *4*, 931-937.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Emara, M.M., Ivanov, P., Hickman, T., Dawra, N., Tisdale, S., Kedersha, N., Hu, G.F., and Anderson, P. (2010). Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. J Biol Chem *285*, 10959-10968.

Feltzin, V.L., Khaladkar, M., Abe, M., Parisi, M., Hendriks, G.J., Kim, J., and Bonini, N.M. (2015). The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in Drosophila. Aging Cell *14*, 443-452.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Freude, K., Hoffmann, K., Jensen, L.R., Delatycki, M.B., des Portes, V., Moser, B., Hamel, B., van Bokhoven, H., Moraine, C., Fryns, J.P., *et al.* (2004). Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. Am J Hum Genet *75*, 305-309.

Frye, M., and Watt, F.M. (2006). The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. Curr Biol *16*, 971-981.

Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., Xing, R., Sun, Z., and Zheng, X. (2009). Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Lett *583*, 437-442.

Fu, Y., Lee, I., Lee, Y.S., and Bao, X. (2015). Small Non-coding Transfer RNA-Derived RNA Fragments (tRFs): Their Biogenesis, Function and Implication in Human Diseases. Genomics Inform *13*, 94-101.

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J.A., and Imler, J.L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol *7*, 590-597.

Garcia-Silva, M.R., Frugier, M., Tosar, J.P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S., and Cayota, A. (2010). A population of tRNA-derived small RNAs is actively produced in Trypanosoma cruzi and recruited to specific cytoplasmic granules. Mol Biochem Parasitol *171*, 64-73.

Genenncher, B., Durdevic, Z., Hanna, K., Zinkl, D., Mobin, M.B., Senturk, N., Da Silva, B., Legrand, C., Carre, C., Lyko, F., *et al.* (2018). Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats. Cell Rep *22*, 1861-1874.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., *et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Goh, W.S., Falciatori, I., Tam, O.H., Burgess, R., Meikar, O., Kotaja, N., Hammell, M., and Hannon, G.J. (2015). piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. Genes Dev *29*, 1032-1044.

Gokcezade, J., Sienski, G., and Duchek, P. (2014). Efficient CRISPR/Cas9 plasmids for rapid and versatile genome editing in Drosophila. G3 (Bethesda) *4*, 2279-2282.

Goodarzi, H., Liu, X., Nguyen, H.C., Zhang, S., Fish, L., and Tavazoie, S.F. (2015). Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell *161*, 790-802.

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014a). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep *15*, 411-418.

Goriaux, C., Theron, E., Brasset, E., and Vaury, C. (2014b). History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in Drosophila melanogaster. Front Genet *5*, 257.

Gou, L.T., Dai, P., Yang, J.H., Xue, Y., Hu, Y.P., Zhou, Y., Kang, J.Y., Wang, X., Li, H., Hua, M.M., *et al.* (2014). Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. Cell Res *24*, 680-700.

Gou, L.T., Kang, J.Y., Dai, P., Wang, X., Li, F., Zhao, S., Zhang, M., Hua, M.M., Lu, Y., Zhu, Y., *et al.* (2017). Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. Cell *169*, 1090-1104 e1013.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2013a). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics *194*, 1029-1035.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics *196*, 961-971.

Gratz, S.J., Wildonger, J., Harrison, M.M., and O'Connor-Giles, K.M. (2013b). CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. Fly (Austin) *7*, 249-255.

Green, D., Fraser, W.D., and Dalmay, T. (2016). Transfer RNA-derived small RNAs in the cancer transcriptome. Pflugers Arch *468*, 1041-1047.

Grentzinger, T., Armenise, C., Brun, C., Mugat, B., Serrano, V., Pelisson, A., and Chambeyron, S. (2012). piRNA-mediated transgenerational inheritance of an acquired trait. Genome Res *22*, 1877-1888.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., *et al.* (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell *36*, 231-244.

Guida, V., Cernilogar, F.M., Filograna, A., De Gregorio, R., Ishizu, H., Siomi, M.C., Schotta, G., Bellenchi, G.C., and Andrenacci, D. (2016). Production of Small Noncoding RNAs from the flamenco Locus Is Regulated by the gypsy Retrotransposon of Drosophila melanogaster. Genetics *204*, 631-644.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Guo, Y., Bosompem, A., Mohan, S., Erdogan, B., Ye, F., Vickers, K.C., Sheng, Q., Zhao, S., Li, C.I., Su, P.F., *et al.* (2015). Transfer RNA detection by small RNA deep sequencing and disease association with myelodysplastic syndromes. BMC Genomics *16*, 727.

Guy, M.P., and Phizicky, E.M. (2014). Two-subunit enzymes involved in eukaryotic post-transcriptional tRNA modification. RNA Biol *11*, 1608-1618.

Guy, M.P., and Phizicky, E.M. (2015). Conservation of an intricate circuit for crucial modifications of the tRNAPhe anticodon loop in eukaryotes. RNA *21*, 61-74.

Guy, M.P., Podyma, B.M., Preston, M.A., Shaheen, H.H., Krivos, K.L., Limbach, P.A., Hopper, A.K., and Phizicky, E.M. (2012). Yeast Trm7 interacts with distinct proteins for critical modifications of the tRNAPhe anticodon loop. RNA *18*, 1921-1933.

Guy, M.P., Shaw, M., Weiner, C.L., Hobson, L., Stark, Z., Rose, K., Kalscheuer, V.M., Gecz, J., and Phizicky, E.M. (2015). Defects in tRNA Anticodon Loop 2'-O-Methylation Are Implicated in Nonsyndromic X-Linked Intellectual Disability due to Mutations in FTSJ1. Hum Mutat *36*, 1176-1187.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J *21*, 4671-4679.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2015). Two classes of short interfering RNA in RNA silencing. EMBO J *34*, 2590.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Han, B.W., Wang, W., Li, C., Weng, Z., and Zamore, P.D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science *348*, 817-821.

Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., and Brennecke, J. (2013). The genetic makeup of the Drosophila piRNA pathway. Mol Cell *50*, 762-777.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J *30*, 3977-3993.

Harris, A.N., and Macdonald, P.M. (2001). Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. Development *128*, 2823-2832.

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 *16*, 673-695.

Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S.L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. Nature *539*, 588-592.

He, G., Chen, L., Ye, Y., Xiao, Y., Hua, K., Jarjoura, D., Nakano, T., Barsky, S.H., Shen, R., and Gao, J.X. (2010). Piwil2 expressed in various stages of cervical neoplasia is a potential complementary marker for p16. Am J Transl Res *2*, 156-169.

Hebert, S.S., and De Strooper, B. (2009). Alterations of the microRNA network cause neurodegenerative disease. Trends Neurosci *32*, 199-206.

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res *34*, 721-733.

Herschhorn, A., and Hizi, A. (2010). Retroviral reverse transcriptases. Cell Mol Life Sci *67*, 2717-2747.

Hirakata, S., and Siomi, M.C. (2016). piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. Biochim Biophys Acta *1859*, 82-92.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat Rev Mol Cell Biol *6*, 318-327.

Homolka, D., Pandey, R.R., Goriaux, C., Brasset, E., Vaury, C., Sachidanandam, R., Fauvarque, M.O., and Pillai, R.S. (2015). PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. Cell Rep *12*, 418-428.

Honda, S., Kawamura, T., Loher, P., Morichika, K., Rigoutsos, I., and Kirino, Y. (2017). The biogenesis pathway of tRNA-derived piRNAs in Bombyx germ cells. Nucleic Acids Res *45*, 9108-9120.

Honda, S., Loher, P., Shigematsu, M., Palazzo, J.P., Suzuki, R., Imoto, I., Rigoutsos, I., and Kirino, Y. (2015). Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. Proc Natl Acad Sci U S A *112*, E3816-3825.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Huang, H., Li, Y., Szulwach, K.E., Zhang, G., Jin, P., and Chen, D. (2014). AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. J Cell Biol *206*, 217-230.

Huang, H.Y., and Hopper, A.K. (2016). Multiple Layers of Stress-Induced Regulation in tRNA Biology. Life (Basel) *6*.

Huang, Y., Zhang, J.L., Yu, X.L., Xu, T.S., Wang, Z.B., and Cheng, X.C. (2013). Molecular functions of small regulatory noncoding RNA. Biochemistry (Mosc) *78*, 221-230.

Ishizu, H., Iwasaki, Y.W., Hirakata, S., Ozaki, H., Iwasaki, W., Siomi, H., and Siomi, M.C. (2015). Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb. Cell Rep *12*, 429-440.

Iwasaki, S., Kawamata, T., and Tomari, Y. (2009). Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. Mol Cell *34*, 58-67.

Iwasaki, Y.W., Murano, K., Ishizu, H., Shibuya, A., Iyoda, Y., Siomi, M.C., Siomi, H., and Saito, K. (2016). Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors Including Histone H1 to Repress Transposons. Mol Cell *63*, 408-419.

Iyengar, B.R., Choudhary, A., Sarangdhar, M.A., Venkatesh, K.V., Gadgil, C.J., and Pillai, B. (2014). Non-coding RNA interact to regulate neuronal development and function. Front Cell Neurosci *8*, 47.

Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. Science *330*, 1824-1827.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kalmykova, A.I., Klenov, M.S., and Gvozdev, V.A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. Nucleic Acids Res *33*, 2052-2059.

Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. EMBO J *29*, 3688-3700.

Kaufmann, J., and Smale, S.T. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. Genes Dev *8*, 821-829.

Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., Yasuda, J., *et al.* (2008). Hidden layers of human small RNAs. BMC Genomics *9*, 157.

Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' end formation of PIWI-interacting RNAs in vitro. Mol Cell *43*, 1015-1022.

Keam, S.P., Young, P.E., McCorkindale, A.L., Dang, T.H., Clancy, J.L., Humphreys, D.T., Preiss, T., Hutvagner, G., Martin, D.I., Cropley, J.E., *et al.* (2014). The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. Nucleic Acids Res *42*, 8984-8995.

Khazaie, Y., and Nasr Esfahani, M.H. (2014). MicroRNA and Male Infertility: A Potential for Diagnosis. Int J Fertil Steril *8*, 113-118.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., *et al.* (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Koch, C.M., Honemann-Capito, M., Egger-Adam, D., and Wodarz, A. (2009). Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. PLoS Genet *5*, e1000644.

Kumar, P., Mudunuri, S.B., Anaya, J., and Dutta, A. (2015). tRFdb: a database for transfer RNA fragments. Nucleic Acids Res *43*, D141-145.

Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A., and Dutta, A. (2018). tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. RNA *24*, 1093-1105.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.C., Luo, Y., Sachidanandam, R., Toth, K.F., *et al.* (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev *28*, 1667-1680.

Lee, J.H., Jung, C., Javadian-Elyaderani, P., Schweyer, S., Schutte, D., Shoukier, M., Karimi-Busheri, F., Weinfeld, M., Rasouli-Nia, A., Hengstler, J.G., *et al.* (2010). Pathways of proliferation and antiapoptosis driven in breast cancer stem cells by stem cell protein piwil2. Cancer Res *70*, 4569-4579.

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.

Lee, S.R., and Collins, K. (2005). Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. J Biol Chem *280*, 42744-42749.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., *et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. EMBO J *23*, 4051-4060.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004b). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Lee, Y.S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev *23*, 2639-2649.

Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frezal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., *et al.* (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol *2*, 174-181.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., *et al.* (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol *15*, 1501-1507.

Li, W., Prazak, L., Chatterjee, N., Gruninger, S., Krug, L., Theodorou, D., and Dubnau, J. (2013a). Activation of transposable elements during aging and neuronal decline in Drosophila. Nat Neurosci *16*, 529-531.

Li, Y., Lu, J., Han, Y., Fan, X., and Ding, S.W. (2013b). RNA interference functions as an antiviral immunity mechanism in mammals. Science *342*, 231-234.

Li, Y., Luo, J., Zhou, H., Liao, J.Y., Ma, L.M., Chen, Y.Q., and Qu, L.H. (2008). Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote Giardia lamblia. Nucleic Acids Res *36*, 6048-6055.

Li, Z., Ender, C., Meister, G., Moore, P.S., Chang, Y., and John, B. (2012). Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Res *40*, 6787-6799.

Liang, C., Xiong, K., Szulwach, K.E., Zhang, Y., Wang, Z., Peng, J., Fu, M., Jin, P., Suzuki, H.I., and Liu, Q. (2013). Sjogren syndrome antigen B (SSB)/La promotes global microRNA expression by binding microRNA precursors through stem-loop recognition. J Biol Chem *288*, 723-736.

Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. Development *120*, 1201-1211.

Lim, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A *104*, 6714-6719.

Lim, S.L., Qu, Z.P., Kortschak, R.D., Lawrence, D.M., Geoghegan, J., Hempfling, A.L., Bergmann, M., Goodnow, C.C., Ormandy, C.J., Wong, L., *et al.* (2015a). Correction: HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. PLoS Genet *11*, e1005782.

Lim, S.L., Qu, Z.P., Kortschak, R.D., Lawrence, D.M., Geoghegan, J., Hempfling, A.L., Bergmann, M., Goodnow, C.C., Ormandy, C.J., Wong, L., *et al.* (2015b). HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. PLoS Genet *11*, e1005620.

Liu, J.J., Shen, R., Chen, L., Ye, Y., He, G., Hua, K., Jarjoura, D., Nakano, T., Ramesh, G.K., Shapiro, C.L., *et al.* (2010). Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. Int J Clin Exp Pathol *3*, 328-337.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Liu, Y. (2016). MicroRNAs and PIWI-interacting RNAs in oncology. Oncol Lett *12*, 2289-2292.

Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., *et al.* (2013). MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res *41*, D262-267.

Machnicka, M.A., Olchowik, A., Grosjean, H., and Bujnicki, J.M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. RNA Biol *11*, 1619-1629.

Maillard, P.V., Ciaudo, C., Marchais, A., Li, Y., Jay, F., Ding, S.W., and Voinnet, O. (2013). Antiviral RNA interference in mammalian cells. Science *342*, 235-238.

Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. Nat Methods *10*, 957-963.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Marchand, V., Blanloeil-Oillo, F., Helm, M., and Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. Nucleic Acids Res *44*, e135.

Marquet, R., Isel, C., Ehresmann, C., and Ehresmann, B. (1995). tRNAs as primer of reverse transcriptases. Biochimie *77*, 113-124.

Martens-Uzunova, E.S., Jalava, S.E., Dits, N.F., van Leenders, G.J., Moller, S., Trapman, J., Bangma, C.H., Litman, T., Visakorpi, T., and Jenster, G. (2012). Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. Oncogene *31*, 978-991.

Martinez, G., Choudury, S.G., and Slotkin, R.K. (2017). tRNA-derived small RNAs target transposable element transcripts. Nucleic Acids Res *45*, 5142-5152.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Mei, Y., Yong, J., Liu, H., Shi, Y., Meinkoth, J., Dreyfuss, G., and Yang, X. (2010). tRNA binds to cytochrome c and inhibits caspase activation. Mol Cell *37*, 668-678.

Meignin, C., and Davis, I. (2008). UAP56 RNA helicase is required for axis specification and cytoplasmic mRNA localization in Drosophila. Dev Biol *315*, 89-98.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., *et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333-338.

Mevel-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615-1624.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science *348*, 812-817.

Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell *157*, 1364-1379.

Molla-Herman, A., Valles, A.M., Ganem-Elbaz, C., Antoniewski, C., and Huynh, J.R. (2015). tRNA processing defects induce replication stress and Chk2-dependent disruption of piRNA transcription. EMBO J *34*, 3009-3027.

Motorin, Y., and Helm, M. (2011). RNA nucleotide methylation. Wiley Interdiscip Rev RNA *2*, 611-631.

Motorin, Y., Muller, S., Behm-Ansmant, I., and Branlant, C. (2007). Identification of modified residues in RNAs by reverse transcription-based methods. Methods Enzymol *425*, 21-53.

Muerdter, F., Guzzardo, P.M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G.J. (2013). A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. Mol Cell *50*, 736-748.

Mugat, B., Akkouche, A., Serrano, V., Armenise, C., Li, B., Brun, C., Fulga, T.A., Van Vactor, D., Pelisson, A., and Chambeyron, S. (2015). MicroRNA-Dependent Transcriptional Silencing of Transposable Elements in Drosophila Follicle Cells. PLoS Genet *11*, e1005194.

Murota, Y., Ishizu, H., Nakagawa, S., Iwasaki, Y.W., Shibata, S., Kamatani, M.K., Saito, K., Okano, H., Siomi, H., and Siomi, M.C. (2014). Yb integrates piRNA intermediates and processing factors into perinuclear bodies to enhance piRISC assembly. Cell Rep *8*, 103-113.

Naar, A.M., Taatjes, D.J., Zhai, W., Nogales, E., and Tjian, R. (2002). Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. Genes Dev *16*, 1339-1344.

Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. RNA *13*, 1911-1922.

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M.K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., *et al.* (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature *491*, 284-287.

Nyswaner, K.M., Checkley, M.A., Yi, M., Stephens, R.M., and Garfinkel, D.J. (2008). Chromatin-associated genes protect the yeast genome from Ty1 insertional mutagenesis. Genetics *178*, 197-214.

Obbard, D.J., Jiggins, F.M., Halligan, D.L., and Little, T.J. (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol *16*, 580-585.

Ohtani, H., Iwasaki, Y.W., Shibuya, A., Siomi, H., Siomi, M.C., and Saito, K. (2013). DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. Genes Dev *27*, 1656-1661.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89-100.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

Olina, A.V., Kulbachinskiy, A.V., Aravin, A.A., and Esyunina, D.M. (2018). Argonaute Proteins and Mechanisms of RNA Interference in Eukaryotes and Prokaryotes. Biochemistry (Mosc) *83*, 483-497.

Olivieri, D., Senti, K.A., Subramanian, S., Sachidanandam, R., and Brennecke, J. (2012). The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. Mol Cell *47*, 954-969.

Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J *29*, 3301-3317.

Olvedy, M., Scaravilli, M., Hoogstrate, Y., Visakorpi, T., Jenster, G., and Martens-Uzunova, E.S. (2016). A comprehensive repertoire of tRNA-derived fragments in prostate cancer. Oncotarget *7*, 24766-24777.

Oom, A.L., Humphries, B.A., and Yang, C. (2014). MicroRNAs: novel players in cancer diagnosis and therapies. Biomed Res Int *2014*, 959461.

Ott, K.M., Nguyen, T., and Navarro, C. (2014). The DExH box helicase domain of spindle-E is necessary for retrotransposon silencing and axial patterning during Drosophila oogenesis. G3 (Bethesda) *4*, 2247-2257.

Ozen, M., Creighton, C.J., Ozdemir, M., and Ittmann, M. (2008). Widespread deregulation of microRNA expression in human prostate cancer. Oncogene *27*, 1788-1793.

Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Front Genet *6*, 2.

Pandey, R.R., and Pillai, R.S. (2014). Primary piRNA biogenesis: caught up in a Maelstrom. EMBO J *33*, 1979-1980.

Pane, A., Wehr, K., and Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell *12*, 851-862.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol *12*, 1484-1495.

Patil, V.S., and Kai, T. (2010). Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. Curr Biol *20*, 724-730.

Pekarsky, Y., Balatti, V., Palamarchuk, A., Rizzotto, L., Veneziano, D., Nigita, G., Rassenti, L.Z., Pass, H.I., Kipps, T.J., Liu, C.G., *et al.* (2016). Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. Proc Natl Acad Sci U S A *113*, 5071-5076.

Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. J Virol *81*, 1951-1960.

Pelisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J *13*, 4401-4411.

Pintard, L., Lecointe, F., Bujnicki, J.M., Bonnerot, C., Grosjean, H., and Lapeyre, B. (2002). Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. EMBO J *21*, 1811-1820.

Png, K.J., Halberg, N., Yoshida, M., and Tavazoie, S.F. (2011). A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. Nature *481*, 190-194.

Preall, J.B., Czech, B., Guzzardo, P.M., Muerdter, F., and Hannon, G.J. (2012). shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA *18*, 1446-1457.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697-711.

Purnell, B.A., Emanuel, P.A., and Gilmour, D.S. (1994). TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. Genes Dev *8*, 830-842.

Qi, H., Watanabe, T., Ku, H.Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. J Biol Chem *286*, 3789-3797.

Qiao, D., Zeeman, A.M., Deng, W., Looijenga, L.H., and Lin, H. (2002). Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. Oncogene *21*, 3988-3999.

Rangan, P., Malone, C.D., Navarro, C., Newbold, S.P., Hayes, P.S., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). piRNA production requires heterochromatin formation in Drosophila. Curr Biol *21*, 1373-1379.

Reese, T.A., Xia, J., Johnson, L.S., Zhou, X., Zhang, W., and Virgin, H.W. (2010). Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts. J Virol *84*, 10344-10353.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

Ren, X., Yang, Z., Mao, D., Chang, Z., Qiao, H.H., Wang, X., Sun, J., Hu, Q., Cui, Y., Liu, L.P., *et al.* (2014). Performance of the Cas9 nickase system in Drosophila melanogaster. G3 (Bethesda) *4*, 1955-1962.

Rouget, C., Papin, C., Boureux, A., Meunier, A.C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467*, 1128-1132.

Rozhkov, N.V., Hammell, M., and Hannon, G.J. (2013). Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev *27*, 400-412.

Rudinger-Thirion, J., Lescure, A., Paulus, C., and Frugier, M. (2011). Misfolded human tRNA isodecoder binds and neutralizes a 3' UTR-embedded Alu element. Proc Natl Acad Sci U S A *108*, E794-802.

Ryder, E., Ashburner, M., Bautista-Llacer, R., Drummond, J., Webster, J., Johnson, G., Morley, T., Chan, Y.S., Blows, F., Coulson, D., *et al.* (2007). The DrosDel deletion collection: a Drosophila genomewide chromosomal deficiency resource. Genetics *177*, 615-629.

Sabin, L.R., Delas, M.J., and Hannon, G.J. (2013). Dogma derailed: the many influences of RNA on the genome. Mol Cell *49*, 783-794.

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M.C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature *461*, 1296-1299.

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev *24*, 2493-2498.

Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev *20*, 2214-2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Saito, K., and Siomi, M.C. (2010). Small RNA-mediated quiescence of transposable elements in animals. Dev Cell *19*, 687-697.

Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics *166*, 1313-1321.

Sato, K., Iwasaki, Y.W., Shibuya, A., Carninci, P., Tsuchizawa, Y., Ishizu, H., Siomi, M.C., and Siomi, H. (2015). Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Mol Cell *59*, 553-563.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., and Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev *24*, 1590-1595.

Schmidt, A., Palumbo, G., Bozzetti, M.P., Tritto, P., Pimpinelli, S., and Schafer, U. (1999). Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. Genetics *151*, 749-760.

Schorn, A.J., Gutbrod, M.J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell *170*, 61-71 e11.

Schorn, A.J., and Martienssen, R. (2018). Tie-Break: Host and Retrotransposons Play tRNA. Trends Cell Biol.

Senti, K.A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet *26*, 499-509.

Shigematsu, M., and Kirino, Y. (2015). tRNA-Derived Short Non-coding RNA as Interacting Partners of Argonaute Proteins. Gene Regul Syst Bio *9*, 27-33.

Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y., and Kalmykova, A. (2014). Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. PLoS Genet *10*, e1004138.

Sienski, G., Batki, J., Senti, K.A., Donertas, D., Tirian, L., Meixner, K., and Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev *29*, 2258-2271.

Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964-980.

Simsek, M., and Adnan, H. (2000). Effect of single mismatches at 3'-end of primers on polymerase chain reaction. J Sci Res Med Sci *2*, 11-14.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Soares, A.R., and Santos, M. (2017). Discovery and function of transfer RNA-derived fragments and their role in disease. Wiley Interdiscip Rev RNA *8*.

Sobala, A., and Hutvagner, G. (2011). Transfer RNA-derived fragments: origins, processing, and functions. Wiley Interdiscip Rev RNA *2*, 853-862.

Speer, J., Gehrke, C.W., Kuo, K.C., Waalkes, T.P., and Borek, E. (1979). tRNA breakdown products as markers for cancer. Cancer *44*, 2120-2123.

St Johnston, D. (2013). Using mutants, knockdowns, and transgenesis to investigate gene function in Drosophila. Wiley Interdiscip Rev Dev Biol *2*, 587-613.

Suzuki, R., Honda, S., and Kirino, Y. (2012). PIWI Expression and Function in Cancer. Front Genet *3*, 204.

Tchurikov, N.A., and Kretova, O.V. (2011). Both piRNA and siRNA pathways are silencing transcripts of the suffix element in the Drosophila melanogaster germline and somatic cells. PLoS One *6*, e21882.

Thompson, D.M., Lu, C., Green, P.J., and Parker, R. (2008). tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA *14*, 2095-2103.

Thompson, D.M., and Parker, R. (2009). Stressing out over tRNA cleavage. Cell *138*, 215-219.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299-308.

Torres, A.G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. Trends Mol Med *20*, 306-314.

Tsuji, T., and Niida, Y. (2008). Development of a simple and highly sensitive mutation screening system by enzyme mismatch cleavage with optimized conditions for standard laboratories. Electrophoresis *29*, 1473-1483.

Urbonavicius, J., Rutkiene, R., Lopato, A., Tauraite, D., Stankeviciute, J., Aucynaite, A., Kaliniene, L., van Tilbeurgh, H., and Meskys, R. (2016). Evolution of tRNAPhe:imG2 methyltransferases involved in the biosynthesis of wyosine derivatives in Archaea. RNA *22*, 1871-1883.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

van den Beek, M., da Silva, B., Pouch, J., Ali Chaouche, M.E.A., Carre, C., and Antoniewski, C. (2018). Dual-layer transposon repression in heads of Drosophila melanogaster. RNA.

van Rij, R.P., Saleh, M.C., Berry, B., Foo, C., Houk, A., Antoniewski, C., and Andino, R. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev *20*, 2985-2995.

Venken, K.J., Carlson, J.W., Schulze, K.L., Pan, H., He, Y., Spokony, R., Wan, K.H., Koriabine, M., de Jong, P.J., White, K.P., *et al.* (2009). Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nat Methods *6*, 431-434.

Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science *314*, 1747-1751.

Vouillot, L., Thelie, A., and Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. G3 (Bethesda) *5*, 407-415.

Vourekas, A., Alexiou, P., Vrettos, N., Maragkakis, M., and Mourelatos, Z. (2016). Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm. Nature *531*, 390-394.

Wang, H., Ma, Z., Niu, K., Xiao, Y., Wu, X., Pan, C., Zhao, Y., Wang, K., Zhang, Y., and Liu, N. (2016). Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in Drosophila. Development *143*, 530-539.

Wang, S.H., and Elgin, S.C. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci U S A *108*, 21164-21169.

Wang, W., Han, B.W., Tipping, C., Ge, D.T., Zhang, Z., Weng, Z., and Zamore, P.D. (2015). Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell *59*, 819-830.

Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., and Ding, S.W. (2006). RNA interference directs innate immunity against viruses in adult Drosophila. Science *312*, 452-454.

Webster, A., Li, S., Hur, J.K., Wachsmuth, M., Bois, J.S., Perkins, E.M., Patel, D.J., and Aravin, A.A. (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. Mol Cell *59*, 564-575.

Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell *135*, 919-932.

Yamanaka, S., Siomi, M.C., and Siomi, H. (2014). piRNA clusters and open chromatin structure. Mob DNA *5*, 22.

Yamasaki, S., Ivanov, P., Hu, G.F., and Anderson, P. (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression. J Cell Biol *185*, 35-42.

Yamashiro, H., and Siomi, M.C. (2018). PIWI-Interacting RNA in Drosophila: Biogenesis, Transposon Regulation, and Beyond. Chem Rev *118*, 4404-4421.

Yeung, M.L., Bennasser, Y., Watashi, K., Le, S.Y., Houzet, L., and Jeang, K.T. (2009). Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. Nucleic Acids Res *37*, 6575-6586.

Yoon, J., Lee, K.S., Park, J.S., Yu, K., Paik, S.G., and Kang, Y.K. (2008). dSETDB1 and SU(VAR)3-9 sequentially function during germline-stem cell differentiation in Drosophila melanogaster. PLoS One *3*, e2234.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932-935.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science *350*, 339-342.

Yudelevich, A. (1971). Specific cleavage of an Escherichia coli leucine transfer RNA following bacteriophage T4 infection. J Mol Biol *60*, 21-29.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development *138*, 4039-4050.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., *et al.* (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871-884.

Zhang, P., Kang, J.Y., Gou, L.T., Wang, J., Xue, Y., Skogerboe, G., Dai, P., Huang, D.W., Chen, R., Fu, X.D., *et al.* (2015). MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. Cell Res *25*, 193-207.

Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell *157*, 1353-1363.

Zhao, H., Bojanowski, K., Ingber, D.E., Panigrahy, D., Pepper, M.S., Montesano, R., and Shing, Y. (1999). New role for tRNA and its fragment purified from human urinary bladder carcinoma conditioned medium: inhibition of endothelial cell growth. J Cell Biochem *76*, 109-117.