JOHANNES GUTENBERG UNIVERSITÄT MAINZ

"SMALL RNA PATHWAYS – EXPLORATION OF NEW PLAYERS, SUPPOSEDLY LOST SITES OF ACTION AND AN UNPERCEIVED DEFENSE STRATEGY"

Dissertation zur Erlangung des Akademischen Grades
„Doktor der Naturwissenschaften“
am Fachbereich Biologie
der Johannes Gutenberg-Universität in Mainz

Vorgelegt von Julia Jehn
geboren am 15. April 1991 in Groß-Umstadt

Mainz, September 2019

Dekan:
1. Gutachter:
2. Gutachter:

Tag der mündlichen Prüfung: 22. November 2019

“Ich respektiere das Gegebene. Daneben aber freilich auch das Werdende, denn eben dies Werdende wird über kurz oder lang abermals ein Gegebenes sein. Alles Alte, soweit es Anspruch darauf hat, sollen wir lieben, aber für das Neue sollen wir recht eigentlich leben. Und vor allem sollen wir, wie der Stechlin uns lehrt, den großen Zusammenhang der Dinge nie vergessen. Sich abschließen heißt sich einmauern, und sich einmauern ist Tod.”

Melusine (Der Stechlin; 29. Kapitel; Theodor Fontane)

Table of Contents

Deutsche Zusammenfassung

Seit ihrer Entdeckung in den 90er Jahren wurden die Wirkungsmechanismen kleiner RNAs intensiv untersucht, wodurch ein komplexes Geflecht aus interagierenden Mechanismen und Organismus-spezifischen Anpassungen erkennbar wurde. Dennoch konzentrieren sich die meisten Studien auf die vermeintlich kanonischen Funktionen der drei großen kleinen RNA-Klassen (miRNAs, siRNAs und piRNAs). Ziel dieser Dissertation ist es verschiedene unerforschte Aspekte kleiner RNA-Pathways zu beleuchten.

Zunächst werden neue Akteure im Universum der kleinen RNAs untersucht: tRNA-Fragmente (tRFs). In Kapitel 1 wird der aktuelle Wissensstand über tRFs dargestellt. In Kapitel 2 werden öffentlich verfügbare und eigene Sequenzierungsdaten kleiner RNAs analysiert, um einen ersten Überblick über die Expression von tRFs in verschiedenen Geweben und Arten zu geben. Diese Analysen deuten darauf hin, dass ein hohes Level an 5' tRNA-Hälften (5' tRHs) ein spezifisches Merkmal des Hippocampus von Primaten ist. Durch Modulation der Regulationskapazität ausgewählter 5' tRHs in HEK293T- und HepG2-Zellen und anschließender Analyse von RNA-Sequenzierungsdaten werden potenzielle Ziel-Transkripte identifiziert, welche zum Teil eine Rolle in der Neurogenese spielen. Darüber hinaus wird ein neuartiger k-mer-Ansatz verwendet, um die *Targeting*-Regel ausgewählter 5' tRHs zu identifizieren. Diese Analysen deuten darauf hin, dass anstelle eines miRNA-ähnlichen 5'-*seeds* eher der mittlere Teil der 5' tRHs für ein effektives *Silencing* wichtig ist.

Darüber hinaus wird der in der Keimbahn von Tieren gut erforschte piRNA-Pathway in dem bislang vernachlässigten Soma untersucht. In Kapitel 3 wird das PIWI-Repertoire und die Expression der jeweiligen PIWI Paraloge in der Keimbahn und in somatischen Geweben von Weichtieren analysiert. Exemplarisch für die Spitzschlammschnecke und die pazifische Auster wird gezeigt, dass PIWI-Proteine in somatischen Geweben flächendeckend exprimiert werden, was vergleichbare, in Arthropoden gewonnen Erkenntnisse bestärkt. Dies deutet darauf hin, dass ein funktioneller somatischer piRNA-Pathway der Ur-Zustand von Protostomen ist und nicht eine Besonderheit bestimmter Taxa. Die putativen piRNAs und piRNA-Cluster verschiedener Entwicklungsstadien werden zudem charakterisiert. Angesichts fehlender Antikörper gegen PIWI-Proteine, um potenzielle somatische piRNAs kopräzipitieren zu können, wird in Kapitel 4 ein neuartiger RNAi-basierter Ansatz beschrieben, um kleiner RNAs zu identifizieren, deren Expression von PLD6 abhängt. PLD6 ist eine Endonuklease, die unabdinglich für die Produktion von primären piRNAs in der tierischen Keimbahn ist.

Zuletzt wird eine mögliche, bisher unbekannte Strategie von einzelsträngigen RNA-Viren, kleinen RNA-basierten Abwehrmechanismen des Wirtes zu entkommen beschrieben, die durch die in Kapitel 5 beschriebenen Analysen der Dimension der strukturellen Selektion von protein-kodierenden Sequenzen offenbart wird. Da insbesondere bei einzelsträngigen RNA-Viren extreme Sekundärstrukturen von kodierende Sequenzen häufiger als erwartet auftreten, wird vermutet, dass sich die viralen Genome zugunsten hochgefalteter Transkripte entwickelt haben, um dem RNAi-System des Wirtes weniger Angriffsfläche zu bieten.

Abstract

Since their discovery in the 1990s, small RNA pathways were intensively studied unravelling a complex mesh of interacting mechanisms and organism-specific adaptations. Still, most studies concentrate on the alleged canonical functions of the three major small RNA classes (miRNAs, siRNAs and piRNAs). This thesis aims to tackle various unexplored aspects of small RNA pathways.

First, the new players in the small RNA universe, the tRNA fragments (tRFs), are investigated. In chapter 1, the current knowledge on tRFs (in the review named tRNA-derived small RNAs; tsRNAs) is reviewed. In chapter 2, publicly available and own small RNA sequencing data is analyzed to give an initial overview on tRF expression across various tissues and species. These analyses indicate that high levels of 5' tRNA-halves (5' tRHs) are a specific trait of the primate hippocampus. By modulating the regulation capacity of selected 5' tRHs in HEK293T and HepG2 cells and subsequent RNA sequencing analyses, potential target transcripts are identified which in part seem to be implicated in neurogenesis. Furthermore, a novel k-mer approach is used to dissect the targeting rule of selected 5' tRHs, indicating that instead of a miRNA-like 5' seed rather the middle part of 5' tRHs is important for effective silencing.

Second, the piRNA pathway, which is well studied in the germline of animals, is examined in the previously neglected soma. In chapter 3, the PIWI repertoire and expression is investigated in the germline and somatic tissues of two representative mollusks (the great pond snail and the pacific oyster) showing that PIWI proteins are ubiquitously expressed in the soma. Reinforcing analogous findings made for arthropods, this suggests that a functional somatic piRNA pathway is the ancestral state of protostomes and not a peculiarity of certain taxa. Furthermore, lineage-specific adaptations of piRNA targets and a dynamic expression of piRNA cluster during different developmental stages are shown. Faced with the non-availability of adequate antibodies against PIWI proteins to co-immunoprecipitate potential somatic piRNAs, in chapter 4, a novel RNAi-based approach is described in order to identify small RNAs whose expression depend on Phospholipase D family member 6 (PLD6), an endonuclease that is critical for the production of primary piRNAs in the animal germline.

Third, a previously unperceived strategy of single-stranded RNA viruses to evade small RNA-based defense mechanisms of the host is revealed by studying the dimension of structural selection of coding sequences, which is described in chapter 5. As especially for single-stranded RNA viruses extreme secondary structures of coding sequences are found to occur more often than expected by chance, it is suspected that viral genomes evolved in favor of highly folded transcripts in order to be less attackable for the host's RNAi system.

Introduction

Precisely regulated gene expression and genome integrity is crucial for the homeostasis of every living cell and the development of organisms. Besides protein-based regulators, small non-coding RNAs govern a big part of this. Small RNAs have several advantages over protein-based regulators, like transcription factors, as they evolve and act faster and their metabolic cost is rather low (Chen and Rajewsky 2007). Additionally, they offer tailor-made regulation as their sequence and structure can be arbitrarily varied (Beisel and Storz 2010).

Small RNAs were long under the radar of scientists, as they were neglected as degradation products of longer transcripts if they were perceived at all. In the 1990s, however, a series of stunning findings changed the view on these tiny molecules. When plant researchers aimed to overexpress a flower pigment gene by introducing a homologous transgene, they achieved silencing of the endogenous gene instead (Napoli et al. 1990; van der Krol et al. 1990). Next, studies in the nematode *Caenorhabditis elegans* showed that sense RNAs silence endogenous homologues equally effective as antisense RNAs (Guo and Kemphues 1995). Few years later, it was revealed that double-stranded RNA is silencing the endogenous gene even more effective than the single strands individually (Fire et al. 1998). Interestingly, only few double-stranded RNA molecules were required for an effective silencing effect, suggesting a catalytic or amplifying mechanism. After its discovery in nematodes, this dsRNA-mediated silencing phenomenon termed RNA interference (RNAi) was found within a broad spectrum of eukaryotes ranging from plants and fungi over planarians and cnidarians to insects (Waterhouse et al. 1998; Cogoni and Macino 1999; Sánchez Alvarado and Newmark 1999; Lohmann et al. 1999; Kennerdell and Carthew 1998).

A study in plants finally discovered that 25 nucleotide (nt) long antisense RNAs accumulate during RNAi, suggesting that these small RNAs are critical for silencing (Hamilton and Baulcombe 1999). It was then shown that cells process double-stranded RNAs to 21-25 nt fractions and that the resulting small antisense RNAs guide an RNA-induced silencing complex (RISC) to specifically cleave target mRNAs within the homologous region (Hammond et al. 2000; Zamore et al. 2000). Shortly after, proteins of the Argonaut family were identified as the core enzymes of the RISC (Tabara et al. 1999; Hammond et al. 2001). Argonauts are classified into three clades: the commonly shared AGOs, the animal-specific PIWIs and the worm-specific WAGOs. They share an N-terminal domain, a PAZ domain, a Mid domain and the PIWI domain (Hutvagner and Simard 2008). While the N-terminal domain is unwinding the duplex during RISC assembly (Kwak and Tomari 2012), the PAZ domain is anchoring the 3’ end of the small RNA and the interface between the Mid and the PIWI domain is binding its 5’ end. In case of catalytic active Argonauts, the endonucleolytic function is executed by the PIWI domain (Hutvagner and Simard 2008). While only few Argonauts like AGO2 are able to silence transcripts by cleavage (Liu et al. 2004), others mediate silencing by decapping or deadenylating target mRNAs or block translation (Behm-Ansmant et al. 2006; Lu et al. 2009; Olsen and Ambros 1999). Even though post-transcriptional silencing is the major mechanism observed for small RNA guided Argonauts, RISCs can also translocate to the nucleus, where

they initiate epigenetic silencing of target genes by recruiting factors remodeling chromatin or methylating DNA (Volpe et al. 2002; Zilberman et al. 2003; Pal-Bhadra et al. 2004; Robert et al. 2005; Bühler et al. 2006). It has additionally become apparent that small RNAs can act as transgenerational transmitters of epigenetic information (Brennecke et al. 2008; Gapp et al. 2014; Rechavi and Lev 2017; Neeb and Nowacki 2018). Only recently, a study revealed that neuronal small RNAs in *C. elegans* transgenerationally control the chemotaxis behavior by silencing transcripts in the germline showing that small RNAs do not care too much about the ‘‘Weismann Barrier’’ (Posner et al. 2019).

Both transcriptional and post-transcriptional RNAi pathways evolved as eukaryote-specific defense mechanisms against RNA viruses and transposable elements that later expanded to regulate the expression of protein-coding genes as well (Waterhouse et al. 2001; Borges and Martienssen 2015). Although prokaryotes do not possess RNAi, they likewise developed small RNA based defense mechanisms like CRISPR that have similar functionalities, but involve different protein effectors (Barrangou et al. 2007). While the key factors of RNAi are conserved in plants, fungi and animals, the mechanisms evolved individually within each organism leading to a plethora of different realizations of RNAi pathways in eukaryotes (Shabalina and Koonin 2008). Despite this, there are three major small RNA classes (classified by their biogenesis and associated Argonaut protein) that are known to regulate genes via RNAi: microRNAs (miRNAs), small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs). Other small RNA classes are small nucleolar RNAs (snoRNAs) and small nuclear RNA (snRNAs), as well as fragments of tRNAs (tRFs) and ribosomal RNAs (rRFs). While snoRNAs and snRNAs are known to be involved in RNA processing (Matera et al. 2007), tRFs and rRFs were long seen as mere degradation products of their abundant parental RNAs and were often depleted from small RNA sequencing data due to this reason. However, this view has changed over the last few years. Especially tRFs are becoming increasingly accepted as regulators of gene expression and virus replication (see chapter 1), but also rRFs are considered to be involved in translation control (Lambert et al. 2019). With the rise of deep sequencing techniques, it is likely that yet unknown small RNA species will be discovered in the near future.

miRNA and siRNA pathways

Interacting with proteins of the AGO-clade, siRNAs and miRNAs act in the soma and the germline of most eukaryotic species (Carthew and Sontheimer 2009). Originating from double stranded RNAs, these 20-25 nt long small RNAs share a similar biogenesis pathway that can be divided in two fundamental processes. First, Dicer cuts double stranded RNA generating a 21-23 nt long duplex with 3’ dinucleotide overhangs (Grishok et al. 2001; Ketting et al. 2001; Hutvágner et al. 2001). In the next step, the RISC assembly, an Argonaut protein first takes up this duplex before ejecting the passenger strand to keep only the guide strand (Leuschner et al. 2006; Yoda et al. 2010; Gu et al. 2011). Besides these shared features, there are many possible sources for double stranded RNA that can serve as precursor.

Canonical miRNA-precursors are transcribed by RNA polymerase II and fold to imperfectly paired hairpin structures that are cut by Drosha in animals and the Dicer homologue Dcl1 in plants (Lee et al. 2003; Basyuk et al. 2003; Kim 2005). In animals, introns that are liberated by splicing, so called mirtrons, can additionally serve as miRNA-precursors (Ruby et al. 2007; Okamura et al. 2007). While in animals the miRNA-precursors are diced in the cytoplasm, in plants, this process takes place in the nucleus and miRNAs are 2'-O-methylated at their 3' end by HEN1 before being exported to the cytoplasm (Kurihara and Watanabe 2004; Yu et al. 2005). For many miRNAs several isoforms exist that vary in length and sequence (Neilsen et al. 2012). These so-called isomiRs arise from variabilities in Drosha and Dicer cleavage, single-nucleotide extensions at the 3' end, exonucleolytic trimming and RNA editing (Morin et al. 2008; Neilsen et al. 2012). The different properties of the isomiRs influence not only the stability of the miRNA (Li et al. 2005; Lu et al. 2010; Ibrahim et al. 2010), but also affect target recognition (Wu et al. 2007; Seitz et al. 2008; Chiang et al. 2010; Starega-Roslan et al. 2011) and AGO loading (Azuma-Mukai et al. 2008; Mi et al. 2008; Montgomery et al. 2008; Takeda et al. 2008; Juvvuna et al. 2012).

Initially, it was assumed that siRNAs are only produced from exogenous sources, such as viruses and foreign genes. Later, however, so called endo-siRNAs were discovered that are generated from endogenous sources (Golden et al. 2008). Their precursors are generated by transcription of inverted repeats or dual strand encoded loci and display perfect base pairing. Plants, fungi and some animals like *C. elegans* additionally possess an RNA-dependent RNA polymerase (RdRP) that is able to synthesize a second strand on single stranded RNA templates (Sugiyama et al. 2005; Zong et al. 2009).

Both siRNAs and miRNAs recognize their targets mainly via sequence complementarity to their 7 nt seed region that ranges from the second to the eighth position of its 5' end (Zamore and Haley 2005; Wang 2014). In contrast to siRNAs and miRNAs of plants, miRNAs of animals do not require perfect complementarity to regulate their targets (Shabalina and Koonin 2008). Because of this imperfect binding, animal miRNAs can regulate many different targets. Target sites in animal mRNAs are predominantly located within the 3' UTR, while plant mRNAs are mostly targeted within the coding sequence (Millar and Waterhouse 2005). While siRNA and miRNAs in plants predominantly cleave their targets, the majority of animal miRNAs silence their targets by deadenylation, decapping and inhibition of translation (He and Hannon 2004). In addition to post-transcriptional silencing, siRNAs were found to regulate genes on a transcriptional level by inducing the methylation of promotors or restructuring chromatin via repressive histone marks (Mette et al. 2000; Volpe et al. 2002; Zilberman et al. 2003; Verdel et al. 2004; Bühler et al. 2006; Fagegaltier et al. 2009).

Even though siRNAs were shown to regulate endogenous genes in plants (Vazquez et al. 2004; Allen et al. 2005), most siRNAs target viruses and transposons (Obbard et al. 2009). Inversely, most miRNAs regulate protein-coding genes, while an additional role in transposon control is controversially discussed (Gebert and Rosenkranz 2015). Since siRNAs co-evolved within each species according to the respective threads and due to the conserved role of many miRNA

families in developmental processes, miRNAs are generally more conserved than siRNAs across large evolutionary distances (Bentwich et al. 2005; Axtell and Bowman 2008).

piRNA pathway

While the miRNA and the siRNA pathways are present in all eukaryotes, the piRNA pathway is specific to animals, where it is mainly described to silence transposons in the germline. piRNAs are with 26-31 nt substantially longer than miRNAs/siRNAs and originate from single stranded transcripts that are transcribed by RNA polymerase II (Aravin et al. 2006; Li et al. 2013). Most piRNA-precursors are derived from genomic regions, where sequence chunks of transposons are overrepresented, so-called piRNA clusters (Brennecke et al. 2007). These clusters either uni- or bidirectionally code for piRNA precursors. While unistrand clusters are the default piRNA source in most organisms, dualstrand clusters that produce piRNA precursors from both genomic strands are predominant in the *Drosophila* germline (Brennecke et al. 2007; Malone et al. 2009).

Precursor transcripts that are derived from unistrand clusters seem to lack common properties that make them distinguishable from protein-coding transcripts. Residing in euchromatic regions, their promotors likewise carry histone 3 lysine 4 dimethylation (H3K4me2) marks and the transcripts are similarly spliced, capped and polyadenylated as mRNAs (Li et al. 2013). By so far unknown mechanisms, however, only piRNA precursors are recruited to the nuage, a perinuclear granular structure in the cytoplasm where piRNA processing takes place (Rogers et al. 2017). Dual-strand clusters in the *Drosophila* germline reside within heterochromatin or in its close proximity and harbor H3K9me3-chromatin marks that get recognized by the Rhino-Deadlock-Cutoff Complex, which initiates transcription and controls transcript processing (Mohn et al. 2014; Chen et al. 2016). How the Rhino-Deadlock-Cutoff Complex specifically binds to piRNA clusters and not to other H3K9me3-marked genomic regions remains unclear. Recently, it has been shown that Bootlegger recruits Nxf3-Nxt1, a variant of the mRNA export receptor dimer Nxf1-Nxt1, to these heterochromatic clusters, which mediates the nuclear export of piRNA precursors and directs their delivery to the nuage (ElMaghraby et al. 2019).

In the nuage (Yb bodies in *Drosophila* follicle cells), piRNA precursors get cut by the mitochondrially-bound endonuclease Zucchini (PLD6 in humans) releasing so-called primary piRNAs that have a pronounced bias for a uridine at the first position of their 5’ end (1U bias) (Huang et al. 2011; Watanabe et al. 2011; Ipsaro et al. 2012; Nishimasu et al. 2012). In *Drosophila*, these primary piRNAs are loaded into P-element induced wimpy testis (Piwi) or Aubergine (Aub) (Vagin et al. 2006) and get 2’ O-methylated at their 3’ end by Hen1 (Horwich et al. 2007; Saito et al. 2007). When loaded on Piwi, the piRNA-Piwi-complex will translocate to the nucleus and silence transposons on a transcriptional level (Le Thomas et al. 2013). piRNA-Aub-complexes, however, stay in the cytoplasm. Here, transposon transcripts, which are recognized by antisense piRNAs via a miRNA-like pairing, get sliced by the associated Aub in a 10-nt offset from the piRNA’s 5’ end (Gunawardane et al. 2007; Brennecke et al. 2007;

Shen et al. 2018). The slicing product complementary to the primary piRNA will therefore likely contain an adenosine at its tenth position (10A bias) and when taken up by Argonaut 3 (Ago3) becomes a so-called secondary piRNA that gets resected by Nibbler and 2'-O-methylated by Hen1 (Hayashi et al. 2016). The secondary piRNAs itself can direct the associated Ago3 to a piRNA precursor, whose cleavage again gives rise to primary piRNAs. As by this process primary and secondary piRNAs get amplified in a self-sustaining loop by consuming piRNA precursors and target transcripts likewise, it is commonly referred to as ping-pong loop (Brennecke et al. 2007). Consistently, the 10-nucleotide sequence overlap between primary and secondary piRNAs, which is a key feature of the piRNA pathway, is termed ping-pong signature. This mechanism only amplifies piRNAs of the same sequence. Triggered by secondary piRNAs, Zucchini can additionally produce piRNAs in 3'-directed phased manner, which allows to diversify the piRNA pool (Mohn et al. 2015; Han et al. 2015). Again, these Zucchini-generated piRNAs have a strong 1U bias. As *in vitro* Zucchini does not show any nucleotide preference, it is likely that a nuage-specific cofactor is driving this specificity (Ipsaro et al. 2012; Nishimasu et al. 2012). While in *Drosophila*, phased Zucchini cleavage is defining the 5' and the 3' end at the same time, in mice the 3' end of piRNAs needs to be resected by PNLDC1 (Mohn et al. 2015; Ding et al. 2017). Generally, piRNAs have different lengths depending on the associated PIWI paralogue (Aravin et al. 2006; Girard et al. 2006; Brennecke et al. 2007; Aravin et al. 2008).

The number of expressed PIWI paralogues varies greatly amongst different metazoan lineages. As already described, *Drosophila melanogaster* is equipped with the three PIWI paralogues Piwi, Aub and Ago3. However, the insect lineage faced diverse duplication and diversification events of RNAi genes that also include changes of the PIWI repertoire (Lewis et al. 2016). The mosquito *Aedes aegypti*, for instance, bears 7 isoforms of Piwi and expresses Ago3 (Campbell et al. 2008), while the silkworm *Bombyx mori* only harbors the two PIWI paralogues SIWI and BmAgo3 (Kawaoka et al. 2008). Not only the amount of PIWI genes varies, but also the expression distribution. While in *D. melanogaster* the PIWI proteins are solely expressed in the germline (with the exception of germline-adjacent follicle cells), most arthropods show additional somatic expression (Lewis et al. 2018). In vertebrates, however, PIWI expression appears to be restricted to the germline (Tosar et al. 2018). While zebrafish only express the two PIWI paralogues Ziwi and Zili (Houwing et al. 2008), mammals generally express four PIWI paralogues, Piwil1-4 (Girard et al. 2006; Aravin et al. 2006; Aravin et al. 2008). Mice, however, lack a homologue of the oocyte-specific Piwil3, whose function is taken over by a Dicer isoform (Flemr et al. 2013; Roovers et al. 2015).

As revealed for the germline development of mice, expression of the different PIWI paralogues is dynamic. While Mili is expressed already in primordial germ cells and persists throughout spermatogenesis in the adult animal, Miwi2 expression is limited to early gonocytes, when de novo methylation takes place and transposons are active (Aravin et al. 2008). In contrast, the expression of Miwi begins in the pachytene stage of meiosis and lasts until the end of meiosis (Deng and Lin 2002; Aravin et al. 2008). The shifted expression is accompanied with different sets of piRNA populations. piRNAs of pre-pachytene germ cells are associated with transposon

control, whereas the more abundant pachytene piRNAs are rather depleted of transposon-derived sequences (Aravin et al. 2006; Girard et al. 2006; Li et al. 2013). The function of pachytene piRNAs remains elusive as most of their sequences map only to the locus producing them. While several studies suggest that pachytene piRNAs are involved in the post-transcriptional regulation of protein-coding genes (Gou et al. 2014; Gebert et al. 2015; Zhang et al. 2015), another study suggests that they stabilize mRNAs in a sequence-independent manner (Vourekas et al. 2012). A piRNA function beyond transposon control was also observed in the *Drosophila* embryo, where the piRNA pathway triggers the degradation of maternal mRNAs by recruiting deadenylation complexes (Rouget et al. 2010). Furthermore, the piRNA pathway was shown to be involved in the viral defense of mosquitos (Morazzani et al. 2012; Schnettler et al. 2013; Miesen et al. 2015) and was found to play a role in the control of memory-related synaptic plasticity in the sea slug *Aplysia californica* (Rajasethupathy et al. 2012). While silencing transposons seems to be the ancestral function of the piRNA pathway (Aravin et al. 2007; Lewis et al. 2018), these findings are probably just the tip of the iceberg and more piRNA pathway functions beyond transposon control will certainly be discovered in the future.

Co-evolution of small RNA pathways and their targets

A fascinating aspect of small RNA regulation is that it comes in plethora of different shapes and styles, which is a consequence of the immense potential of small RNA pathways to adapt quickly to both endogenous and exogenous forces. Thus, in up to 1.6 billion years of eukaryotic evolution (Wang et al. 1999; Simpson and Roger 2004), manifold duplication and specification events of the core components of small RNA pathways took independently place in the eukaryotic lineages giving rise to different realizations of small RNA biogenesis and effector factors (Chapman and Carrington 2007). A prime example for the adaptation to species-specific threads is the piRNA pathway. While the transposon classes represented in the piRNA clusters are conserved among *Drosophila* species, the individual elements of these classes are not, suggesting that piRNA clusters co-evolve with the respective transposon repertoire (Malone et al. 2009). Similar observations were made for the nematode *C. elegans*, where piRNA clusters evolve to target active transposons (Bagijn et al. 2012). Additionally to piRNA cluster adaptation, there is evidence that RNAi genes are under recurrent positive selection to keep genomic parasites in check (Kolaczkowski et al. 2011; Obbard et al. 2011). However, it is not only the piRNA pathway evolving to catch up with emerging transposable elements. Evolutionary forces also act on the prosecuted transposons to evade piRNA-mediated repression leading to an evolutionary arms race between small RNA silencing pathways and transposons. For instance, it has been noted that during the evolution of transposons of the LINE-1 (L1) family, the 5’ UTRs of L1s frequently changed (Khan et al. 2006). Since MILI and MIWI2 repress L1s by establishing de novo DNA marks at the 5’ UTR of L1s in mice (Kuramochi-Miyagawa et al. 2008), exchanging the 5’ UTR might be a silencing-avoiding strategy. In other cases, transposons were found to evade the silencing process by either expressing proteins that are able to revert repressing methylation marks from its promotor (Cui and Fedoroff

2002), or capturing host gene sequences to potentially avoid methylation by confusing the host defense system (Juretic et al. 2005; Hollister et al. 2011).

Relentless co-evolutionary struggle with one another has also been noted for plants and viruses. While plants produce virus-derived siRNAs to protect themselves from viral infection, plant viruses have evolved several suppressors of small RNA-mediated silencing (Ding and Voinnet 2007). But also animal viruses were found to express proteins that counteract RNAi control (Li et al. 2002; Li et al. 2004; Lecellier et al. 2005; Otsuka et al. 2007). Such viral suppressors of RNA silencing (VSRs) inhibit different steps of small RNA pathways. For instance, the B2 proteins of alphanoda- and betanodaviruses protect double-stranded RNA from being processed by Dicer, which impedes the production of anti-viral small RNAs (Lingel et al. 2005; Fenner et al. 2007). Other plant VSRs like P1/HC-Pro, p21 and p19 trigger the degradation of anti-viral small RNAs by inhibiting microRNA methylation (Yu et al. 2006), while the 2b protein of the cucumber mosaic virus inhibits silencing by associating with the PAZ-domain of siRNA-loaded AGO1 (Zhang et al. 2006). For the respiratory syncytial virus (RSV), it has been additionally shown that small RNA pathways of the host can be exploited by viruses to abrogate anti-viral defense mechanism (Deng et al. 2015).

Apart from transposons and viruses, small RNA pathways have also co-evolved with protein-coding targets. This is best observable for miRNA-mediated gene regulation, whose development has been hypothesized to be a crucial step for multicellular organisms to evolve due to their role in spatiotemporal gene regulation (Moran et al. 2017). It has been shown that nonconserved miRNA target sites most likely evolve in the 3’ UTR of genes, which are expressed in tissues where the miRNA is absent. In contrast, genes that are preferentially expressed spatiotemporal with a miRNA evolved to selectively lack target sites for this miRNA (Farh et al. 2005). In order not to accidently provide a target site for miRNAs, genes that carry out basal cellular functions, so called housekeeping genes, have chosen a special strategy. In order to avoid fortuitously nonconserved miRNA target sites to arise, they developed short 3’ UTRs (Stark et al. 2005). By the means of conserved targeting, selective avoidance and emerging nonconserved targeting, miRNAs may have shaped the evolution of nearly all mammalian mRNAs (Bartel 2009).

Aims of the thesis

One focus of this thesis is to shed light on a less well-studied small RNA class, the tRNA-derived fragments. Here, an initial overview of the tRF expression profile across tissues and species will be given and the gene regulatory mechanism of a certain subclass, the 5’ tRNA-halves, will be studied. Besides the identification of potential tRF targets, it will be investigated whether 5’ tRHs regulate their targets by miRNA- or piRNA-like mechanisms and whether targeting is conserved across species.

Another aim of this thesis is to study a putative somatic function of the piRNA pathway in the bilateralian branch. In contrast to non-bilaterian species, where PIWI proteins were found to

be expressed in somatic stem cells (Funayama et al. 2010; Alié et al. 2011; Seipel et al. 2004; Juliano et al. 2014), in bilaterian species the piRNA pathway was considered to be restricted to the germline and early developmental processes until recently (Lewis et al. 2018). Here, a deeper analysis of molluskan species, which had yet been unstudied in this context, shall unravel whether a somatic piRNA pathway is a conserved feature of the molluskan lineage.

While there is growing evidence that a somatic function of the piRNA pathway has been the ancestral state that later got lost in some species, especially in the branch of vertebrates (Lewis et al. 2018; Palakodeti et al. 2008; Perrat et al. 2013; Jones et al. 2016; Teixeira et al. 2017; Juliano et al. 2014; Funayama et al. 2010), some studies suggest that a somatic function persists in vertebrates in few special niches (Nandi et al. 2016; Ross et al. 2014). Faced with the non-availability of adequate antibodies for pull-down experiments with PIWI proteins, this hypothesis is only based on indications by bioinformatics analyses so far. Hence, another aim of this thesis is to develop an antibody-independent *in vitro* method that enables the detection and further characterization of putative piRNAs in somatic cells.

It was shown that natural selection can act directly on the level of DNA or RNA, favoring specific structures of single stranded DNA or RNA molecules (Katz and Burge 2003; Chamary and Hurst 2005; Hoede et al. 2006; Fricke et al. 2018). Currently it is unknown whether this is a peculiarity of only a few genomic loci in a limited number of species, or rather a widespread phenomenon throughout life. Therefore, another aim of this thesis is to evaluate on a genome-wide scale if the synonymous codons (codons that do not change the amino acid sequence of the encoded protein) within the open reading frame are selected towards higher or lower folded mRNA structures. Furthermore, potential explanations for the selection of extreme secondary structures will be discussed.

References

Alié, Alexandre; Leclère, Lucas; Jager, Muriel; Dayraud, Cyrielle; Chang, Patrick; Le Guyader, Hervé et al. (2011): Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: ancient association of "germline genes" with stemness. In *Developmental biology* 350 (1), pp. 183–197. DOI: 10.1016/j.ydbio.2010.10.019.

Allen, Edwards; Xie, Zhixin; Gustafson, Adam M.; Carrington, James C. (2005): microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. In *Cell* 121 (2), pp. 207–221. DOI: 10.1016/j.cell.2005.04.004.

Aravin, Alexei; Gaidatzis, Dimos; Pfeffer, Sébastien; Lagos-Quintana, Mariana; Landgraf, Pablo; Iovino, Nicola et al. (2006): A novel class of small RNAs bind to MILI protein in mouse testes. In *nature* 442 (7099), pp. 203–207. DOI: 10.1038/nature04916.

Aravin, Alexei A.; Hannon, Gregory J.; Brennecke, Julius (2007): The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. In *Science (New York, N.Y.)* 318 (5851), pp. 761–764. DOI: 10.1126/science.1146484.

Aravin, Alexei A.; Sachidanandam, Ravi; Bourc'his, Deborah; Schaefer, Christopher; Pezic, Dubravka; Toth, Katalin Fejes et al. (2008): A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. In *Molecular cell* 31 (6), pp. 785–799. DOI: 10.1016/j.molcel.2008.09.003.

Axtell, Michael J.; Bowman, John L. (2008): Evolution of plant microRNAs and their targets. In *Trends in Plant Science* 13 (7), pp. 343–349. DOI: 10.1016/j.tplants.2008.03.009.

Azuma-Mukai, Asuka; Oguri, Hideo; Mituyama, Toutai; Qian, Zhi Rong; Asai, Kiyoshi; Siomi, Haruhiko; Siomi, Mikiko C. (2008): Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. In *PNAS* 105 (23), pp. 7964–7969. DOI: 10.1073/pnas.0800334105.

Bagijn, Marloes P.; Goldstein, Leonard D.; Sapetschnig, Alexandra; Weick, Eva-Maria; Bouasker, Samir; Lehrbach, Nicolas J. et al. (2012): Function, targets, and evolution of Caenorhabditis elegans piRNAs. In *Science (New York, N.Y.)* 337 (6094), pp. 574–578. DOI: 10.1126/science.1220952.

Barrangou, Rodolphe; Fremaux, Christophe; Deveau, Hélène; Richards, Melissa; Boyaval, Patrick; Moineau, Sylvain et al. (2007): CRISPR provides acquired resistance against viruses in prokaryotes. In *Science (New York, N.Y.)* 315 (5819), pp. 1709–1712. DOI: 10.1126/science.1138140.

Bartel, David P. (2009): MicroRNAs: Target Recognition and Regulatory Functions. In *Cell* 136 (2), pp. 215–233. DOI: 10.1016/j.cell.2009.01.002.

Basyuk, Eugenia; Suavet, Florence; Doglio, Alain; Bordonné, Rémy; Bertrand, Edouard (2003): Human let-7 stem-loop precursors harbor features of RNase III cleavage products. In *Nucleic acids research* 31 (22), pp. 6593–6597. DOI: 10.1093/nar/gkg855.

Behm-Ansmant, Isabelle; Rehwinkel, Jan; Doerks, Tobias; Stark, Alexander; Bork, Peer; Izaurralde, Elisa (2006): mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. In *Genes & Development* 20 (14), pp. 1885–1898. DOI: 10.1101/gad.1424106.

Beisel, Chase L.; Storz, Gisela (2010): Base pairing small RNAs and their roles in global regulatory networks. In *FEMS microbiology reviews* 34 (5), pp. 866–882. DOI: 10.1111/j.1574-6976.2010.00241.x.

Bentwich, Isaac; Avniel, Amir; Karov, Yael; Aharonov, Ranit; Gilad, Shlomit; Barad, Omer et al. (2005): Identification of hundreds of conserved and nonconserved human microRNAs. In *Nature genetics* 37 (7), p. 766. DOI: 10.1038/ng1590.

Borges, Filipe; Martienssen, Robert A. (2015): The expanding world of small RNAs in plants. In *Nature reviews. Molecular cell biology* 16 (12), pp. 727–741. DOI: 10.1038/nrm4085.

Brennecke, Julius; Aravin, Alexei A.; Stark, Alexander; Dus, Monica; Kellis, Manolis; Sachidanandam, Ravi; Hannon, Gregory J. (2007): Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. In *Cell* 128 (6), pp. 1089–1103. DOI: 10.1016/j.cell.2007.01.043.

Brennecke, Julius; Malone, Colin D.; Aravin, Alexei A.; Sachidanandam, Ravi; Stark, Alexander; Hannon, Gregory J. (2008): An epigenetic role for maternally inherited piRNAs in transposon silencing. In *Science (New York, N.Y.)* 322 (5906), pp. 1387–1392. DOI: 10.1126/science.1165171.

Bühler, Marc; Verdel, André; Moazed, Danesh (2006): Tethering RITS to a Nascent Transcript Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. In *Cell* 125 (5), pp. 873–886. DOI: 10.1016/j.cell.2006.04.025.

Campbell, Corey L.; Black, William C.; Hess, Ann M.; Foy, Brian D. (2008): Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. In *BMC Genomics* 9 (1), pp. 1–16. DOI: 10.1186/1471-2164-9-425.

Carthew, Richard W.; Sontheimer, Erik J. (2009): Origins and Mechanisms of miRNAs and siRNAs. In *Cell* 136 (4), pp. 642–655. DOI: 10.1016/j.cell.2009.01.035.

Chamary, J. V.; Hurst, Laurence D. (2005): Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. In *Genome biology* 6 (9), R75.

Chapman, Elisabeth J.; Carrington, James C. (2007): Specialization and evolution of endogenous small RNA pathways. In *nrg* 8 (11), pp. 884–896. DOI: 10.1038/nrg2179.

Chen, Kevin; Rajewsky, Nikolaus (2007): The evolution of gene regulation by transcription factors and microRNAs. In *nrg* 8 (2), pp. 93–103. DOI: 10.1038/nrg1990.

Chen, Yung-Chia Ariel; Stuwe, Evelyn; Luo, Yicheng; Ninova, Maria; Le Thomas, Adrien; Rozhavskaya, Ekaterina et al. (2016): Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. In *Molecular cell* 63 (1), pp. 97–109. DOI: 10.1016/j.molcel.2016.05.010.

Chiang, H. Rosaria; Schoenfeld, Lori W.; Ruby, J. Graham; Auyeung, Vincent C.; Spies, Noah; Baek, Daehyun et al. (2010): Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. In *Genes & Development* 24 (10), pp. 992–1009. DOI: 10.1101/gad.1884710.

Cogoni, Carlo; Macino, Giuseppe (1999): Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. In *Nature* 399 (6732), p. 166.

Cui, Hongchang; Fedoroff, Nina V. (2002): Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein. In *The Plant cell* 14 (11), pp. 2883–2899. DOI: 10.1105/tpc.006163.

Deng, Junfang; Ptashkin, Ryan N.; Chen, Yu; Cheng, Zhi; Liu, Guangliang; Phan, Thien et al. (2015): Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism. In *Molecular therapy : the journal of the American Society of Gene Therapy* 23 (10), pp. 1622–1629. DOI: 10.1038/mt.2015.124.

Deng, Wei; Lin, Haifan (2002): miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. In *Developmental cell* 2 (6), pp. 819–830. DOI: 10.1016/S1534-5807(02)00165-X.

Ding, Deqiang; Liu, Jiali; Dong, Kunzhe; Midic, Uros; Hess, Rex A.; Xie, Huirong et al. (2017): PNLDC1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. In *Nature communications* 8 (1), p. 819. DOI: 10.1038/s41467-017-00854-4.

Ding, Shou-Wei; Voinnet, Olivier (2007): Antiviral immunity directed by small RNAs. In *Cell* 130 (3), pp. 413–426. DOI: 10.1016/j.cell.2007.07.039.

ElMaghraby, Mostafa F.; Andersen, Peter Refsing; Pühringer, Florian; Hohmann, Ulrich; Meixner, Katharina; Lendl, Thomas et al. (2019): A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. In *Cell* 178 (4), 964-979.e20. DOI: 10.1016/j.cell.2019.07.007.

Fagegaltier, Delphine; Bougé, Anne-Laure; Berry, Bassam; Poisot, Émilie; Sismeiro, Odile; Coppée, Jean-Yves et al. (2009): The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. In *PNAS* 106 (50), pp. 21258–21263. DOI: 10.1073/pnas.0809208105.

Farh, Kyle Kai-How; Grimson, Andrew; Jan, Calvin; Lewis, Benjamin P.; Johnston, Wendy K.; Lim, Lee P. et al. (2005): The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. In *Science (New York, N.Y.)* 310 (5755), pp. 1817–1821. DOI: 10.1126/science.1121158.

Fenner, Beau J.; Goh, Winnie; Kwang, Jimmy (2007): Dissection of double-stranded RNA binding protein B2 from betanodavirus. In *Journal of virology* 81 (11), pp. 5449–5459. DOI: 10.1128/JVI.00009-07.

Fire, Andrew; Xu, SiQun; Montgomery, Mary K.; Kostas, Steven A.; Driver, Samuel E.; Mello, Craig C. (1998): Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. In *Nature* 391 (6669), p. 806.

Flemr, Matyas; Malik, Radek; Franke, Vedran; Nejepinska, Jana; Sedlacek, Radislav; Vlahovicek, Kristian; Svoboda, Petr (2013): A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. In *Cell* 155 (4), pp. 807–816. DOI: 10.1016/j.cell.2013.10.001.

Fricke, Markus; Gerst, Ruman; Ibrahim, Bashar; Niepmann, Michael; Marz, Manja (2018): Global importance of RNA secondary structures in protein-coding sequences. In *Bioinformatics (Oxford, England)* 35 (4), pp. 579–583.

Funayama, Noriko; Nakatsukasa, Mikiko; Mohri, Kurato; Masuda, Yoshiki; Agata, Kiyokazu (2010): Piwi expression in archeocytes and choanocytes in demosponges: insights into the stem cell system in demosponges. In *Evolution & development* 12 (3), pp. 275–287. DOI: 10.1111/j.1525-142X.2010.00413.x.

Gapp, Katharina; Jawaid, Ali; Sarkies, Peter; Bohacek, Johannes; Pelczar, Pawel; Prados, Julien et al. (2014): Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. In *Nature neuroscience* 17 (5), pp. 667–669. DOI: 10.1038/nn.3695.

Gebert, Daniel; Ketting, René F.; Zischler, Hans; Rosenkranz, David (2015): piRNAs from Pig Testis Provide Evidence for a Conserved Role of the Piwi Pathway in Post-Transcriptional Gene Regulation in Mammals. In *PloS one* 10 (5), e0124860. DOI: 10.1371/journal.pone.0124860.

Gebert, Daniel; Rosenkranz, David (2015): RNA-based regulation of transposon expression. In *Wiley interdisciplinary reviews. RNA* 6 (6), pp. 687–708. DOI: 10.1002/wrna.1310.

Girard, Angélique; Sachidanandam, Ravi; Hannon, Gregory J.; Carmell, Michelle A. (2006): A germline-specific class of small RNAs binds mammalian Piwi proteins. In *nature* 442 (7099), pp. 199–202. DOI: 10.1038/nature04917.

Golden, Daniel E.; Gerbasi, Vincent R.; Sontheimer, Erik J. (2008): An Inside Job for siRNAs. In *Molecular cell* 31 (3), pp. 309–312. DOI: 10.1016/j.molcel.2008.07.008.

Gou, Lan-Tao; Dai, Peng; Yang, Jian-Hua; Xue, Yuanchao; Hu, Yun-Ping; Zhou, Yu et al. (2014): Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. In *Cell research* 24, 680 EP -. DOI: 10.1038/cr.2014.41.

Grishok, Alla; Pasquinelli, Amy E.; Conte, Darryl; Li, Na; Parrish, Susan; Ha, Ilho et al. (2001): Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing. In *Cell* 106 (1), pp. 23–34. DOI: 10.1016/S0092-8674(01)00431-7.

Gu, Shuo; Jin, Lan; Zhang, Feijie; Huang, Yong; Grimm, Dirk; Rossi, John J.; Kay, Mark A. (2011): Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (22), pp. 9208–9213. DOI: 10.1073/pnas.1018023108.

Gunawardane, Lalith S.; Saito, Kuniaki; Nishida, Kazumichi M.; Miyoshi, Keita; Kawamura, Yoshinori; Nagami, Tomoko et al. (2007): A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. In *Science (New York, N.Y.)* 315 (5818), pp. 1587–1590. DOI: 10.1126/science.1140494.

Guo, Su; Kemphues, Kenneth J. (1995): par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. In *Cell* 81 (4), pp. 611–620.

Hamilton, A. J.; Baulcombe, D. C. (1999): A species of small antisense RNA in posttranscriptional gene silencing in plants. In *Science (New York, N.Y.)* 286 (5441), pp. 950–952. DOI: 10.1126/science.286.5441.950.

Hammond, S. M.; Boettcher, S.; Caudy, A. A.; Kobayashi, R.; Hannon, G. J. (2001): Argonaute2, a link between genetic and biochemical analyses of RNAi. In *Science (New York, N.Y.)* 293 (5532), pp. 1146–1150. DOI: 10.1126/science.1064023.

Hammond, Scott M.; Bernstein, Emily; Beach, David; Hannon, Gregory J. (2000): An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. In *Nature* 404 (6775), p. 293.

Han, Bo W.; Wang, Wei; Li, Chengjian; Weng, Zhiping; Zamore, Phillip D. (2015): piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. In *Science (New York, N.Y.)* 348 (6236), pp. 817–821. DOI: 10.1126/science.aaa1264.

Hayashi, Rippei; Schnabl, Jakob; Handler, Dominik; Mohn, Fabio; Ameres, Stefan L.; Brennecke, Julius (2016): Genetic and mechanistic diversity of piRNA 3'-end formation. In *Nature* 539 (7630), pp. 588–592. DOI: 10.1038/nature20162.

He, Lin; Hannon, Gregory J. (2004): MicroRNAs: small RNAs with a big role in gene regulation. In *Nature Reviews Genetics* 5 (7), p. 522. DOI: 10.1038/nrg1379.

Hoede, Claire; Denamur, Erick; Tenaillon, Olivier (2006): Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. In *PLoS genetics* 2 (11), e176.

Hollister, Jesse D.; Smith, Lisa M.; Guo, Ya-Long; Ott, Felix; Weigel, Detlef; Gaut, Brandon S. (2011): Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (6), pp. 2322–2327. DOI: 10.1073/pnas.1018222108.

Horwich, Michael D.; Li, Chengjian; Matranga, Christian; Vagin, Vasily; Farley, Gwen; Wang, Peng; Zamore, Phillip D. (2007): The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. In *Current Biology* 17 (14), pp. 1265–1272. DOI: 10.1016/j.cub.2007.06.030.

Houwing, Saskia; Berezikov, Eugene; Ketting, René F. (2008): Zili is required for germ cell differentiation and meiosis in zebrafish. In *The EMBO journal* 27 (20), pp. 2702–2711. DOI: 10.1038/emboj.2008.204.

Huang, Huiyan; Gao, Qun; Peng, Xiaoxue; Choi, Seok-Yong; Sarma, Krishna; Ren, Hongmei et al. (2011): piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. In *Developmental cell* 20 (3), pp. 376–387. DOI: 10.1016/j.devcel.2011.01.004.

Hutvagner, Gyorgy; Simard, Martin J. (2008): Argonaute proteins: key players in RNA silencing. In *Nature reviews. Molecular cell biology* 9 (1), pp. 22–32. DOI: 10.1038/nrm2321.

Hutvágner, György; McLachlan, Juanita; Pasquinelli, Amy E.; Bálint, Éva; Tuschl, Thomas; Zamore, Phillip D. (2001): A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. In *Science* 293 (5531), pp. 834–838. DOI: 10.1126/science.1062961.

Ibrahim, Fadia; Rymarquis, Linda A.; Kim, Eun-Jeong; Becker, James; Balassa, Eniko; Green, Pamela J.; Cerutti, Heriberto (2010): Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. In *PNAS* 107 (8), pp. 3906–3911. DOI: 10.1073/pnas.0912632107.

Ipsaro, Jonathan J.; Haase, Astrid D.; Knott, Simon R.; Joshua-Tor, Leemor; Hannon, Gregory J. (2012): The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. In *Nature* 491 (7423), pp. 279–283. DOI: 10.1038/nature11502.

Jones, Brian C.; Wood, Jason G.; Chang, Chengyi; Tam, Austin D.; Franklin, Michael J.; Siegel, Emily R.; Helfand, Stephen L. (2016): A somatic piRNA pathway in the Drosophila fat body ensures metabolic homeostasis and normal lifespan. In *Nat Commun* 7 (1), p. 691. DOI: 10.1038/ncomms13856.

Juliano, Celina E.; Reich, Adrian; Liu, Na; Götzfried, Jessica; Zhong, Mei; Uman, Selen et al. (2014): PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (1), pp. 337–342. DOI: 10.1073/pnas.1320965111.

Juretic, Nikoleta; Hoen, Douglas R.; Huynh, Michael L.; Harrison, Paul M.; Bureau, Thomas E. (2005): The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. In *Genome research* 15 (9), pp. 1292–1297. DOI: 10.1101/gr.4064205.

Juvvuna, Prasanna Kumar; Khandelia, Piyush; Lee, Li Ming; Makeyev, Eugene V. (2012): Argonaute identity defines the length of mature mammalian microRNAs. In *Nucleic Acids Res* 40 (14), pp. 6808–6820. DOI: 10.1093/nar/gks293.

Katz, Luba; Burge, Christopher B. (2003): Widespread selection for local RNA secondary structure in coding regions of bacterial genes. In *Genome research* 13 (9), pp. 2042–2051.

Kawaoka, Shinpei; Minami, Kosuke; Katsuma, Susumu; Mita, Kazuei; Shimada, Toru (2008): Developmentally synchronized expression of two Bombyx mori Piwi subfamily genes, SIWI and BmAGO3 in germ-line cells. In *Biochemical and biophysical research communications* 367 (4), pp. 755–760. DOI: 10.1016/j.bbrc.2008.01.013.

Kennerdell, Jason R.; Carthew, Richard W. (1998): Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. In *Cell* 95 (7), pp. 1017–1026.

Ketting, René F.; Fischer, Sylvia E.J.; Bernstein, Emily; Sijen, Titia; Hannon, Gregory J.; Plasterk, Ronald H.A. (2001): Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. In *Genes Dev.* 15 (20), pp. 2654–2659. DOI: 10.1101/gad.927801.

Khan, Hameed; Smit, Arian; Boissinot, Stéphane (2006): Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. In *Genome research* 16 (1), pp. 78–87. DOI: 10.1101/gr.4001406.

Kim, V. Narry (2005): MicroRNA biogenesis: coordinated cropping and dicing. In *Nature Reviews Molecular Cell Biology* 6 (5), p. 376. DOI: 10.1038/nrm1644.

Kolaczkowski, Bryan; Hupalo, Daniel N.; Kern, Andrew D. (2011): Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. In *Molecular biology and evolution* 28 (2), pp. 1033–1042. DOI: 10.1093/molbev/msq284.

Kuramochi-Miyagawa, Satomi; Watanabe, Toshiaki; Gotoh, Kengo; Totoki, Yasushi; Toyoda, Atsushi; Ikawa, Masahito et al. (2008): DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. In *Genes & Development* 22 (7), pp. 908–917. DOI: 10.1101/gad.1640708.

Kurihara, Yukio; Watanabe, Yuichiro (2004): Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. In *Proceedings of the National Academy of Sciences* 101 (34), pp. 12753–12758. DOI: 10.1073/pnas.0403115101.

Kwak, Pieter Bas; Tomari, Yukihide (2012): The N domain of Argonaute drives duplex unwinding during RISC assembly. In *Nature structural & molecular biology* 19 (2), pp. 145–151. DOI: 10.1038/nsmb.2232.

Lambert, Marine; Benmoussa, Abderrahim; Provost, Patrick (2019): Small Non-Coding RNAs Derived From Eukaryotic Ribosomal RNA. In *Non-coding RNA* 5 (1). DOI: 10.3390/ncrna5010016.

Le Thomas, Adrien; Rogers, Alicia K.; Webster, Alexandre; Marinov, Georgi K.; Liao, Susan E.; Perkins, Edward M. et al. (2013): Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. In *Genes & Development* 27 (4), pp. 390–399. DOI: 10.1101/gad.209841.112.

Lecellier, Charles-Henri; Dunoyer, Patrice; Arar, Khalil; Lehmann-Che, Jacqueline; Eyquem, Stephanie; Himber, Christophe et al. (2005): A cellular microRNA mediates antiviral defense in human cells. In *Science (New York, N.Y.)* 308 (5721), pp. 557–560. DOI: 10.1126/science.1108784.

Lee, Yoontae; Ahn, Chiyoung; Han, Jinju; Choi, Hyounjeong; Kim, Jaekwang; Yim, Jeongbin et al. (2003): The nuclear RNase III Drosha initiates microRNA processing. In *Nature* 425 (6956), p. 415. DOI: 10.1038/nature01957.

Leuschner, Philipp J. F.; Ameres, Stefan L.; Kueng, Stephanie; Martinez, Javier (2006): Cleavage of the siRNA passenger strand during RISC assembly in human cells. In *EMBO reports* 7 (3), pp. 314–320. DOI: 10.1038/sj.embor.7400637.

Lewis, Samuel H.; Quarles, Kaycee A.; Yang, Yujing; Tanguy, Melanie; Frézal, Lise; Smith, Stephen A. et al. (2018): Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. In *Nature ecology & evolution* 2 (1), pp. 174–181. DOI: 10.1038/s41559-017-0403-4.

Lewis, Samuel H.; Salmela, Heli; Obbard, Darren J. (2016): Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. In *Genome biology and evolution* 8 (3), pp. 507–518. DOI: 10.1093/gbe/evw018.

Li, Hongwei; Li, Wan Xiang; Ding, Shou Wei (2002): Induction and suppression of RNA silencing by an animal virus. In *Science (New York, N.Y.)* 296 (5571), pp. 1319–1321. DOI: 10.1126/science.1070948.

Li, Junjie; Yang, Zhiyong; Yu, Bin; Liu, Jun; Chen, Xuemei (2005): Methylation Protects miRNAs and siRNAs from a 3′-End Uridylation Activity in Arabidopsis. In *Current Biology* 15 (16), pp. 1501–1507. DOI: 10.1016/j.cub.2005.07.029.

Li, Wan-Xiang; Li, Hongwei; Lu, Rui; Li, Feng; Dus, Monica; Atkinson, Peter et al. (2004): Interferon antagonist proteins of influenza and vaccinia viruses are suppressors of RNA silencing. In *Proceedings of the National Academy of Sciences* 101 (5), pp. 1350–1355. DOI: 10.1073/pnas.0308308100.

Li, Xin Zhiguo; Roy, Christian K.; Dong, Xianjun; Bolcun-Filas, Ewelina; Wang, Jie; Han, Bo W. et al. (2013): An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. In *Molecular cell* 50 (1), pp. 67–81. DOI: 10.1016/j.molcel.2013.02.016.

Lingel, Andreas; Simon, Bernd; Izaurralde, Elisa; Sattler, Michael (2005): The structure of the flock house virus B2 protein, a viral suppressor of RNA interference, shows a novel mode of double-stranded RNA recognition. In *EMBO reports* 6 (12), pp. 1149–1155. DOI: 10.1038/sj.embor.7400583.

Liu, Jidong; Carmell, Michelle A.; Rivas, Fabiola V.; Marsden, Carolyn G.; Thomson, J. Michael; Song, Ji-Joon et al. (2004): Argonaute2 Is the Catalytic Engine of Mammalian RNAi. In *Science* 305 (5689), pp. 1437–1441. DOI: 10.1126/science.1102513.

Lohmann, Jan U.; Endl, Ingrid; Bosch, Thomas C. G. (1999): Silencing of developmental genes in Hydra. In *Developmental biology* 214 (1), pp. 211–214.

Lu, Shanfa; Sun, Ying-Hsuan; Chiang, Vincent L. (2009): Adenylation of plant miRNAs. In *Nucleic Acids Res* 37 (6), pp. 1878–1885. DOI: 10.1093/nar/gkp031.

Lu, Yilu; Li, Chao; Zhang, Kun; Sun, Huaqin; Tao, Dachang; Liu, Yunqiang et al. (2010): Identification of piRNAs in Hela cells by massive parallel sequencing. In *BMB reports* 43 (9), pp. 635–641. DOI: 10.5483/BMBRep.2010.43.9.635.

Malone, Colin D.; Brennecke, Julius; Dus, Monica; Stark, Alexander; McCombie, W. Richard; Sachidanandam, Ravi; Hannon, Gregory J. (2009): Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. In *Cell* 137 (3), pp. 522–535. DOI: 10.1016/j.cell.2009.03.040.

Matera, A. Gregory; Terns, Rebecca M.; Terns, Michael P. (2007): Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. In *Nature Reviews Molecular Cell Biology* 8 (3), p. 209. DOI: 10.1038/nrm2124.

Mette, M. F.; Aufsatz, W.; van der Winden, J.; Matzke, M. A.; Matzke, A. J. (2000): Transcriptional silencing and promoter methylation triggered by double-stranded RNA. In *The EMBO journal* 19 (19), pp. 5194–5201. DOI: 10.1093/emboj/19.19.5194.

Mi, Shijun; Cai, Tao; Hu, Yugang; Chen, Yemiao; Hodges, Emily; Ni, Fangrui et al. (2008): Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. In *Cell* 133 (1), pp. 116–127. DOI: 10.1016/j.cell.2008.02.034.

Miesen, Pascal; Girardi, Erika; van Rij, Ronald P. (2015): Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. In *Nucleic acids research* 43 (13), pp. 6545–6556. DOI: 10.1093/nar/gkv590.

Millar, Anthony A.; Waterhouse, Peter M. (2005): Plant and animal microRNAs: similarities and differences. In *Functional & Integrative Genomics* 5 (3), pp. 129–135. DOI: 10.1007/s10142-005-0145-2.

Mohn, Fabio; Handler, Dominik; Brennecke, Julius (2015): Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. In *Science (New York, N.Y.)* 348 (6236), pp. 812–817. DOI: 10.1126/science.aaa1039.

Mohn, Fabio; Sienski, Grzegorz; Handler, Dominik; Brennecke, Julius (2014): The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. In *Cell* 157 (6), pp. 1364–1379. DOI: 10.1016/j.cell.2014.04.031.

Montgomery, Taiowa A.; Howell, Miya D.; Cuperus, Josh T.; Li, Dawei; Hansen, Jesse E.; Alexander, Amanda L. et al. (2008): Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. In *Cell* 133 (1), pp. 128–141. DOI: 10.1016/j.cell.2008.02.033.

Moran, Yehu; Agron, Maayan; Praher, Daniela; Technau, Ulrich (2017): The evolutionary origin of plant and animal microRNAs. In *natecolevol* 1 (3), p. 27. DOI: 10.1038/s41559-016-0027.

Morazzani, Elaine M.; Wiley, Michael R.; Murreddu, Marta G.; Adelman, Zach N.; Myles, Kevin M. (2012): Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. In *PLOS Pathogens* 8 (1), e1002470. DOI: 10.1371/journal.ppat.1002470.

Morin, Ryan D.; O'Connor, Michael D.; Griffith, Malachi; Kuchenbauer, Florian; Delaney, Allen; Prabhu, Anna-Liisa et al. (2008): Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. In *Genome Res.* 18 (4), pp. 610–621. DOI: 10.1101/gr.7179508.

Nandi, Sayan; Chandramohan, Dhruva; Fioriti, Luana; Melnick, Ari M.; Hébert, Jean M.; Mason, Christopher E. et al. (2016): Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. In *Proceedings of the National Academy of Sciences of the United States of America* 113 (45), pp. 12697–12702. DOI: 10.1073/pnas.1609287113.

Napoli, C.; Lemieux, C.; Jorgensen, R. (1990): Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. In *The Plant cell* 2 (4), pp. 279–289. DOI: 10.1105/tpc.2.4.279.

Neeb, Zachary T.; Nowacki, Mariusz (2018): RNA-mediated transgenerational inheritance in ciliates and plants. In *Chromosoma* 127 (1), pp. 19–27. DOI: 10.1007/s00412-017-0655-4.

Neilsen, Corine T.; Goodall, Gregory J.; Bracken, Cameron P. (2012): IsomiRs – the overlooked repertoire in the dynamic microRNAome. In *Trends in Genetics* 28 (11), pp. 544–549. DOI: 10.1016/j.tig.2012.07.005.

Nishimasu, Hiroshi; Ishizu, Hirotsugu; Saito, Kuniaki; Fukuhara, Satoshi; Kamatani, Miharu K.; Bonnefond, Luc et al. (2012): Structure and function of Zucchini endoribonuclease in piRNA biogenesis. In *Nature* 491 (7423), pp. 284–287. DOI: 10.1038/nature11509.

Obbard, Darren J.; Gordon, Karl H. J.; Buck, Amy H.; Jiggins, Francis M. (2009): The evolution of RNAi as a defence against viruses and transposable elements. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364 (1513), pp. 99–115. DOI: 10.1098/rstb.2008.0168.

Obbard, Darren J.; Jiggins, Francis M.; Bradshaw, Nicholas J.; Little, Tom J. (2011): Recent and recurrent selective sweeps of the antiviral RNAi gene Argonaute-2 in three species of Drosophila. In *Molecular biology and evolution* 28 (2), pp. 1043–1056. DOI: 10.1093/molbev/msq280.

Okamura, Katsutomo; Hagen, Joshua W.; Duan, Hong; Tyler, David M.; Lai, Eric C. (2007): The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. In *Cell* 130 (1), pp. 89–100. DOI: 10.1016/j.cell.2007.06.028.

Olsen, Philip H.; Ambros, Victor (1999): The lin-4 Regulatory RNA Controls Developmental Timing in Caenorhabditis elegans by Blocking LIN-14 Protein Synthesis after the Initiation of Translation. In *Developmental biology* 216 (2), pp. 671–680. DOI: 10.1006/dbio.1999.9523.

Otsuka, Motoyuki; Jing, Qing; Georgel, Philippe; New, Liguo; Chen, Jianming; Mols, Johann et al. (2007): Hypersusceptibility to Vesicular Stomatitis Virus Infection in Dicer1-Deficient Mice Is Due to Impaired miR24 and miR93 Expression. In *Immunity* 27 (1), pp. 123–134. DOI: 10.1016/j.immuni.2007.05.014.

Palakodeti, Dasaradhi; Smielewska, Magda; Lu, Yi-Chien; Yeo, Gene W.; Graveley, Brenton R. (2008): The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. In *RNA (New York, N.Y.)* 14 (6), pp. 1174–1186. DOI: 10.1261/rna.1085008.

Pal-Bhadra, Manika; Leibovitch, Boris A.; Gandhi, Sumit G.; Chikka, Madhusudana Rao; Rao, Madhusudana; Bhadra, Utpal et al. (2004): Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. In *Science (New York, N.Y.)* 303 (5658), pp. 669–672. DOI: 10.1126/science.1092653.

Perrat, Paola N.; DasGupta, Shamik; Wang, Jie; Theurkauf, William; Weng, Zhiping; Rosbash, Michael; Waddell, Scott (2013): Transposition-driven genomic heterogeneity in the Drosophila brain. In *Science (New York, N.Y.)* 340 (6128), pp. 91–95. DOI: 10.1126/science.1231965.

Posner, Rachel; Toker, Itai Antoine; Antonova, Olga; Star, Ekaterina; Anava, Sarit; Azmon, Eran et al. (2019): Neuronal Small RNAs Control Behavior Transgenerationally. In *Cell* 177 (7), 1814-1826.e15. DOI: 10.1016/j.cell.2019.04.029.

Rajasethupathy, Priyamvada; Antonov, Igor; Sheridan, Robert; Frey, Sebastian; Sander, Chris; Tuschl, Thomas; Kandel, Eric R. (2012): A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. In *Cell* 149 (3), pp. 693–707. DOI: 10.1016/j.cell.2012.02.057.

Rechavi, Oded; Lev, Itamar (2017): Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans. In *Current biology : CB* 27 (14), R720-R730. DOI: 10.1016/j.cub.2017.05.043.

Robert, Valérie J. P.; Sijen, Titia; van Wolfswinkel, Josien; Plasterk, Ronald H. A. (2005): Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences. In *Genes & Development* 19 (7), pp. 782–787. DOI: 10.1101/gad.332305.

Rogers, Alicia K.; Situ, Kathy; Perkins, Edward M.; Toth, Katalin Fejes (2017): Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. In *Genes & Development* 31 (18), pp. 1858–1869. DOI: 10.1101/gad.303214.117.

Roovers, Elke F.; Rosenkranz, David; Mahdipour, Mahdi; Han, Chung-Ting; He, Nannan; Chuva de Sousa Lopes, Susana M. et al. (2015): Piwi proteins and piRNAs in mammalian oocytes and early embryos. In *Cell reports* 10 (12), pp. 2069–2082. DOI: 10.1016/j.celrep.2015.02.062.

Ross, Robert J.; Weiner, Molly M.; Lin, Haifan (2014): PIWI proteins and PIWI-interacting RNAs in the soma. In *Nature* 505 (7483), pp. 353–359. DOI: 10.1038/nature12987.

Rouget, Christel; Papin, Catherine; Boureux, Anthony; Meunier, Anne-Cécile; Franco, Bénédicte; Robine, Nicolas et al. (2010): Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. In *Nature* 467 (7319), pp. 1128–1132. DOI: 10.1038/nature09465.

Ruby, J. Graham; Jan, Calvin H.; Bartel, David P. (2007): Intronic microRNA precursors that bypass Drosha processing. In *Nature* 448 (7149), p. 83. DOI: 10.1038/nature05983.

Saito, Kuniaki; Sakaguchi, Yuriko; Suzuki, Takeo; Suzuki, Tsutomu; Siomi, Haruhiko; Siomi, Mikiko C. (2007): Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. In *Genes & Development* 21 (13), pp. 1603–1608. DOI: 10.1101/gad.1563607.

Sánchez Alvarado, A.; Newmark, P. A. (1999): Double-stranded RNA specifically disrupts gene expression during planarian regeneration. In *Proceedings of the National Academy of Sciences* 96 (9), pp. 5049–5054. DOI: 10.1073/pnas.96.9.5049.

Schnettler, Esther; Donald, Claire L.; Human, Stacey; Watson, Mick; Siu, Ricky W. C.; McFarlane, Melanie et al. (2013): Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. In *The Journal of general virology* 94 (Pt 7), pp. 1680–1689. DOI: 10.1099/vir.0.053850-0.

Seipel, Katja; Yanze, Nathalie; Schmid, Volker (2004): The germ line and somatic stem cell gene Cniwi in the jellyfish Podocoryne carnea. In *Int. J. Dev. Biol.* 48 (1), pp. 1–7. DOI: 10.1387/ijdb.15005568.

Seitz, Hervé; Ghildiyal, Megha; Zamore, Phillip D. (2008): Argonaute Loading Improves the 5′ Precision of Both MicroRNAs and Their miRNA∗ Strands in Flies. In *Current Biology* 18 (2), pp. 147–151. DOI: 10.1016/j.cub.2007.12.049.

Shabalina, Svetlana A.; Koonin, Eugene V. (2008): Origins and evolution of eukaryotic RNA interference. In *Trends in ecology & evolution* 23 (10), pp. 578–587. DOI: 10.1016/j.tree.2008.06.005.

Shen, En-Zhi; Chen, Hao; Ozturk, Ahmet R.; Tu, Shikui; Shirayama, Masaki; Tang, Wen et al. (2018): Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline. In *Cell* 172 (5), 937-951.e18. DOI: 10.1016/j.cell.2018.02.002.

Simpson, Alastair G.B.; Roger, Andrew J. (2004): The real 'kingdoms' of eukaryotes. In *Current Biology* 14 (17), R693-R696. DOI: 10.1016/j.cub.2004.08.038.

Starega-Roslan, Julia; Krol, Jacek; Koscianska, Edyta; Kozlowski, Piotr; Szlachcic, Wojciech J.; Sobczak, Krzysztof; Krzyzosiak, Wlodzimierz J. (2011): Structural basis of microRNA length variety. In *Nucleic Acids Res* 39 (1), pp. 257–268. DOI: 10.1093/nar/gkq727.

Stark, Alexander; Brennecke, Julius; Bushati, Natascha; Russell, Robert B.; Cohen, Stephen M. (2005): Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3'UTR Evolution. In *Cell* 123 (6), pp. 1133–1146. DOI: 10.1016/j.cell.2005.11.023.

Sugiyama, Tomoyasu; Cam, Hugh; Verdel, André; Moazed, Danesh; Grewal, Shiv I. S. (2005): RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. In *Proceedings of the National Academy of Sciences* 102 (1), pp. 152–157. DOI: 10.1073/pnas.0407641102.

Tabara, Hiroaki; Sarkissian, Madathia; Kelly, William G.; Fleenor, Jamie; Grishok, Alla; Timmons, Lisa et al. (1999): The rde-1 gene, RNA interference, and transposon silencing in C. elegans. In *Cell* 99 (2), pp. 123–132.

Takeda, Atsushi; Iwasaki, Shintaro; Watanabe, Toshiaki; Utsumi, Maki; Watanabe, Yuichiro (2008): The Mechanism Selecting the Guide Strand from Small RNA Duplexes is Different Among Argonaute Proteins. In *Plant Cell Physiol* 49 (4), pp. 493–500. DOI: 10.1093/pcp/pcn043.

Teixeira, Felipe Karam; Okuniewska, Martyna; Malone, Colin D.; Coux, Rémi-Xavier; Rio, Donald C.; Lehmann, Ruth (2017): piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. In *Nature* 552 (7684), pp. 268–272. DOI: 10.1038/nature25018.

Tosar, Juan Pablo; Rovira, Carlos; Cayota, Alfonso (2018): Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. In *Communications biology* 1, p. 2. DOI: 10.1038/s42003-017-0001-7.

Vagin, Vasily V.; Sigova, Alla; Li, Chengjian; Seitz, Hervé; Gvozdev, Vladimir; Zamore, Phillip D. (2006): A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. In *Science* 313 (5785), pp. 320–324. DOI: 10.1126/science.1129333.

van der Krol, A. R.; Mur, L. A.; Beld, M.; Mol, J. N.; Stuitje, A. R. (1990): Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. In *The Plant cell* 2 (4), pp. 291–299. DOI: 10.1105/tpc.2.4.291.

Vazquez, Franck; Vaucheret, Hervé; Rajagopalan, Ramya; Lepers, Christelle; Gasciolli, Virginie; Mallory, Allison C. et al. (2004): Endogenous trans-Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. In *Molecular cell* 16 (1), pp. 69–79. DOI: 10.1016/j.molcel.2004.09.028.

Verdel, André; Jia, Songtao; Gerber, Scott; Sugiyama, Tomoyasu; Gygi, Steven; Grewal, Shiv I. S.; Moazed, Danesh (2004): RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. In *Science* 303 (5658), pp. 672–676. DOI: 10.1126/science.1093686.

Volpe, Thomas A.; Kidner, Catherine; Hall, Ira M.; Teng, Grace; Grewal, Shiv I. S.; Martienssen, Robert A. (2002): Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. In *Science* 297 (5588), pp. 1833–1837. DOI: 10.1126/science.1074973.

Vourekas, Anastassios; Zheng, Qi; Alexiou, Panagiotis; Maragkakis, Manolis; Kirino, Yohei; Gregory, Brian D.; Mourelatos, Zissimos (2012): Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. In *Nature structural & molecular biology* 19 (8), pp. 773–781. DOI: 10.1038/nsmb.2347.

Wang, D. Y.; Kumar, S.; Hedges, S. B. (1999): Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. In *Proceedings. Biological sciences* 266 (1415), pp. 163–171. DOI: 10.1098/rspb.1999.0617.

Wang, Xiaowei (2014): Composition of seed sequence is a major determinant of microRNA targeting patterns. In *Bioinformatics (Oxford, England)* 30 (10), pp. 1377–1383. DOI: 10.1093/bioinformatics/btu045.

Watanabe, Toshiaki; Chuma, Shinichiro; Yamamoto, Yasuhiro; Kuramochi-Miyagawa, Satomi; Totoki, Yasushi; Toyoda, Atsushi et al. (2011): MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. In *Developmental cell* 20 (3), pp. 364–375. DOI: 10.1016/j.devcel.2011.01.005.

Waterhouse, P. M.; Graham, M. W.; Wang, M. B. (1998): Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. In *Proceedings of the National Academy of Sciences* 95 (23), pp. 13959–13964. DOI: 10.1073/pnas.95.23.13959.

Waterhouse, Peter M.; Wang, Ming-Bo; Lough, Tony (2001): Gene silencing as an adaptive defence against viruses. In *Nature* 411 (6839), p. 834.

Wu, Haoquan; Neilson, Joel R.; Kumar, Priti; Manocha, Monika; Shankar, Premlata; Sharp, Phillip A.; Manjunath, N. (2007): miRNA Profiling of Naïve, Effector and Memory CD8 T Cells. In *PLoS ONE* 2 (10), e1020. DOI: 10.1371/journal.pone.0001020.

Yoda, Mayuko; Kawamata, Tomoko; Paroo, Zain; Ye, Xuecheng; Iwasaki, Shintaro; Liu, Qinghua; Tomari, Yukihide (2010): ATP-dependent human RISC assembly pathways. In *nsmb* 17 (1), pp. 17–23. DOI: 10.1038/nsmb.1733.

Yu, Bin; Chapman, Elisabeth J.; Yang, Zhiyong; Carrington, James C.; Chen, Xuemei (2006): Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in Arabidopsis. In *FEBS letters* 580 (13), pp. 3117–3120. DOI: 10.1016/j.febslet.2006.04.063.

Yu, Bin; Yang, Zhiyong; Li, Junjie; Minakhina, Svetlana; Yang, Maocheng; Padgett, Richard W. et al. (2005): Methylation as a Crucial Step in Plant microRNA Biogenesis. In *Science* 307 (5711), pp. 932–935. DOI: 10.1126/science.1107130.

Zamore, Phillip D.; Haley, Benjamin (2005): Ribo-gnome: the big world of small RNAs. In *Science (New York, N.Y.)* 309 (5740), pp. 1519–1524. DOI: 10.1126/science.1111444.

Zamore, Phillip D.; Tuschl, Thomas; Sharp, Phillip A.; Bartel, David P. (2000): RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. In *Cell* 101 (1), pp. 25–33.

Zhang, Peng; Kang, Jun-Yan; Gou, Lan-Tao; Wang, Jiajia; Xue, Yuanchao; Skogerboe, Geir et al. (2015): MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. In *Cell research* 25 (2), pp. 193–207. DOI: 10.1038/cr.2015.4.

Zhang, Xiuren; Yuan, Yu-Ren; Pei, Yi; Lin, Shih-Shun; Tuschl, Thomas; Patel, Dinshaw J.; Chua, Nam-Hai (2006): Cucumber mosaic virus-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. In *Genes & Development* 20 (23), pp. 3255–3268. DOI: 10.1101/gad.1495506.

Zilberman, Daniel; Cao, Xiaofeng; Jacobsen, Steven E. (2003): ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. In *Science (New York, N.Y.)* 299 (5607), pp. 716–719. DOI: 10.1126/science.1079695.

Zong, Jie; Yao, Xuan; Yin, Jinyuan; Zhang, Dabing; Ma, Hong (2009): Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. In *Gene* 447 (1), pp. 29–39. DOI: 10.1016/j.gene.2009.07.004.

1. tRNA fragments in health and disease

Julia Jehn[1], David Rosenkranz[1,2]

[1]Institute of Organismic and Molecular Evolution, Anthropology, Johannes Gutenberg University Mainz, Anselm-Franz-von-Bentzel-Weg 7, 55099 Mainz, Germany

[2]Senckenberg Zentrum für Humangenetik, Weismüllerstraße 50, 60314 Frankfurt am Main, Germany

This chapter was published as a Review in Med One under the title "tRNA-Derived Small RNAs: The Good, the Bad and the Ugly" (Jehn and Rosenkranz, Med One 2019 4:e190015). Numbering of figures and tables as well as the citation style have been adjusted.

1.1. Abstract

Due to the abundance and conserved role of tRNAs, fragments thereof were considered as mere degradation products for a long time. Lately, however, it was unveiled that these 15–35 nucleotides-long tRNA (-precursor) -derived small RNAs (tsRNAs) can modulate gene expression by different mechanisms and act in a variety of contexts. While some tsRNAs inhibit translation globally by impeding the formation of the translation initiation complex, many studies find tsRNAs to silence target genes in a sequence-specific manner that is potentially mediated by Argonaute proteins. This function plays a role in transposon control, but was also found to be exploited by viruses and trypanosoma to regulate host genes. Beyond their involvement in infectious disease, aberrant tsRNA expression is linked to several other diseases such as cancer or neurological disorders. Furthermore, it was recently shown that tsRNAs residing in sperm of high-fat or low-protein diet mice can act as transgenerational transmitters that induce metabolic disorders and addictive behavior in the offspring. A better understanding of tsRNA-mediated gene regulation pathways, will not only expand our knowledge on how parental lifestyle influences the epigenome of the progeny, but may also enable the development of new drugs and biomarkers.

1.2. Introduction

1.2.1. Small non-coding RNAs

A major advance in understanding gene expression was the discovery of regulatory non-coding RNAs (ncRNAs). ncRNAs are transcripts that, unlike messenger RNA (mRNAs), do not code for proteins. Apart from mutations within protein-coding genes, perturbations in ncRNA pathways can lead to severe disease.

The most abundant and conserved ncRNAs are ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). While rRNAs together with ribosomal proteins form the ribosome, the macromolecular machine where protein synthesis takes place, tRNAs translate the codon information of the mRNAs during the protein synthesis as tRNAs with the matching anticodon

carry the respective amino acid to the end of the growing protein chain. rRNAs and tRNAs thus constitute central factors in protein translation.

Other ncRNAs such as small ncRNAs (<200 nt) and long ncRNAs (>200 nt) add yet another layer to the control of gene expression, as they can enhance or repress either the transcription (transcriptional level) or the translation (post-transcriptional level) of genes by various mechanism. As their expression is often specific to certain tissues or developmental stages, their regulatory potential critically contributes to cell-specific gene expression. While long ncRNAs can modulate gene expression and protein modifications by acting as a scaffold to bring molecular interaction partners together or to sponge up either antisense mRNAs or other regulators of gene expression, small ncRNAs usually guide effector proteins to their targets. The three most prominent small ncRNA classes, the micro RNAs (miRNAs), the small-interfering RNAs (siRNAs) and the PIWI-interacting RNAs (piRNAs), are loaded onto members of the Argonaute family, which can be subdivided into the AGO and the PIWI clade. While the 26–32 nt long piRNAs interact with proteins of the PIWI clade, the 22–24 nt long miRNAs and 21–22 nt long siRNAs bind to proteins of the AGO clade. Together with the respective Argonaute protein, these small ncRNAs form the heart of the RNA-induced silencing complex (RISC). Depending on the AGO/PIWI paralogue involved, the RISC can act either at the transcriptional level, inducing DNA methylation and histone modifications that lead to changes in chromatin structure, or at the post-transcriptional level by slicing mRNA, inducing its decay via decapping and deadenylation or inhibiting translation.

miRNAs and siRNAs typically post-transcriptionally regulate transcripts that have a target site complementary to the small ncRNA. While siRNAs require almost perfect complementarity to the target mRNA, miRNAs allow more imperfect complementarity for target regulation, requiring near-perfect match only within the 7 nt seed region located in the 5’ portion of the miRNA (Hutvágner and Zamore 2002). Similar to this, the piRNA seed is located at position second to seventh of the 5’ end and effective targeting is established for complementary target sites permitting also GU wobble base pairing (Zhang et al. 2018a).

While their ability to regulate gene expression bases on similar principles, the biogenesis of small ncRNAs differs substantially. miRNAs and siRNAs are generated from RNA polymerase II transcribed double-stranded precursors, that get cut by Drosha in the nucleus and are further processed by Dicer in the cytoplasm. The resulting duplex is then taken up by an AGO protein which only keeps the main strand while releasing the so-called passenger strand (Ha and Kim 2014). piRNAs, however, derive from single-stranded RNA pol II transcripts that mainly originate from conserved genomic loci called piRNA clusters. In a 3’-directed phased manner these precursors are cut by Zucchini that generates the 5’ end of primary piRNAs (Mohn et al. 2015; Han et al. 2015). Pre-mature piRNAs with a bias for a uridine at the 5’ end are loaded on a PIWI protein and get first trimmed and then 2’O-methylated at the 3’ end (Hayashi et al. 2016; Saxe et al. 2013; Ding et al. 2017). In case of post-transcriptional silencing, the piRNA guides the PIWI protein to a target site of a transcript, which gets sliced with an offset of 10 nt starting from the 5’ end of the guiding piRNA. This induces the generation of secondary, target-derived piRNAs that are again loaded onto PIWI proteins and in turn trigger the production of

piRNAs that resemble primary piRNAs, as they can target and slice complementary piRNA precursors. This results in a self-sustaining so-called ping-pong-loop that, while silencing the target transcripts, generates more piRNAs at the same time (Czech et al. 2018).

Given their important roles in normal physiology, it is not surprising that the dysregulation of small ncRNA pathways is involved in a variety of diseases. The piRNA pathway was discovered, as its disruption led to a strong infertility phenotype in fruit flies (Lin and Spradling 1997). An active piRNA pathway ensures genomic integrity by silencing transposable elements (so-called selfish, jumping genes). miRNAs, on the other hand, were found to act as oncogenes or tumor suppressors and are potential biomarkers for specific cancer types (Lan et al. 2015). Pathologic miRNA expression also plays a role in diabetes and cardiovascular disease (Paul et al. 2018).

1.2.2. tRNA-derived small RNAs as emerging small ncRNAs

The recently established high-throughput sequencing techniques of small ncRNA transcriptomes did not only expand the knowledge on the three prominent small ncRNA classes, but also opened up the field to study small ncRNA classes whose functions are far less well understood. One of these emerging new players in the small ncRNA zoo are tRNA-derived small RNAs (tsRNAs). While miRNA, siRNA and piRNA pathways were intensively studied over the last decades, tsRNAs were considered as mere degradation products of the abundant tRNAs until recently. First indications, that tsRNAs have a functional role came from studies in *Escherichia coli*, where tRNA-halves generated upon bacteriophage T4 infection were found to orchestrate a molecular defense response (Levitz et al. 1990).

In this review, we want to give an overview of how tsRNAs are generated and how they regulate genes, highlighting similarities and differences to other small ncRNA pathways. As tsRNAs were very recently shown to be important players in transgenerational epigenetics of metabolic disorders and addiction behavior, we focus especially on their role in disease.

1.3. Biogenesis and classification

tRNA precursors are transcribed by RNA polymerase III. Still in the nucleus, the so-called leader and trailer sequences are cut off from the tRNA precursors by the endonucleases RNase P and RNase Z, respectively. The trailer sequence being liberated from the 3’ end of the precursor gives rise to a first class of tsRNAs: the tsRNA1 series (also called 3’ U tsRNAs; figure 1.1). tsRNAs of the tsRNA1 series are approximately 15–22 nt in size and have a poly-U-stretch at their 3’ end resembling the Pol-III transcription termination signal (Haussecker et al. 2010; Kumar et al. 2014). If present, intronic sequences are spliced. The mature tRNA sequence is then folded and receives several chemical base modifications. Finally, a CCA signature is attached to its 3’ end before it is exported to the cytoplasm. The structure of tRNAs is highly conserved. With its four arms (D loop, anticodon loop, variable region, TψC loop) and the acceptor stem, the secondary structure of a tRNA resembles a cloverleaf. These secondary

structure elements are winded up to an L-shaped tertiary structure. By yet not fully elucidated mechanisms, mature tRNAs can get sliced into tRNA halves (tRHs) and smaller tsRNAs (figure 1.1). Please note that within this review the umbrella term tsRNA includes tRNA halves as well as the shorter tRNA-derived small RNAs. Depending from which end of the tRNA the tsRNAs are originating, they are part of the 5′ or the 3′ tsRNA series. Most tRHs are generated by bisecting tRNAs near the anticodon. Smaller tsRNAs can be produced by slicing either mature tRNAs or tRHs. In individual cases of potentially misfolded tRNAs this is done by Dicer (Cole et al. 2009; Babiarz et al. 2008), but generally their production is Dicer-independent (Kuscu et al. 2018; Li et al. 2012). tsRNAs seem to have the small RNA typical 5′ terminal phosphate group and 3′ terminal hydroxyl group (Couvillion et al. 2010, p. 2742; Haussecker et al. 2010). As 5′ and 3′ tsRNAs derive from mature tRNAs, they likewise carry plenty of base modifications, lack introns and in case of the 3′ tsRNAs (also called 3′ CCA tsRNAs) have the post-transcriptionally added 3′ CCA sequence. Importantly, the pool of tsRNAs within a cell does not at all reflect the abundance of the corresponding paternal tRNAs (Kumar et al. 2014; Reifur et al. 2012). There is emerging evidence that isodecoder-tRNAs (tRNAs having the same anticodon sequence, but bearing sequence variabilities) are tissue-specifically expressed and that their abundance influences the tsRNA pool generated thereof (Torres et al. 2019, p. 8453). Additionally, the cleavage does not seem to occur randomly as certain cleavage sites are enriched for the respective parental tRNAs (Kumar et al. 2014).

Apart from tsRNAs that are expressed under physiological conditions, a different tRH type is produced under stress conditions such as hypoxia. These tRNA-derived, stress-induced small RNAs (tiRNAs; 31–40 nt long) are generated by Angiogenin in mammalian cells (RNY1 in yeast). Angiogenin is a RNase Type A ribonuclease, which under stress conditions preferably slices mature tRNAs within the anticodon loop (Yamasaki et al. 2009). This creates 5′ tiRNAs with a 3′ phosphate group and 3′ tiRNAs with an 5′ hydroxyl group (Rybak and Vallee 1988). Notably, standard small ncRNA library preparation protocols that rely on the ligation of 5′ and 3′ adapter molecules will necessarily fail to include 5′ tiRNAs, carrying a phosphate group at both ends, and 3′ tiRNAs, carrying a hydroxyl group at both ends. This might have considerably contributed to the fact that these tsRNAs remain under the radar of a special scientific attention.

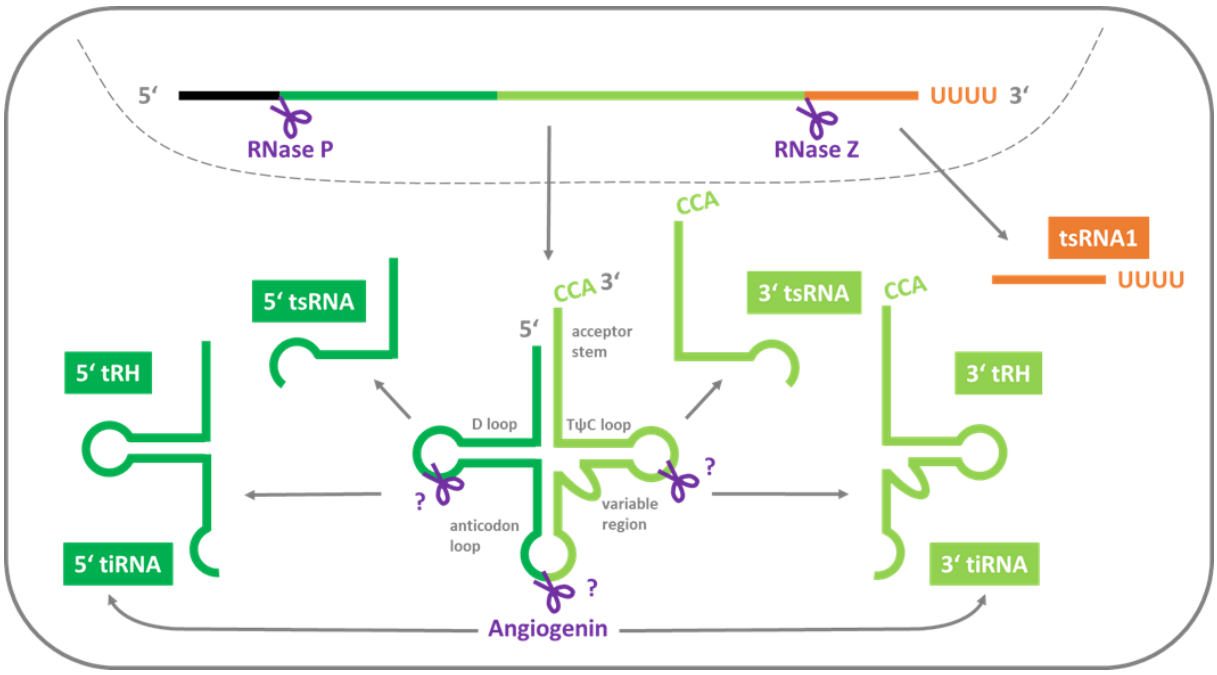


Figure 1.1: Biogenesis of tRNA-derived small RNAs. The trailer sequences being clipped off from the 3' end of the tRNA precursors by RNase Z are tsRNAs of the tsRNA1 series (orange). Additionally, RNase P cuts off the leader sequence of the 5' end of the tRNA precursor. The pre-tRNA is properly folded and extensively post-transcriptionally modified before it gets exported to the cytoplasm. Slicing of mature tRNAs by enzymes such as Angiogenin and Dicer produces tRNA halves and shorter tsRNAs. Depending on their derivation within the parental tRNA, these tsRNAs and tRhs are classified as 5' tsRNAs/tRHs (dark green) or 3' tsRNAs/tRHs (light green).

Most tsRNAs are generated from tRNAs that are encoded in the nucleus. However, tsRNAs can also derive from tRNAs that are encoded by organelles such as mitochondria and chloroplasts (Hirose et al. 2015; Telonis et al. 2015; Cognat et al. 2017). While in species like *Triops cancriformis*, nuclear-derived tsRNAs are about 400 times more prevalent than mitochondrial-derived tsRNAs (confirmed by Northern Blot analysis) (Hirose et al. 2015), organellar tsRNAs make up to one quarter of the tsRNA population in *Arabidopsis thaliana* and are exclusively located in the cytoplasm (confirmed by Northern Blot analysis) (Cognat et al. 2017). In case of *Triops cancriformis*, the pool of mitochondrial-derived tsRNAs is more diverse in terms of the region of origin within the parental tRNA, while the nuclear-derived tsRNA pool predominantly consists of 5' tRHs (Hirose et al. 2015). Thus, organellar tsRNAs might regulate other biological processes than tsRNAs derived from nuclear encoded tRNAs.

1.4. Gene regulatory function

Indications that tsRNAs are generated by defined cutting of tRNAs imply that these small ncRNAs might have a relevant function in the cells. As the major small ncRNA classes are known to regulate gene expression, it seems natural to assume that tsRNAs may also be involved in modulating gene expression. When studying gene regulation mechanisms, an important hint towards the mechanism of action is the subcellular localization of the involved

factors. Even though tsRNAs of the tsRNA1 series are generated in the nucleus, they seem to be widely distributed in the cytoplasm (confirmed by Northern Blot analysis) (Liao et al. 2010; Haussecker et al. 2010). This suggests that they play a role in post-transcriptional gene regulation. 5′ and 3′ tsRNAs are also predominantly localized in the cytoplasm (confirmed by Northern Blot analysis) (Haussecker et al. 2010; Kumar et al. 2014). Contrasting this, 5′ tsRNAs were found to be enriched in the nucleoli deep sequencing data of HeLa cells, suggesting a role in transcriptional regulation, after being reimported to the nucleus (Kumar et al. 2014). In general, the subcellular distribution of the different tsRNA series is not sufficiently supported by data and requires further investigation.

tsRNAs of all three series could be identified in all domains of life (Levitz et al. 1990; Gebetsberger et al. 2012; Couvillion et al. 2012; Garcia-Silva et al. 2010; Hsieh et al. 2010; Kumar et al. 2014; Kumar et al. 2015). It is therefore reasonable to assume that they have similar and conserved modes of action in the different organisms. Studies exploring the gene regulatory mechanism of tsRNAs commonly point to two mechanisms: while one cohort of studies describes a sequence-specific post-transcriptional silencing that closely resembles the miRNA pathway, other studies describe a tsRNA-induced global translation repression.

1.4.1. Global translation repression

Under stress conditions, cells globally shut down protein synthesis to quickly adapt their gene expression to the present situation. Part of that is induced by cleaving tRNAs with Angiogenin (Yamasaki et al. 2009). This not only decreases the amount of available tRNAs, which slows down translation, but more importantly generates tiRNAs that trigger a stress response program that inhibits the assembly of the translation initiation machinery at mRNAs. Ivanov *et al.* showed that stress-induced 5′ tiRNAs with a 5′ terminal oligoguanine (TOG; four to five guanine residues) motif such as 5′ tiRNA-Ala displace components of the protein complex eukaryotic initiation factor 4F (eIF4F) from the 5′ cap of mRNAs. This impedes the assembly of translation initiation complexes around mRNAs and in consequence blocks the translation of mRNAs (Ivanov et al. 2011). They could further show that displacing eIF4E and eIF4G is not driven by sequence specificity, but depends on the assembly of four 5′ TOG-containing tiRNAs to a RNA G-quadruplex (RG4) structure (figure 1.2A) that might recruit additional factors (Lyons et al. 2017). This RG4 structure and an intrinsic stem-loop structure of these 5′ tiRNAs are also required for the association with the cold shock domain of Y-Box Binding Protein 1 (YBX1), which induces the formation of stress granules in a phospho-eIF2α-independent manner (Emara et al. 2010; Ivanov et al. 2014). In these cytoplasmic foci, the mRNAs with attached stalled translation preinitiation complexes are transiently stored and can be released upon stress resolution to quickly reestablish homeostatic translation (Panas et al. 2016).

Later, Guzzi *et al.* discovered that 18 nt long TOG-containing tsRNAs, which they refer to as mini TOGs (mTOGs), are enriched in human embryonic stem cells (Guzzi et al. 2018). They could show that only mTOGs that are pseudouridylated at their eighth position (mediated by the Pseudouridine Synthase 7 (PUS7)) repress translation and that their loss impairs early

embryogenesis and hematopoietic stem cell differentiation. As they found pseudouridylated mTOGs to strongly interact with Polyadenylate Binding Protein 1 (PABPC1), a central translation initiation factor, while YBX1 predominantly binds to unmodified mTOGs, they hypothesize that pseudouridylation may lead to structural rearrangements that allows mTOGs to orchestrate different cell processes by binding to the respective protein partner.

Similar results were obtained by Krishna *et al.* (Krishna et al. 2019). They observed that a specific set of 5′ tRHs (30–35 nt) that is generated independently of Angiogenin is specifically enriched during retinoic acid induced differentiation of mouse embryonic stem cells (confirmed by Northern Blot analysis). As these 5′ tRHs were found to associate with individual subunits of polysomes as well as with completely assembled polysomes this suggests that 5′ tRHs are able to regulate translation at various stages. However, how this is accomplished remains an open question. Further, they could show that depending on the differentiation state these 5′ tRHs are interacting with distinct effector proteins and bind to different mRNA pools suggesting that these 5′ tRHs contribute to stem cell differentiation. As the differentiation-dependent 5′ tRHs mainly interact with the same protein, it is likely that they exert similar functions. Exemplarily for 5′ tRH-Gln-CTG, they provide a mechanistic explanation how these 5′ tRHs drive differentiation: by sequestering the ribosomal binding protein Igf2bp1 the mRNA of cMyc, a potent regulator of pluripotency, is no longer stabilized and gets degraded, which in turn perpetuates differentiation processes.

In case of the model archaeon *Haloferax volcanii*, protein synthesis is inhibited by binding of a 5′ tsRNA (tsRNA-Val; 26 nt) that is generated under alkaline stress conditions to the small subunit of the ribosome (confirmed by Northern Blot analysis) (Gebetsberger et al. 2012). The endonuclease that generates this tsRNA is not known. In a follow-up study Gebetsberger *et al.* introduced synthetic Val-tsRNAs in *S. cerevisiae* and *E. coli*, to show that the inhibitory potential of this tsRNA on mRNA translation is conserved in all three domains of life (Gebetsberger et al. 2017). They further showed that the 5′ tsRNA-Val specifically binds in close proximity to the mRNA channel of the 30S ribosomal subunit. This impedes the binding of this ribosomal subunit to mRNAs and thereby interrupts the formation of translation initiation complexes (figure 1.2C).

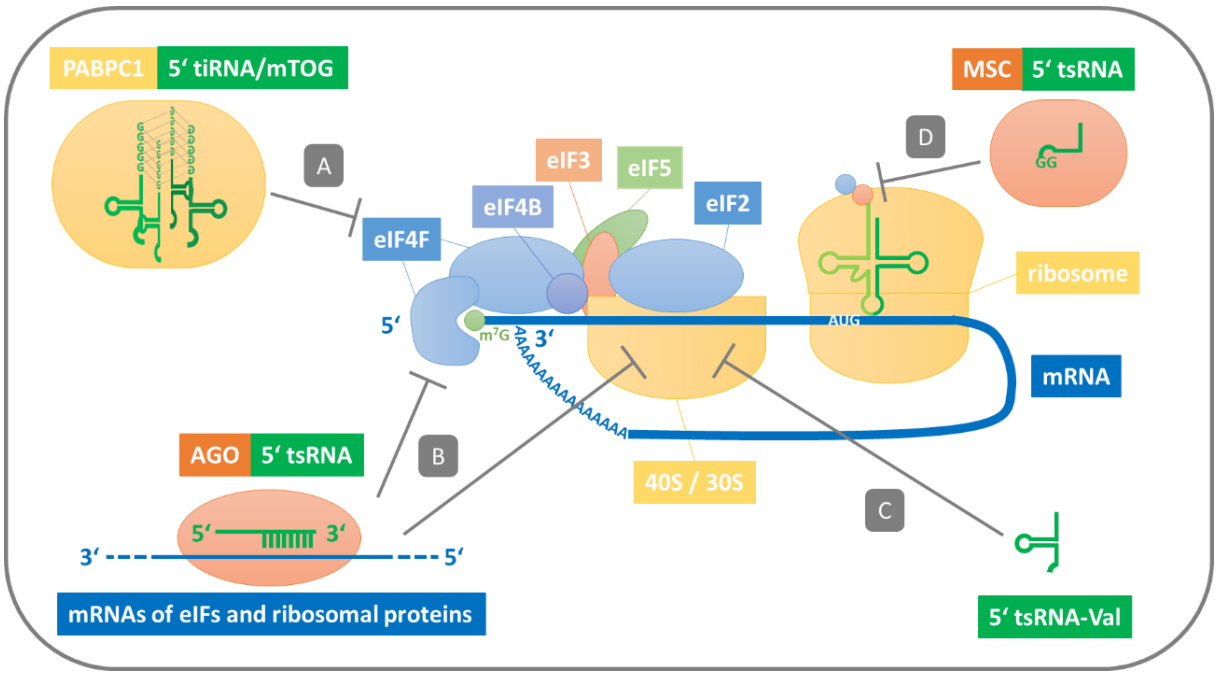


Figure 1.2: Mechanisms of tsRNA-mediated repression of translation initiation and polypeptide elongation. (A) 5′ tiRNAs with a 5′ TOG motif form a R4G structure that is required to displace the translation initiation complex eIF4F from the 5′ cap of mRNAs. In case of embryonic stem cells this is mediated by the protein PABPC1 that interacts with R4G structures of smaller mTOCs. (B) 5′ tsRNAs associated with Argonaute proteins sequence-specifically silence the mRNAs of translation initiation factors and ribosomal proteins. Target sites are distributed all over the respective mRNAs and efficient silencing is not restricted to a miRNA-like complementarity (see also figure 1.3B). (C) 5′ tsRNA-Val binds nearby the mRNA channel of the small ribosomal subunit (eukaryotic 40S and archaeal/bacterial 30S), which impedes binding to mRNA. (D) 5′ tsRNAs with a 3′ terminal GG motif block translation. At least 5′ tsRNA-Gln is primarily associated with the protein complex MSC that potentially inhibits polypeptide elongation.

With luciferase reporter assays in HeLa cells, Sobala and Hutvagner could show an inhibitory effect on translation for 5′ tsRNAs (e.g., tsRNA-Gln; 19 nt) that are generated independent of Angiogenin (figure 1.2D). Here, effective translation inhibition was dependent on a 3′ "GG" dinucleotide sequence in the tsRNA, while a sequence complementarity to the silenced mRNA was not required (Sobala and Hutvagner 2013). As 5′ tsRNAs were found to associate with actively elongating polysomes (confirmed by Northern Blot analysis) and overexpression of 5′ tsRNA-Gln did not reduce the number of reporter mRNAs being present in polysomes, the authors suggest that ribosome assembly is not impaired but the elongation of the peptide chain. By combining isotopic labeling SILAC mass spectrometry analysis with RNA immunoprecipitation, they later identified the Multisynthetase Complex (MSC) as major binding partner of 5′ tsRNA-Gln (Keam et al. 2017). As MSC is involved in the processivity of translation, binding of 5′ tsRNA-Gln might impede translation by blocking polypeptide elongation (Negrutskii and Deutscher 1991; Kaminska et al. 2009; Mirande 2010; David et al. 2011).

1.4.2. Sequence-specific gene regulation

A different mode of translation suppression was revealed by Luo *et al.* via mRNA sequencing and ribosome profiling of tsRNA mimic transfected S2 cells (Luo et al. 2018). Here, 5′ tsRNAs (20–22 nt) do not inhibit translation by displacing or blocking proteins of the translation machinery, but instead silence the mRNAs of key components of the translation machinery (e.g., ribosomal proteins and translation initiation factors) by sequence-specific binding (figure 1.2B and 3B). The 7 nt long target sites that were associated with a reduced translation efficiency were perfectly complementary to the respective 5′ tsRNAs and were conserved across different *Drosophila* genomes. The tsRNA-mediated target silencing was further found to be dependent on AGO2 and the fact that 5′ tsRNAs were co-immunoprecipitated with AGO2 suggests that tsRNAs inhibit specific targets via RNA interference (RNAi). However, unlike miRNAs, these 5′ tsRNAs bind equally effective all over the mRNA and a complementary 5′-portion of the tsRNA is not critical for the silencing effect. A miRNA-untypical gene targeting was also observed by Deng *et al.* in virus infected A549 cells, where the 3′-portion of a 5′ tRH (tRH-Glu-CTC; 31 nt) suppresses anti-viral target genes by sequence-specific binding to the 3′ UTR of their mRNAs (Deng et al. 2015). These findings contradict RNAi-typical Argonaute-induced silencing, as the structural and biophysical environment of the Argonaute MID-PIWI domain favors a tight binding of the 5′-portion of the small RNA, which is then used for target recognition (Boland et al. 2011).

On the other hand, several studies showed that tsRNAs from different series (18 nt 3′ tsRNA, 5′ and 3′ tsRNAs, 22 nt 3′ tsRNA) not only associate with the different Argonaute paralogs (confirmed by Northern Blot analysis in (Kuscu et al. 2018)), but also silence mRNAs via sequence complementary matching of their 5′-portion to target sites within the 3′ UTR (figure 1.3A) (Maute et al. 2013; Kumar et al. 2014; Kuscu et al. 2018). As Kuscu *et al.* noted, 3′ tsRNA-mediated target silencing mainly depends on a miRNA-like 5′ seed, but sequence complementarity beyond the seed sequence is also required (Kuscu et al. 2018). By performing Renilla luciferase reporter assays in HCT116 cells, Haussecker *et al.* observed a sequence-dependent target inhibition for 5′ tsRNAs and 3′ tsRNAs, but not tsRNA1s (Haussecker et al. 2010). In line with this, Kumar *et al.* found 5′ and 3′ tsRNAs, but not tsRNA1s, to associate with the four human AGO paralogs when analyzing PAR-CLIP data from HEK293 cells (Kumar et al. 2014). While 5′ and 3′ tsRNAs had read counts in the range of miRNAs for AGO1, AGO3 and AGO4, there were almost no read counts for AGO2.

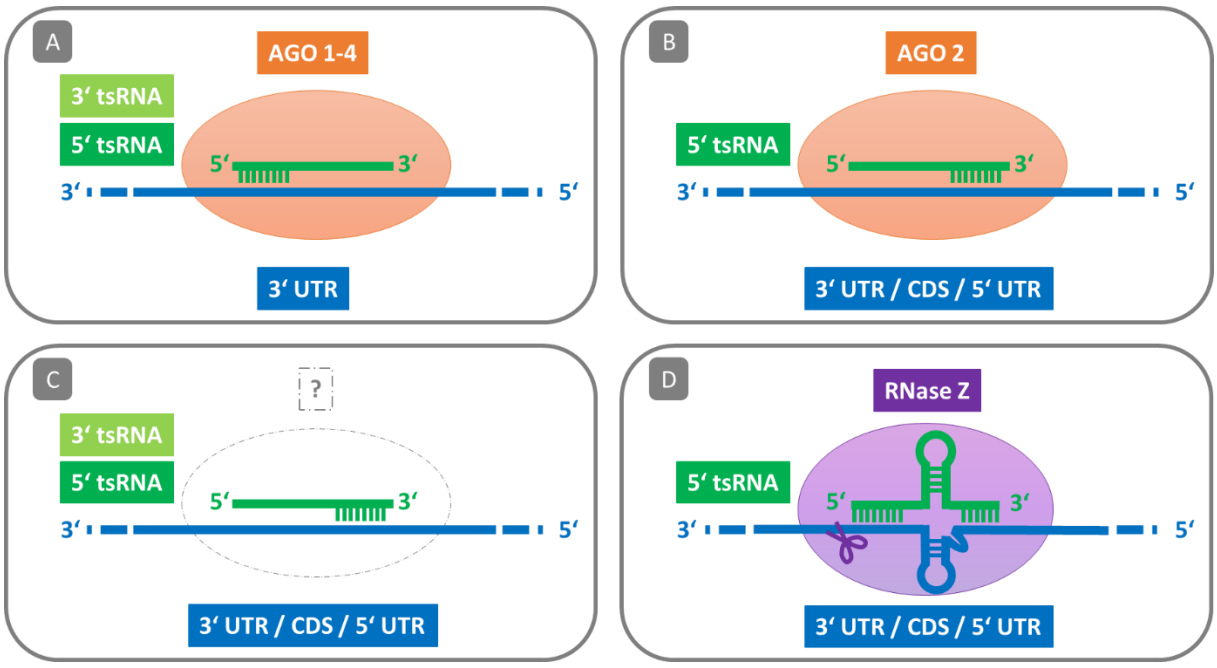


Figure 1.3: Mechanisms of tsRNA-mediated sequence-specific gene silencing. (A) Typical AGO-mediated silencing: 3’ tsRNAs and 5’tsRNAs, but not tsRNA1s, were found to associate with all four human AGO paralogs. The tsRNA-AGO complexes silence mRNAs that have a target site within the 3’ UTR that is complementary to the 5’ portion of the tsRNA. (B) Untypical AGO2-mediated silencing: 5’ tsRNAs associate with AGO2. Target sites are complementary to various portions of the tsRNA (e.g., the 3’ portion) and distributed all over the mRNA. (C) tsRNAs interact with yet unknown effector proteins or act independent of effectors to silence target mRNAs. Target recognition may be miRNA-untypical. (D) RNase Z mediated silencing: 5’ tRHs can form hybrids with mRNAs that structurally resemble pre-tRNAs. RNase Z recognizes such RNA hybrids and cleaves the mRNA.

While Argonaute-mediated RNAi is restricted to eukaryotic cells, tsRNAs are found in all domains of life. It was therefore speculated whether tsRNAs are part of an ancient regulatory pathway that was later complemented by the powerful RNAi pathways in the eukaryotic lineage (Keam and Hutvagner 2015). If tsRNA-mediated gene regulation is not a eukaryotic-specific trait, this would either require other tsRNA-interacting effector proteins, or regulatory mechanisms that do not rely on tsRNA-interacting protein partners (figure 1.3C). One potential tsRNA-directed modulator is the full-length form of the tRNA 3’ processing endoribonuclease tRNase Z, which cuts off the trailer sequence of pre-tRNAs. Due to its highly conserved function, it is expressed in all domains of life. As shown by Elbarbary *et al.*, RNase Z can cut RNA hybrids that have secondary structures like tRNA-precursors. The existence of such RNase Z-interacting small ncRNAs was confirmed by Northern Blot analysis. More importantly, they could show that 5’ tRHs can form such tRNA-precursor-like secondary structures with mRNAs that get recognized by cytoplasmic RNase Z, leading to the endonucleolytic cleavage and thus degradation of these mRNAs (figure 1.3D) (Elbarbary et al. 2009).

Besides these examples of tsRNA-mediated post-transcriptional gene repression, there is also a study by Kim *et al.* showing that binding of tsRNAs to mRNAs can also have the contrary effect of enhancing translation (Kim et al. 2017). They showed that a 3’ tsRNA (tsRNA-Leu-

CAG; 22 nt; confirmed by Northern Blot analysis) increases the translation of at least two ribosomal proteins by binding to sequence-complementary target sites that presumably lies within the coding sequence of their mRNA and hypothesize that this binding removes translation-inhibiting secondary structures.

In summary, present studies strongly indicate that tsRNAs can regulate mRNA targets by sequence-specific recognition. However, given the displayed objections, it is still unclear which the effector protein of this tsRNA-mediated gene regulation pathway is and if such an effector is needed in every case. Thus, more research is needed to pin down the exact mechanisms of target regulation.

1.5. Association with diseases and infections

1.5.1. tsRNAs as transgenerational epigenetic transmitters of metabolic disorders and addiction behavior

Increasing evidence is found that environmental stimuli such as an imbalanced diet or stress can influence the gene expression of next generations, which may foster the development of certain disease. This can be mediated by passing on parental small RNAs to the emerging offspring. So far strongest evidence for small RNA mediated transgenerational effects are found in nematodes, plants and ciliates (Rechavi and Lev 2017; Neeb and Nowacki 2018), but such phenomena could be observed also in mammals (Gapp et al. 2014; Grandjean et al. 2015). First detected for miRNAs, several studies could now show that tsRNAs may also act as transmitters of epigenetic information (figure 1.4A). Peng *et al.* first observed that mature mouse sperm is extremely enriched for 5' tRHs (29-34 nt) suggesting that these tsRNAs may play a role in early embryo development and transgenerational epigenetic inheritance (Peng et al. 2012). Later, Chen *et al.* could show that male mice subjected to a high-fat diet inherit a susceptibility for metabolic disorders, which is transmitted by 5' tRHs (30–34 nt) in their sperm (Chen et al. 2016). They could trace this back to 5' tRHs, as microinjection of the 30–40 nt fraction (predominantly 5' tRHs) from sperm of high-fat diet mice into normal-diet zygotes did phenocopy this effect in the F1 offspring. In a parallel-published study, Sharma *et al.* aimed to unveil the mechanism of how paternal low-protein diet alters the hepatic cholesterol biosynthesis in the offspring (Sharma et al. 2016). They could show that sperm of low-protein diet mice contains higher levels of 5' tRHs (28–34 nt) and indicate that these 5' tRHs are loaded via vesicles called epididymosomes into the maturing sperm (supported by Northern Blot analysis). Exemplarily for one of the low-protein diet-upregulated tRHs, they show that, in the early embryo, 5' tRH-Gly-GCC can downregulate the expression of ribosomal proteins as well as transcripts that are driven by the long terminal repeat retrotransposon MERVL. However, it remains unclear whether the repression of MERVL-driven transcripts or the slower growth kinetics due to the impaired ribosomal biogenesis leads to the observed metabolic changes in the offspring of low-protein diet mice.

Chen *et al.* additionally observed that injecting unmodified tsRNA into zygotes failed to phenocopy the high-fat diet-induced metabolic disturbances, while RNA modifications such as m^5C and m^2G are enriched in the sperm small RNA population of high-fat diet mice (Chen et al. 2016). In a follow-up study, they could show that the tRNA methyltransferase DNMT2 is upregulated in the caput epididymis under high-fat diet conditions (Zhang et al. 2018b). DNMT2 has been shown in previous studies to 5'-methylate the cytosine of certain tRNAs at position 38 (Goll et al. 2006; Tuorto et al. 2012). By generating *Dnmt2*-knockout mice, they could demonstrate that DNMT2 is required for intergenerational transmission of high-fat diet-induced metabolic disorders via sperm small RNAs (Zhang et al. 2018b). They showed that DNMT2 elevates the level of m^5C modifications in high-fat diet-induced sperm small RNAs and that it alters the composition of the sperm small RNA population (supported by Northern Blot analysis) stressing that RNA modifications are an important feature of small RNA mediated transgenerational effects.

Besides transgenerationally induced metabolic disorders, tsRNAs are also suspected to be a transmitter of inherited behavior. A study from Short *et al.* suggests that altered levels of miRNAs in mouse sperm, but also altered tsRNA levels (more 5' tRHs, less 3' tsRNAs), are related to inherited affective behaviors, which is reflected by a reduced anxiety and fear memory in the male offspring of physical active mice (Short et al. 2017). A more compelling evidence that sperm tsRNAs also induces behavioral phenotypes across generations was presented by Sarker *et al.* (Sarker et al. 2019). Here, they show that feeding pregnant female mice with a high-fat diet led to tsRNA-mediated disturbances of the fat metabolism and addictive-like behaviors in the paternal lineage for at least three generations. Again, microinjection of sperm tsRNAs from F1-males of high-fat diet mice into normal-diet zygotes did phenocopy the observed effects.

1.5.2. Host regulation by vesicle-transferred parasite tsRNAs

As suggested by Sharma *et al.*, tsRNAs are generated in the epididymis and shuttled via extracellular vesicles into maturing sperm (Sharma et al. 2016). Another form of vesicle-mediated tsRNA-transfer was observed in the infection process of the parasite *Trypanosoma cruzi* (*T. cruzi*). *T. cruzi* is a parasitic protozoan that has a life cycle with two main developmental stages (replicative or infective) each in an invertebrate and a vertebrate host (Jimenez 2014). Transmission usually occurs when an insect vector (typically members of the Triatominae) feeds on the blood of an infected mammal, thereby taking up the non-replicative trypomastigotes, which are circulating in the bloodstream. Inside insects, the trypomastigotes differentiate into epimastigotes that actively divide before they differentiate into non-replicative metacyclic trypomastigotes. During blood feeding on the next person or animal, the insects release the infective metacyclic trypomastigotes by defecation. Through the bite wound or mucous membranes the parasites can enter the vertebrate host, where they differentiate into replicating amastigotes. After several replication rounds the amastigotes differentiate into bloodstream trypomastigotes that can invade other tissues or can be taken

up by new blood feeding insects. Humans infected with this parasite develop the Chagas disease, which is endemic in most parts of Latin America (Moncayo and Silveira 2017).

Although *T. cruzi* lacks canonical Argonaute-mediated small RNA pathways (Ullu et al. 2004), it can apparently post-transcriptionally regulate gene expression (Clayton and Shapira 2007). It was therefore asked whether alternative small RNA pathways have evolved in *T. cruzi* in order to orchestrate the differentiation processes and adjust gene expression to the changing environment (Garcia-Silva et al. 2010). Thus, several studies analyzed the small RNA transcriptome of *T. cruzi* and found distinct tsRNA proportions and types in the different developmental forms (confirmed by Northern Blot analysis) (Reifur et al. 2012; Garcia-Silva et al. 2010). While tsRNAs (majorly 5' tRHs) make up only 26% of the small RNA population in the epimastigote form of *T. cruzi*, 63% of the small RNAs of the infective metacyclic form are tsRNAs (majorly 3' tRHs). The developmental stage-dependent tsRNAs not only differ in their abundance and tsRNA class, but also originate from a different set of parental tRNAs and have a distinct subcellular localization pattern (posterior cytoplasmic granules in the epimastigote, even cytoplasmic distribution in the metacyclic form). Later it was revealed that *T. cruzi* secretes extracellular vesicles that are highly loaded with tsRNAs and TcPIWI-tryp proteins (supported by fluorescence in situ hybridization assays) (Garcia-Silva et al. 2014b). TcPIWI-tryp is the only Argonaute protein reported in *T. cruzi* which was shown to bind tsRNAs (Garcia-Silva et al. 2014c). These extracellular vesicles can fuse with other parasites but can also invade mammalian cells. By incorporating extracellular vesicles of *T. cruzi* into HeLa cells, Garcia-Silva *et al.* could demonstrate that the tsRNA cargo is able to downregulate HeLa cell genes (Garcia-Silva et al. 2014a). As most differentially expressed host genes are related with structural cell components and immune responses pathways, this suggests that vesicle-transmitted tsRNAs play a major role in promoting the susceptibility of the mammalian host cells (figure 1.4B). A better understanding of this tsRNA-mediated host-parasite interaction may allow the development of effective anti-parasitic drugs. As trypanosomatids lack orthologues of the mammalian enzymes for tRNA half processing such as Angiogenin (Fricker et al. 2019), targeting Trypanosoma-specific tsRNA biogenesis factors might be a promising approach.

1.5.3. tsRNAs in of virus infections and retrotransposition

Another example of tsRNA-mediated host gene regulation was observed for the respiratory syncytial virus (RSV) (Deng et al. 2015). By studying the small RNA population of RSV-infected A549 cells, Wang *et al.* revealed that virus infection triggers predominantly the production of ~30 nt long 5' tRHs (confirmed by Northern Blot analysis) (Wang et al. 2013). Exemplarily for one of the RSV-induced tsRNAs (5' tsRNA-Glu-CTC), they could further show that this Angiogenin-dependent tsRNA is able to repress target mRNAs and promote RSV replication (figure 1.4B). In a follow-up study, Deng *et al.* were able to characterize the underlying molecular mechanism (Deng et al. 2015). With apolipoprotein E receptor 2 (APOER2) they could not only identify a target gene of 5' tRH-Glu-CTC, but could also show that this target is

an anti-viral protein. Thus, RSV exploits tsRNA-generation of the host to shut down an anti-viral mechanism. APOER2 has a target site for 5′ tRH-Glu-CTC in the 3′ UTR of its mRNA. However, as already indicated previously, they observed a miRNA-untypical target silencing. By performing luciferase assays of target site mutations, they demonstrated that the 3′ end of the 5′ tRH-Glu-CTC is crucial for target repression, while its 5′ end is only contributing to target recognition.

Using 5′ tRHs to modulate host genes might not be a unique mechanism of RSV, as Selitsky *et al.* observed increased levels of 5′ tRHs in liver samples of human adults that have chronical hepatitis B virus and hepatitis C virus infections (Selitsky et al. 2015). While this study did not investigate whether these 5′ tRHs have a similar role in virus progression, *in silico* analysis of 5′ tRHs that get produced by Angiogenin upon *Rickettsia conorii* infection (confirmed by Northern Blot analysis) suggest that endothelial factors of the host are targeted, which would facilitate this bacteria to overcome the hosts endothelial barrier (Gong et al. 2013).

While the mechanism described for RSV progression bases on a quite specific tsRNA-mediated repression of a host anti-viral protein, tsRNAs may have a more general role in the promotion of retroviruses and retrotransposons, since tRNAs serve as primers for their reverse transcription (figure 1.4C) (Marquet et al. 1995, p. 114). This may be the case especially for 3′ tsRNAs, as the primer binding site (PBS) is usually complementary to the 3′ end of the priming tRNA. Indeed, Ruggero *et al.* provided evidence that Human T-Cell Leukemia Virus Type 1 (HTLV-1) exploits host 3′ tsRNAs (18 nt, 3′ tsRNA-Pro) as primer for reverse transcription (Ruggero et al. 2014). When analyzing the small RNA population of $CD4^+$ T cells infected with HTLV-1, they observed that 3′ tsRNA-Pro was the most abundant tsRNA. Previously, its parental tRNA (tRNA-Pro) has been suggested to act as a primer for reverse transcription of HTLV-1 by binding via its 3′ portion to the PBS of HTLV-1 transcripts (Seiki et al. 1982). In a reverse transcription assay, Ruggero *et al.* could confirm that 3′ tsRNA-Pro can serve as a primer for HTLV-1 reverse transcriptase (Ruggero et al. 2014). They could further show that both the parental tRNA-Pro as well as the 3′ tsRNA-Pro are enriched in virus particles suggesting an important role for virus progression.

Contrasting conclusions were drawn by Yeung *et al.*, who found high levels of a 18 nt 3′ tsRNA (3′ tsRNA-Lys; confirmed by Northern Blot analysis) in human immunodeficiency virus type 1 (HIV-1) infected MT4 T-cells (Yeung et al. 2009). 3′ tsRNA-Lys is antisense to the PBS of the HIV-1. As they saw this 3′ tsRNA to be associated with AGO2, they suggested that 3′ tsRNA-Lys directs an RNAi-based defense mechanism of the host against HIV-1. Supporting this hypothesis, they observed a 3′ tsRNA-Lys-dose-dependent depletion of HIV-1 transcripts and further showed that HIV-1 replication is increased in cells where 3′ tsRNA-Lys is antagomir-inhibited.

Similar observations were made for long terminal repeat (LTR)-retrotransposons, for which Schorn *et al.* could show that 3′ tsRNAs play a role in inhibiting their retrotransposition in mouse stem cells (Schorn et al. 2017). Retrotransposons are “parasitic genes” within a genome that amplify themselves and jump to other genome loci utilizing a mechanism that involves

reverse transcription. Schorn *et al.* noted that 3′ tsRNAs are especially abundant in mouse stem cells that are depleted for histone H3 lysine 9 tri-methylation and therefore are impaired in transcriptional silencing of retrotransposons. Intriguingly, the 3′ tsRNAs that were found to be enriched are antisense to the 18 nt PBS of the most active mouse transposon families. Using transposition reporter assays, they could not only show that 3′ tsRNAs inhibit retrotransposition, but present two different mechanisms that depend on the size of the 3′ tsRNA. While 18 nt 3′ tsRNAs are only able to block the reverse transcription by binding to the PBS of LTR-retrotransposon transcripts, 22 nt 3′ tsRNAs post-transcriptionally silence the LTR-retrotransposons by reducing the retrotransposon transcript levels. The authors suggest that this effect is mediated by an RNAi-mechanism, as in a previous study 22 nt 3′ tsRNAs were shown to cleave target RNA in association with AGO2 (Li et al. 2012).

Instead of complementarity to the 3′ portion of tRNAs, few retrotransposons have a PBS that is complementary to an internal region of the primer tRNA. In case of the fruit fly retrotransposon *copia*, which is such an exceptions, it was suggested that the 39 nt 5′ tRH-Met, but not its parental tRNA, serves as a primer for reverse transcription (Marquet et al. 1995). Similar to the situation in retroviruses, 5′ tRHs may not only promote retrotransposon activity, but may also inhibit it. As described above, 5′ tRHs transmitted via sperm (confirmed by Northern Blot analysis) are able to downregulate transcripts that are driven by the LTR-retrotransposon MERVL in the developing mouse embryo (Sharma et al. 2016). Other indications come from plants, where 19 nt 5′ tsRNAs that are enriched in the pollen of *Arabidopsis* (confirmed by Northern Blot analysis) associate with AGO1 and cleave LTR-retrotransposons of the *gypsy* family (Martinez et al. 2017).

In conclusion, tsRNAs can be accounted to the pool of small ncRNAs that repress reactivated retrotransposons, thereby safeguarding the genome of germline cells and stem cells. While a tsRNA was also shown to inhibit virus progression by degrading viral transcripts, several studies found retroviruses to exploit tsRNAs to foster their own replication, or to downregulate the host's defense mechanisms. Targeting infection-promoting tsRNAs may therefore be a promising attempt to develop anti-viral treatment.

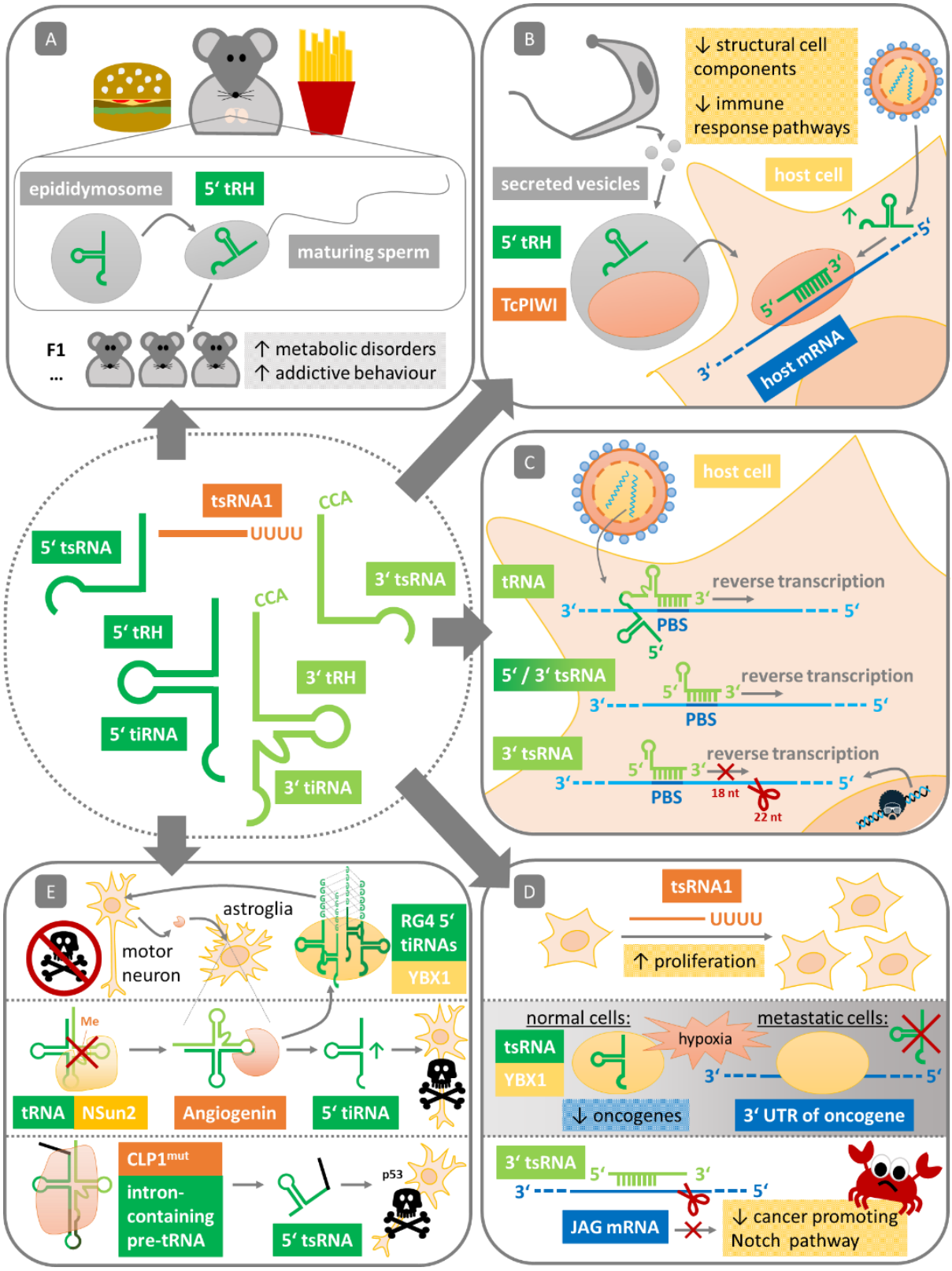


Figure 1.4: Association of tsRNAs with diseases. (A) 5′ tRHs as transgenerational transmitters inducing metabolic disorders and addictive behavior. 5′ tRHs are loaded into maturing sperm via epididymosomes. (B) tsRNAs repress host genes to facilitate infection. T. cruzi secretes vesicles containing tRHs and TcPIWI-tryp proteins that can fuse with host cells. RSV infection triggers the accumulation of host-generated 5′ tRHs. (C) Generally, the 3′ portion of tRNAs primes reverse transcription of retroviruses. Similarly, tsRNAs can act as primers, but can also block primer binding sites (PBS) of RNA from retroviruses and retrotransposons. (D) Proliferation-enhancing tsRNA1s can foster cancer progression (upper part), while e.g. specific 3′ tsRNAs act anti-tumorigenic. tsRNAs

generated under hypoxia sequester YBX1, which prevents stabilization of pro-oncogenic transcripts. As metastatic cells inhibit generation of these tsRNAs, YBX1 is able to stabilize tumor-promoting mRNAs (middle part). By degrading the mRNA of JAG, a 3′ tsRNA inactivates the cancer promoting Notch pathway (lower part). (E) Angiogenin secreted by stressed motor neurons is endocytosed by adjacent astroglia where tRNAs are cleaved to generate tiRNAs. 5′ tiRNAs forming R4G structures can associate with YBX1, which triggers stress granule assembly that contributes protecting motor neurons from undergoing apoptosis (upper part). Angiogenin-mediated tiRNA-generation can also promote apoptosis when the tRNA-methylase NSun2 is defective (middle part). Mutations in the Kinase CLP1 induce the accumulation of leader-sequence-containing 5′ tsRNAs, which trigger p53-dependent apoptosis of motor neurons (lower part).

1.5.4. tsRNAs in cancer progression

Initially, tsRNAs were deeper characterized and recognized as functional small ncRNAs in the context of cancer. In 2009 Lee *et al.* analyzed small RNA sequencing data of the human prostate cancer cell line HCT116 (Lee et al. 2009). Here, they found tsRNAs to be the second most abundant small ncRNA class following miRNAs. They deeper investigated a tsRNA1 (tsRNA1-Ser-TGA, 19 nt), which was more abundant in small RNA libraries of a variety of cancer cell lines compared to normal tissue libraries (confirmed by Northern Blot analysis). By siRNA-mediated knockdown and rescue co-transfection, they could show that tsRNA1-Ser-TGA is required for cell cycle propagation and proliferation of HCT116 cells (figure 1.4D, upper part). As luciferase reporter assays did not reveal any modulation effect, the authors suggest an RNAi-independent mechanism, which is in line with the above-mentioned PAR-Clip data analysis of Kumar *et al.* (Kumar et al. 2014).

So far, the best mechanistic explanation for a role of tsRNAs in cancer progression was found for a specific set of hypoxia-induced tsRNAs (figure 1.4D, middle part) (Goodarzi et al. 2015). As hypoxia is a major stress encountered by progressing cancerous cells, Goodarzi *et al.* analyzed the tsRNA levels in breast cancer cells under low oxygen conditions. While hypoxia induced the expression of a group of tsRNAs in breast cancer cells (MDA-231 cells) as well as in normal mammary epithelial cells, this specific set of tsRNAs was not upregulated in highly metastatic breast cancer cells (MDA-LM2 cells), indicating that these tsRNAs might have a tumor suppressive role. Interestingly, these particular tsRNAs (tsRNA-Glu, -Asp, -Gly, and -Tyr) share a sequence motif (“SCUBYC”) suggesting that they interact with a common *trans* factor. Indeed, by using a streptavidin-coupled tsRNA-Gly mimetic, they could identify the RNA-binding protein YBX1 as a specific interactor. As mentioned previously, YBX1 was found to induce stress granule formation when interacting with 5′ tiRNAs (Ivanov et al. 2011). Unlike the 5′ tiRNAs of the study conducted by Ivanov *et al.*, the tsRNAs here are not generated by a cleavage in the anticodon loop of the parental tRNA (Ivanov et al. 2011), but originate from a cleavage site within the D loop and cover at least the complete anticodon loop (Goodarzi et al. 2015). High-throughput sequencing of the YBX1-crosslinked small ncRNAs confirmed a specific interaction with this subset of tsRNAs. The fact that relatively low expressed tsRNAs (tsRNA-Glu and -Asp) were found to be enriched in this YBX1-crosslink sequencing data, but not highly expressed tsRNAs that are not part of this potentially tumor suppressive tsRNA set, suggests that the relative abundance of tsRNAs not necessarily represents an indicator for

functional activity. Goodarzi *et al.* could further show that this specific set of hypoxia-induced tsRNAs competes with pro-oncogenic transcripts for YBX1 binding. As YBX1 stabilizes pro-oncogenic transcripts by binding to their 3' UTR, displacement of YBX1 from these mRNAs leads to their degradation, which in turn inhibits tumor progression. In contrast, highly metastatic breast cancer cells evade this mechanism, as they are able to block the induction of these tumor suppressive tsRNAs. Here, YBX1 continues to stabilize pro-oncogenic transcripts, which is fostering metastasis of these breast cancer cells even more.

Another tumor suppressive tsRNA was found to be involved in colorectal cancer (Huang et al. 2017). Since this specific 17 nt 3' tsRNA can be derived from both, the 3' end of tRNA-Leu, as well as from the pre-miRNA-1280 (supported by Northern Blot analysis), 3' tsRNA-Leu is also listed as miR-1280 in popular databases. Several studies indicated a tumor suppressive activity of 3' tsRNA-Leu/miR-1280 in various cancer types such as melanoma, medulloblastoma, bladder cancer and thyroid carcinoma (Sun et al. 2015; Majid et al. 2012; Wang et al. 2015; Meng et al. 2016; Xu et al. 2015; Piepoli et al. 2012). In line with this, Huang *et al.* demonstrated that 3' tsRNA-Leu/miR-1280 inhibits the proliferation and colony formation of colorectal cancer cells (figure 1.4D, lower part) (Huang et al. 2017). They could mechanistically explain this as they identified Jagged Canonical Notch Ligand 2 (JAG2) as a direct target of 3' tsRNA-Leu/miR-1280. Upon binding of 3' tsRNA-Leu/miR-1280 to the 3' UTR of JAG2, the JAG2 transcript gets degraded, which in turn inactivates the Notch signaling pathway. Notch signaling is important for key functions of cancer progression such as proliferation, metastasis and the maintenance of cancer stem cell phenotypes. 3' tsRNA-Leu/miR-1280 can therefore be considered as a powerful tumor suppressive factor.

In contrast to that, a slightly bigger tsRNA originating from the 3' end of tRNA-Leu (tsRNA-Leu-CAG; 22 nt; confirmed by Northern Blot analysis) was shown to be essential for viability of the cancer cell lines HCT116 and HeLa (Kim et al. 2017). As inhibition of this 3' tsRNA in a hepatocellular carcinoma mouse model induced apoptosis of the tumorigenic cells, 3' tsRNA-Leu-CAG might be used as a potential drug target. The fact that two similar tsRNAs may induce completely converse functions is astonishing, but may be a prevalent feature of tsRNA-mediated processes in tumor progression. In line with this are the results of a study that compared the small RNA profiles of different prostate cancer states. While samples from organ-confined prostate cancer predominantly had 18 nt long tsRNAs, samples from metastatic lymph node prostate cancer expressed mainly 27 nt long tsRNAs (Martens-Uzunova et al. 2011). It is therefore likely that specific tsRNA populations play different roles in tumor progression.

Beside these reports on the implication of individual tsRNAs in cancer progression, tiRNAs generated by Angiogenin may play a more general role, as this ribonuclease is upregulated in almost all types of cancer (Tello-Montoliu et al. 2006). This assumption is supported by the finding, that the ribonucleolylitc activity of Angiogenin is required for angiogenesis and thus tumor formation (Shapiro and Vallee 1989; Kao et al. 2002). How exactly tiRNAs are involved in cancer progression is not clear yet.

1.5.5. tsRNAs in neurodegeneration

While its relevance for cancer progression is yet unclear, several studies could demonstrate that Angiogenin treatment is able to protect stressed motor neurons from degeneration (Kieran et al. 2008; Aparicio-Erriu and Prehn 2012; Thiyagarajan et al. 2012; Sebastià et al. 2009). Mutations of Angiogenin are strongly associated with amyotrophic lateral sclerosis (ALS) and Parkinson disease, where the degeneration of motor neurons is a major symptom (.van Es et al. 2011; Greenway et al. 2006). Notably, most of the mutations that are associated with ALS impair the RNase function of Angiogenin (Greenway et al. 2006; Crabtree et al. 2007). Thus, it was suggested that Angiogenin-generated tiRNAs play an important role in the survival of motor neurons. Skorupa *et al.* revealed that stressed motor neurons secrete Angiogenin, which is endocytosed by astroglia in close proximity (Skorupa et al. 2012). Astroglia that are deficient in Angiogenin under normal conditions then generate tiRNAs, which seem to mediate neuroprotection of motor neurons in paracrine. Ivanov *et al.* could later demonstrate that 5′ tiRNA-Ala and 5′ tiRNA-Cys forming RG4 structures via their 5′ TOG motif can enter motor neurons, where they interact with YBX1 to induce stress granule formation, which might be a central part of the neuroprotective response (figure 1.4E, upper part) (Ivanov et al. 2014). As the pathological GGGGCC repeats in the *C9ORF72* gene, which is the most common genetic cause of ALS (Mori et al. 2013), are able to form similar RG4 structures, Ivanov *et al.* suggest that these repeats interfere with the tiRNA-YBX1-mediated processes such as stress granule formation that in turn induces apoptosis (Ivanov et al. 2014).

Angiogenin-treatment was also shown to protect cortical neurons that are exposed to hyperosmotic stress (Saikia et al. 2014). Even though not validated for the neurons, the study by Saikia *et al.* provided a mechanism for cell survival of similarly stressed mouse embryonic fibroblasts. They showed that a set of 20 Angiogenin-dependent tiRNAs can form complexes with Cytochrome C (Cyt C) that gets released by the mitochondria in order to trigger the apoptosis cascade. The tiRNA-Cyt C interaction, however, blocks the formation of apoptosomes, which prevents that cells undergo apoptosis.

Contrasting effects were observed by Blanco *et al.*, where accumulation of Angiogenin-generated 5′ tiRNAs in NSun2-mutants leads to reduced cell size and increased apoptosis of cortical, hippocampal and striatal neurons (figure 1.4E, middle part) (Blanco et al. 2014). NSun2 is a cytosine-5 RNA methyltransferase, whose mutation can cause hereditary intellectual disability and the microcephaly causing Dubowitz-like syndrome (Abbasi-Moheb et al. 2012; Khan et al. 2012; Martinez et al. 2012). Blanco *et al.* could demonstrate that functional NSun2 methylates all tRNAs carrying a cytosine in their variable loop at position 48 or 49 and that this m^5C modification protects the tRNA to be cleaved by Angiogenin (Blanco et al. 2014). In NSun2-deficient cells, target tRNAs remain hypo-methylated and are cleaved by Angiogenin, which leads to an accumulation of 5′ tiRNAs. As 5′ tiRNAs are able to block translation globally (Ivanov et al. 2011), the lower translation rates may explain the reduction of neuronal cell size

and synapse formation in NSun2 knockout brains (Blanco et al. 2014). Importantly, these neurodegenerative phenotypes can be reverted by inhibiting Angiogenin.

Another report where tsRNAs are involved in degeneration of motor neurons links distortions in tRNA biogenesis to neurological defects in the peripheral and the central nervous system (Hanada et al. 2013; Karaca et al. 2014; Schaffer et al. 2014). Hanada *et al.* showed that mutations in the RNA kinase CLP1 leads to defects in the processing of intron-containing pre-tRNAs, which gives rise to untypical 5′ tsRNAs that contain the 5′ leader sequence (confirmed by Northern Blot analysis) (Hanada et al. 2013). Accumulation of these 5′ leader exon tsRNAs in motor neurons can trigger the activation of p53-dependent apoptosis (figure 1.4E, lower part). As shown by Schaffer *et al.* these neurodegeneration effects of mutant CLP1 is a conserved feature in vertebrate neurogenesis (Schaffer et al. 2014).

1.6. Perspectives, open questions, and considerations

Given the broad range of known tsRNA-associated diseases, it is likely that more diseases will be linked to aberrant tsRNA expression in the near future. To be able to develop tsRNA-based biomarkers or therapies, a deeper understanding of tsRNA pathways is required. Major remaining questions are:

- Which enzymes generate tsRNAs in which context?
- How are certain tsRNAs exported/imported to the nucleus?
- What are the protein interactors of tsRNAs involved in gene regulation?
- How does this work mechanistically?
- Which role do post-transcriptional modifications of mature tRNAs and tsRNAs play?

While answering these questions, future studies should consider the following aspects. First of all, in most of the reviewed studies conclusions are drawn from experiments that focus on individual tsRNAs. However, as we learn from the overall picture these studies give, different subsets of tsRNAs even from the same biogenesis-pathway may exert different functions in distinct contexts. It is therefore essential to investigate the mechanisms of different tsRNA subpopulations conjunctively within one experimental setting. Furthermore, most of the studies test the regulatory potential of tsRNAs using synthetic mimetics. As tsRNAs are heavily modified and it was shown that RNA modifications can be important features of small RNA function, the use of mimetics without the endogenous base modifications might lead to artifact results and wrong conclusions. It will be interesting to see whether mimetics including the natural base modifications will have, e.g., different affinities to certain protein interactors. In line with this, future studies should investigate differences in post-transcriptional modifications of tsRNAs and their parental tRNAs. This may not only explain why specific tsRNAs are enriched compared to their source tRNA, but also reveal why certain tsRNAs have different effects under distinct conditions. It is likely that aberrant tsRNA modifications are a major feature in the development and progression of disease. Moreover, future studies

should also investigate the role of organelle-derived tsRNAs. This might be especially relevant as mitochondrial-derived tsRNAs can be differentially expressed during development and as RNA modifications within mitochondrial tRNAs were shown to be associated with mitochondrial disease (Asano et al. 2018; Hirose et al. 2015). Additionally, library preparation protocols should be improved in order to obtain unbiased sequencing data. Currently, post-transcriptional base modifications as well as untypical 5' end 3' end modifications (e.g., tiRNAs) are factors that prevent tsRNAs from being faithfully represented in RNA sequence libraries and can cause sequencing arrests of longer transcripts that would then be erroneously interpreted as tsRNAs. Thus, the development of more sophisticated adapter ligation and reverse transcription protocols is required to obtain an unbiased picture of the real tsRNA population within biological samples. In addition, it advisable to verify small RNA-sequencing based expression data with Northern Blot analysis. Taking these considerations into account, research in this emerging field will open possibilities to develop therapies against certain viruses or parasites or to improve the diagnosis of severe disease such as cancers or neurodegenerative disorders.

1.7. Author Contributions

The manuscript was written by **JJ** and DR. Figures were designed by **JJ**.

1.8. Declarations and Acknowledgements

The authors declare that there is no conflict of interest.

This work was supported by the International PhD Programme (IPP) coordinated by the Institute of Molecular Biology IMB, Mainz, Germany, funded by the Boehringer Ingelheim Foundation.

We thank Daniel Gebert, Katharina Mauer, Sacha Heerschop, and Hans Zischler for helpful comments and discussion.

1.9. References

van Es, Michael A.; Schelhaas, Helenius J.; van Vught, Paul W. J.; Ticozzi, Nicola; Andersen, Peter M.; Groen, Ewout J. N. et al. (2011): Angiogenin variants in Parkinson disease and amyotrophic lateral sclerosis. In *Annals of neurology* 70 (6), pp. 964–973. DOI: 10.1002/ana.22611.

Abbasi-Moheb, Lia; Mertel, Sara; Gonsior, Melanie; Nouri-Vahid, Leyla; Kahrizi, Kimia; Cirak, Sebahattin et al. (2012): Mutations in NSUN2 cause autosomal-recessive intellectual disability. In *The American Journal of Human Genetics* 90 (5), pp. 847–855.

Aparicio-Erriu, Isabela M.; Prehn, Jochen H. M. (2012): Molecular Mechanisms in Amyotrophic Lateral Sclerosis: The Role of Angiogenin, a Secreted RNase. In *Frontiers in neuroscience* 6, p. 167. DOI: 10.3389/fnins.2012.00167.

Asano, Kana; Suzuki, Takeo; Saito, Ayaka; Wei, Fan-Yan; Ikeuchi, Yoshiho; Numata, Tomoyuki et al. (2018): Metabolic and chemical regulation of tRNA modification associated with taurine deficiency and human disease. In *Nucleic Acids Res* 46 (4), pp. 1565–1583. DOI: 10.1093/nar/gky068.

Babiarz, Joshua E.; Ruby, J. Graham; Wang, Yangming; Bartel, David P.; Blelloch, Robert (2008): Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. In *Genes & Development* 22 (20), pp. 2773–2785. DOI: 10.1101/gad.1705308.

Blanco, Sandra; Dietmann, Sabine; Flores, Joana V.; Hussain, Shobbir; Kutter, Claudia; Humphreys, Peter et al. (2014): Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. In *The EMBO journal* 33 (18), pp. 2020–2039. DOI: 10.15252/embj.201489282.

Boland, Andreas; Huntzinger, Eric; Schmidt, Steffen; Izaurralde, Elisa; Weichenrieder, Oliver (2011): Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (26), pp. 10466–10471. DOI: 10.1073/pnas.1103946108.

Chen, Qi; Yan, Menghong; Cao, Zhonghong; Li, Xin; Zhang, Yunfang; Shi, Junchao et al. (2016): Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. In *Science (New York, N.Y.)* 351 (6271), pp. 397–400. DOI: 10.1126/science.aad7977.

Clayton, Christine; Shapira, Michal (2007): Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. In *Molecular and biochemical parasitology* 156 (2), pp. 93–101. DOI: 10.1016/j.molbiopara.2007.07.007.

Cognat, Valérie; Morelle, Geoffrey; Megel, Cyrille; Lalande, Stéphanie; Molinier, Jean; Vincent, Timothée et al. (2017): The nuclear and organellar tRNA-derived RNA fragment population in Arabidopsis thaliana is highly dynamic. In *Nucleic acids research* 45 (6), pp. 3460–3472. DOI: 10.1093/nar/gkw1122.

Cole, Christian; Sobala, Andrew; Lu, Cheng; Thatcher, Shawn R.; Bowman, Andrew; Brown, John W. S. et al. (2009): Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. In *RNA (New York, N.Y.)* 15 (12), pp. 2147–2160. DOI: 10.1261/rna.1738409.

Couvillion, Mary T.; Bounova, Gergana; Purdom, Elizabeth; Speed, Terence P.; Collins, Kathleen (2012): A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. In *Molecular cell* 48 (4), pp. 509–520. DOI: 10.1016/j.molcel.2012.09.010.

Couvillion, Mary T.; Sachidanandam, Ravi; Collins, Kathleen (2010): A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. In *Genes & Development* 24 (24), pp. 2742–2747. DOI: 10.1101/gad.1996210.

Crabtree, Benedict; Thiyagarajan, Nethaji; Prior, Stephen H.; Wilson, Peter; Iyer, Shalini; Ferns, Tyrone et al. (2007): Characterization of human angiogenin variants implicated in amyotrophic lateral sclerosis. In *Biochemistry* 46 (42), pp. 11810–11818. DOI: 10.1021/bi701333h.

Czech, Benjamin; Munafò, Marzia; Ciabrelli, Filippo; Eastwood, Evelyn L.; Fabry, Martin H.; Kneuss, Emma; Hannon, Gregory J. (2018): piRNA-Guided Genome Defense: From Biogenesis to Silencing. In *Annual review of genetics* 52, pp. 131–157. DOI: 10.1146/annurev-genet-120417-031441.

David, Alexandre; Netzer, Nir; Strader, Michael Brad; Das, Suman R.; Chen, Cai Yun; Gibbs, James et al. (2011): RNA binding targets aminoacyl-tRNA synthetases to translating ribosomes. In *The Journal of biological chemistry* 286 (23), pp. 20688–20700. DOI: 10.1074/jbc.M110.209452.

Deng, Junfang; Ptashkin, Ryan N.; Chen, Yu; Cheng, Zhi; Liu, Guangliang; Phan, Thien et al. (2015): Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism. In *Molecular therapy : the journal of the American Society of Gene Therapy* 23 (10), pp. 1622–1629. DOI: 10.1038/mt.2015.124.

Ding, Deqiang; Liu, Jiali; Dong, Kunzhe; Midic, Uros; Hess, Rex A.; Xie, Huirong et al. (2017): PNLDC1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. In *Nature communications* 8 (1), p. 819. DOI: 10.1038/s41467-017-00854-4.

Elbarbary, Reyad A.; Takaku, Hiroaki; Uchiumi, Naoto; Tamiya, Hiroko; Abe, Mayumi; Takahashi, Masayuki et al. (2009): Modulation of gene expression by human cytosolic tRNase Z(L) through 5'-half-tRNA. In *PloS one* 4 (6), e5908. DOI: 10.1371/journal.pone.0005908.

Emara, Mohamed M.; Ivanov, Pavel; Hickman, Tyler; Dawra, Nemisha; Tisdale, Sarah; Kedersha, Nancy et al. (2010): Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. In *The Journal of biological chemistry* 285 (14), pp. 10959–10968. DOI: 10.1074/jbc.M109.077560.

Fricker, Roger; Brogli, Rebecca; Luidalepp, Hannes; Wyss, Leander; Fasnacht, Michel; Joss, Oliver et al. (2019): A tRNA half modulates translation as stress response in Trypanosoma brucei. In *Nature communications* 10 (1), p. 118. DOI: 10.1038/s41467-018-07949-6.

Gapp, Katharina; Jawaid, Ali; Sarkies, Peter; Bohacek, Johannes; Pelczar, Pawel; Prados, Julien et al. (2014): Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. In *Nature neuroscience* 17 (5), pp. 667–669. DOI: 10.1038/nn.3695.

Garcia-Silva, Maria R.; Cabrera-Cabrera, Florencia; das Neves, Roberta Ferreira Cura; Souto-Padrón, Thaís; Souza, Wanderley de; Cayota, Alfonso (2014a): Gene expression changes induced by Trypanosoma cruzi shed microvesicles in mammalian host cells: relevance of tRNA-derived halves. In *BioMed research international* 2014, p. 305239. DOI: 10.1155/2014/305239.

Garcia-Silva, Maria R.; das Neves, Roberta Ferreira Cura; Cabrera-Cabrera, Florencia; Sanguinetti, Julia; Medeiros, Lia C.; Robello, Carlos et al. (2014b): Extracellular vesicles shed by Trypanosoma cruzi are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells. In *Parasitology research* 113 (1), pp. 285–304. DOI: 10.1007/s00436-013-3655-1.

Garcia-Silva, Maria Rosa; Frugier, Magali; Tosar, Juan Pablo; Correa-Dominguez, Alejandro; Ronalte-Alves, Lysangela; Parodi-Talice, Adriana et al. (2010): A population of tRNA-derived small RNAs is actively produced in Trypanosoma cruzi and recruited to specific cytoplasmic granules. In *Molecular and biochemical parasitology* 171 (2), pp. 64–73. DOI: 10.1016/j.molbiopara.2010.02.003.

Garcia-Silva, Maria Rosa; Sanguinetti, Julia; Cabrera-Cabrera, Florencia; Franzén, Oscar; Cayota, Alfonso (2014c): A particular set of small non-coding RNAs is bound to the distinctive Argonaute protein of Trypanosoma cruzi: insights from RNA-interference deficient organisms. In *Gene* 538 (2), pp. 379–384. DOI: 10.1016/j.gene.2014.01.023.

Gebetsberger, Jennifer; Wyss, Leander; Mleczko, Anna M.; Reuther, Julia; Polacek, Norbert (2017): A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. In *RNA biology* 14 (10), pp. 1364–1373. DOI: 10.1080/15476286.2016.1257470.

Gebetsberger, Jennifer; Zywicki, Marek; Künzi, Andrea; Polacek, Norbert (2012): tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii. In *Archaea* 2012.

Goll, Mary Grace; Kirpekar, Finn; Maggert, Keith A.; Yoder, Jeffrey A.; Hsieh, Chih-Lin; Zhang, Xiaoyu et al. (2006): Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. In *Science (New York, N.Y.)* 311 (5759), pp. 395–398. DOI: 10.1126/science.1120976.

Gong, Bin; Lee, Yong Sun; Lee, Inhan; Shelite, Thomas R.; Kunkeaw, Nawapol; Xu, Guang et al. (2013): Compartmentalized, functional role of angiogenin during spotted fever group rickettsia-induced endothelial barrier dysfunction: evidence of possible mediation by host tRNA-derived small noncoding RNAs. In *BMC infectious diseases* 13, p. 285. DOI: 10.1186/1471-2334-13-285.

Goodarzi, Hani; Liu, Xuhang; Nguyen, Hoang C. B.; Zhang, Steven; Fish, Lisa; Tavazoie, Sohail F. (2015): Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. In *Cell* 161 (4), pp. 790–802. DOI: 10.1016/j.cell.2015.02.053.

Grandjean, Valérie; Fourré, Sandra; De Abreu, Diana Andrea Fernandes; Derieppe, Marie-Alix; Remy, Jean-Jacques; Rassoulzadegan, Minoo (2015): RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. In *Scientific reports* 5, 18193 EP -. DOI: 10.1038/srep18193.

Greenway, Matthew J.; Andersen, Peter M.; Russ, Carsten; Ennis, Sean; Cashman, Susan; Donaghy, Colette et al. (2006): ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. In *Nature genetics* 38 (4), p. 411. DOI: 10.1038/ng1742.

Guzzi, Nicola; Cieśla, Maciej; Ngoc, Phuong Cao Thi; Lang, Stefan; Arora, Sonali; Dimitriou, Marios et al. (2018): Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. In *Cell* 173 (5), 1204-1216.e26. DOI: 10.1016/j.cell.2018.03.008.

Ha, Minju; Kim, V. Narry (2014): Regulation of microRNA biogenesis. In *Nat Rev Mol Cell Biol* 15 (8), p. 509.

Han, Bo W.; Wang, Wei; Li, Chengjian; Weng, Zhiping; Zamore, Phillip D. (2015): piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. In *Science (New York, N.Y.)* 348 (6236), pp. 817–821. DOI: 10.1126/science.aaa1264.

Hanada, Toshikatsu; Weitzer, Stefan; Mair, Barbara; Bernreuther, Christian; Wainger, Brian J.; Ichida, Justin et al. (2013): CLP1 links tRNA metabolism to progressive motor-neuron loss. In *Nature* 495 (7442), p. 474.

Haussecker, Dirk; Huang, Yong; Lau, Ashley; Parameswaran, Poornima; Fire, Andrew Z.; Kay, Mark A. (2010): Human tRNA-derived small RNAs in the global regulation of RNA silencing. In *RNA (New York, N.Y.)* 16 (4), pp. 673–695. DOI: 10.1261/rna.2000810.

Hayashi, Rippei; Schnabl, Jakob; Handler, Dominik; Mohn, Fabio; Ameres, Stefan L.; Brennecke, Julius (2016): Genetic and mechanistic diversity of piRNA 3'-end formation. In *Nature* 539 (7630), pp. 588–592. DOI: 10.1038/nature20162.

Hirose, Yuka; Ikeda, Kahori T.; Noro, Emiko; Hiraoka, Kiriko; Tomita, Masaru; Kanai, Akio (2015): Precise mapping and dynamics of tRNA-derived fragments (tRFs) in the development of Triops cancriformis (tadpole shrimp). In *BMC genetics* 16 (1), p. 83. DOI: 10.1186/s12863-015-0245-5.

Hsieh, Li-Ching; Lin, Shu-I; Kuo, Hui-Fen; Chiou, Tzyy-Jen (2010): Abundance of tRNA-derived small RNAs in phosphate-starved Arabidopsis roots. In *Plant signaling & behavior* 5 (5), pp. 537–539. DOI: 10.4161/psb.11029.

Huang, Bingqing; Yang, Huipeng; Cheng, Xixi; Wang, Dan; Fu, Shuyu; Shen, Wencui et al. (2017): tRF/miR-1280 Suppresses Stem Cell-like Cells and Metastasis in Colorectal Cancer. In *Cancer research* 77 (12), pp. 3194–3206. DOI: 10.1158/0008-5472.CAN-16-3146.

Hutvágner, György; Zamore, Phillip D. (2002): A microRNA in a multiple-turnover RNAi enzyme complex. In *Science (New York, N.Y.)* 297 (5589), pp. 2056–2060. DOI: 10.1126/science.1073827.

Ivanov, Pavel; Emara, Mohamed M.; Villen, Judit; Gygi, Steven P.; Anderson, Paul (2011): Angiogenin-induced tRNA fragments inhibit translation initiation. In *Molecular cell* 43 (4), pp. 613–623. DOI: 10.1016/j.molcel.2011.06.022.

Ivanov, Pavel; O'Day, Elizabeth; Emara, Mohamed M.; Wagner, Gerhard; Lieberman, Judy; Anderson, Paul (2014): G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (51), pp. 18201–18206. DOI: 10.1073/pnas.1407361111.

Jimenez, Veronica (2014): Dealing with environmental challenges: mechanisms of adaptation in Trypanosoma cruzi. In *Research in microbiology* 165 (3), pp. 155–165.

Kaminska, Monika; Havrylenko, Svitlana; Decottignies, Paulette; Le Maréchal, Pierre; Negrutskii, Boris; Mirande, Marc (2009): Dynamic Organization of Aminoacyl-tRNA Synthetase Complexes in the Cytoplasm of Human Cells. In *The Journal of biological chemistry* 284 (20), pp. 13746–13754. DOI: 10.1074/jbc.M900480200.

Kao, Richard Y. T.; Jenkins, Jeremy L.; Olson, Karen A.; Key, Marc E.; Fett, James W.; Shapiro, Robert (2002): A small-molecule inhibitor of the ribonucleolytic activity of human angiogenin that possesses antitumor activity. In *Proceedings of the National Academy of Sciences* 99 (15), pp. 10066–10071. DOI: 10.1073/pnas.152342999.

Karaca, Ender; Weitzer, Stefan; Pehlivan, Davut; Shiraishi, Hiroshi; Gogakos, Tasos; Hanada, Toshikatsu et al. (2014): Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. In *Cell* 157 (3), pp. 636–650.

Keam, Simon P.; Hutvagner, Gyorgy (2015): tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. In *Life (Basel, Switzerland)* 5 (4), pp. 1638–1651. DOI: 10.3390/life5041638.

Keam, Simon P.; Sobala, Andrew; Have, Sara ten; Hutvagner, Gyorgy (2017): tRNA-Derived RNA Fragments Associate with Human Multisynthetase Complex (MSC) and Modulate Ribosomal Protein Translation. In *Journal of proteome research* 16 (2), pp. 413–420. DOI: 10.1021/acs.jproteome.6b00267.

Khan, Muzammil Ahmad; Rafiq, Muhammad Arshad; Noor, Abdul; Hussain, Shobbir; Flores, Joana V.; Rupp, Verena et al. (2012): Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. In *The American Journal of Human Genetics* 90 (5), pp. 856–863.

Kieran, Dairín; Sebastia, Jordi; Greenway, Matthew J.; King, Matthew A.; Connaughton, Dervla; Concannon, Caoimhin G. et al. (2008): Control of motoneuron survival by angiogenin. In *Journal of Neuroscience* 28 (52), pp. 14056–14061.

Kim, Hak Kyun; Fuchs, Gabriele; Wang, Shengchun; Wei, Wei; Zhang, Yue; Park, Hyesuk et al. (2017): A transfer-RNA-derived small RNA regulates ribosome biogenesis. In *Nature* 552 (7683), p. 57. DOI: 10.1038/nature25005.

Krishna, Srikar; Yim, Daniel Gr; Lakshmanan, Vairavan; Tirumalai, Varsha; Koh, Judice Ly; Park, Jung Eun et al. (2019): Dynamic expression of tRNA-derived small RNAs define cellular states. In *EMBO reports* 20 (7), e47789. DOI: 10.15252/embr.201947789.

Kumar, Pankaj; Anaya, Jordan; Mudunuri, Suresh B.; Dutta, Anindya (2014): Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. In *BMC Biol* 12 (1), pp. 12–78. DOI: 10.1186/PREACCEPT-5867533061403216.

Kumar, Pankaj; Mudunuri, Suresh B.; Anaya, Jordan; Dutta, Anindya (2015): tRFdb: a database for transfer RNA fragments. In *Nucleic acids research* 43 (Database issue), D141-5. DOI: 10.1093/nar/gku1138.

Kuscu, Canan; Kumar, Pankaj; Kiran, Manjari; Su, Zhangli; Malik, Asrar; Dutta, Anindya (2018): tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. In *RNA (New York, N.Y.)* 24 (8), pp. 1093–1105. DOI: 10.1261/rna.066126.118.

Lan, Huiyin; Lu, Haiqi; Wang, Xian; Jin, Hongchuan (2015): MicroRNAs as potential biomarkers in cancer: opportunities and challenges. In *BioMed research international* 2015, p. 125094. DOI: 10.1155/2015/125094.

Lee, Yong Sun; Shibata, Yoshiyuki; Malhotra, Ankit; Dutta, Anindya (2009): A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). In *Genes & Development* 23 (22), pp. 2639–2649. DOI: 10.1101/gad.1837609.

Levitz, Ruth; Chapman, Daphne; Amitsur, Michal; Green, Robert; Snyder, Larry; Kaufmann, Gabriel (1990): The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. In *The EMBO journal* 9 (5), pp. 1383–1389.

Li, Zhihua; Ender, Christine; Meister, Gunter; Moore, Patrick S.; Chang, Yuan; John, Bino (2012): Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. In *Nucleic acids research* 40 (14), pp. 6787–6799. DOI: 10.1093/nar/gks307.

Liao, Jian-You; Ma, Li-Ming; Guo, Yan-Hua; Zhang, Yu-Chan; Zhou, Hui; Shao, Peng et al. (2010): Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. In *PloS one* 5 (5), e10563. DOI: 10.1371/journal.pone.0010563.

Lin, Haifan; Spradling, Allan C. (1997): A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. In *Development (Cambridge, England)* 124 (12), pp. 2463–2476.

Luo, Shiqi; He, Feng; Luo, Junjie; Dou, Shengqian; Wang, Yirong; Guo, Annan; Lu, Jian (2018): Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response. In *Nucleic acids research* 46 (10), pp. 5250–5268. DOI: 10.1093/nar/gky189.

Lyons, Shawn M.; Gudanis, Dorota; Coyne, Steven M.; Gdaniec, Zofia; Ivanov, Pavel (2017): Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. In *Nature communications* 8 (1), p. 1127. Available online at https://www.nature.com/articles/s41467-017-01278-w.pdf.

Majid, Shahana; Dar, Altaf A.; Saini, Sharanjot; Shahryari, Varahram; Arora, Sumit; Zaman, Mohd Saif et al. (2012): MicroRNA-1280 Inhibits Invasion and Metastasis by Targeting ROCK1 in Bladder Cancer. In *PloS one* 7 (10). DOI: 10.1371/journal.pone.0046743.

Marquet, R.; Isel, C.; Ehresmann, C.; Ehresmann, B. (1995): tRNAs as primer of reverse transcriptases. In *Biochimie* 77 (1-2), pp. 113–124.

Martens-Uzunova, E. S.; Jalava, S. E.; Dits, N. F.; van Leenders, G J L H; Møller, S.; Trapman, J. et al. (2011): Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. In *Oncogene* 31, 978 EP -. DOI: 10.1038/onc.2011.304.

Martinez, Fernando Jose; Lee, Jeong Ho; Lee, Ji Eun; Blanco, Sandra; Nickerson, Elizabeth; Gabriel, Stacey et al. (2012): Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. In *Journal of medical genetics* 49 (6), pp. 380–385. DOI: 10.1136/jmedgenet-2011-100686.

Martinez, German; Choudury, Sarah G.; Slotkin, R. Keith (2017): tRNA-derived small RNAs target transposable element transcripts. In *Nucleic acids research* 45 (9), pp. 5142–5152. DOI: 10.1093/nar/gkx103.

Maute, Roy L.; Schneider, Christof; Sumazin, Pavel; Holmes, Antony; Califano, Andrea; Basso, Katia; Dalla-Favera, Riccardo (2013): tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (4), pp. 1404–1409. DOI: 10.1073/pnas.1206761110.

Meng, D.; Li, Z.; Ma, X.; Fu, L.; Qin, G. (2016): MicroRNA-1280 modulates cell growth and invasion of thyroid carcinoma through targeting estrogen receptor α. In *Cellular and Molecular Biology* 62 (3), pp. 1–6.

Mirande, Marc (2010): Processivity of translation in the eukaryote cell: Role of aminoacyl-tRNA synthetases. In *FEBS letters* 584 (2), pp. 443–447. DOI: 10.1016/j.febslet.2009.11.027.

Mohn, Fabio; Handler, Dominik; Brennecke, Julius (2015): Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. In *Science (New York, N.Y.)* 348 (6236), pp. 812–817. DOI: 10.1126/science.aaa1039.

Moncayo, Álvaro; Silveira, Antonio Carlos (2017): Current epidemiological trends of Chagas disease in Latin America and future challenges: Epidemiology, surveillance, and health policies. In : American Trypanosomiasis Chagas Disease: Elsevier, pp. 59–88.

Mori, Kohji; Weng, Shih-Ming; Arzberger, Thomas; May, Stephanie; Rentzsch, Kristin; Kremmer, Elisabeth et al. (2013): The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. In *Science (New York, N.Y.)* 339 (6125), pp. 1335–1338. DOI: 10.1126/science.1232927.

Neeb, Zachary T.; Nowacki, Mariusz (2018): RNA-mediated transgenerational inheritance in ciliates and plants. In *Chromosoma* 127 (1), pp. 19–27. DOI: 10.1007/s00412-017-0655-4.

Negrutskii, B. S.; Deutscher, M. P. (1991): Channeling of aminoacyl-tRNA for protein synthesis in vivo. In *Proceedings of the National Academy of Sciences* 88 (11), pp. 4991–4995. DOI: 10.1073/pnas.88.11.4991.

Panas, Marc D.; Ivanov, Pavel; Anderson, Paul (2016): Mechanistic insights into mammalian stress granule dynamics. In *The Journal of cell biology* 215 (3), pp. 313–323. DOI: 10.1083/jcb.201609081.

Paul, Prosenjit; Chakraborty, Anindya; Sarkar, Debasree; Langthasa, Malobika; Rahman, Musfhia; Bari, Minakshi et al. (2018): Interplay between miRNAs and human diseases. In *Journal of cellular physiology* 233 (3), pp. 2007–2018. DOI: 10.1002/jcp.25854.

Peng, Hongying; Shi, Junchao; Zhang, Ying; Zhang, He; Liao, Shangying; Li, Wei et al. (2012): A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. In *Cell research* 22, 1609 EP -. DOI: 10.1038/cr.2012.141.

Piepoli, Ada; Tavano, Francesca; Copetti, Massimiliano; Mazza, Tommaso; Palumbo, Orazio; Panza, Anna et al. (2012): Mirna expression profiles identify drivers in colorectal and pancreatic cancers. In *PloS one* 7 (3), e33663. DOI: 10.1371/journal.pone.0033663.

Rechavi, Oded; Lev, Itamar (2017): Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans. In *Current biology : CB* 27 (14), R720-R730. DOI: 10.1016/j.cub.2017.05.043.

Reifur, Larissa; Garcia-Silva, Maria Rosa; Poubel, Saloê Bispo; Alves, Lysangela Ronalte; Arauco, Paulo; Buiar, Diane Kelly et al. (2012): Distinct subcellular localization of tRNA-derived fragments in the infective metacyclic forms of Trypanosoma cruzi. In *Memorias do Instituto Oswaldo Cruz* 107 (6), pp. 816–819.

Ruggero, Katia; Guffanti, Alessandro; Corradin, Alberto; Sharma, Varun Kumar; Bellis, Gianluca de; Corti, Giorgio et al. (2014): Small noncoding RNAs in cells transformed by human T-cell leukemia virus type 1: a role for a tRNA fragment as a primer for reverse transcriptase. In *Journal of virology* 88 (7), pp. 3612–3622. DOI: 10.1128/JVI.02823-13.

Rybak, Susanna M.; Vallee, Bert L. (1988): Base cleavage specificity of angiogenin with Saccharomyces cerevisiae and Escherichia coli 5S RNAs. In *Biochemistry* 27 (7), pp. 2288–2294.

Saikia, Mridusmita; Jobava, Raul; Parisien, Marc; Putnam, Andrea; Krokowski, Dawid; Gao, Xing-Huang et al. (2014): Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress. In *Molecular and cellular biology* 34 (13), pp. 2450–2463. DOI: 10.1128/MCB.00136-14.

Sarker, Gitalee; Sun, Wenfei; Rosenkranz, David; Pelczar, Pawel; Opitz, Lennart; Efthymiou, Vissarion et al. (2019): Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. In *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1820810116.

Saxe, Jonathan P.; Chen, Mengjie; Zhao, Hongyu; Lin, Haifan (2013): Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. In *The EMBO journal* 32 (13), pp. 1869–1885. DOI: 10.1038/emboj.2013.121.

Schaffer, Ashleigh E.; Eggens, Veerle R. C.; Caglayan, Ahmet Okay; Reuter, Miriam S.; Scott, Eric; Coufal, Nicole G. et al. (2014): CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. In *Cell* 157 (3), pp. 651–663.

Schorn, Andrea J.; Gutbrod, Michael J.; LeBlanc, Chantal; Martienssen, Rob (2017): LTR-Retrotransposon Control by tRNA-Derived Small RNAs. In *Cell* 170 (1), 61-71.e11. DOI: 10.1016/j.cell.2017.06.013.

Sebastià, J.; Kieran, D.; Breen, B.; King, M. A.; Netteland, D. F.; Joyce, D. et al. (2009): Angiogenin protects motoneurons against hypoxic injury. In *Cell Death and Differentiation* 16 (9), p. 1238. DOI: 10.1038/cdd.2009.52.

Seiki, M.; Hattori, S.; Yoshida, M. (1982): Human adult T-cell leukemia virus: molecular cloning of the provirus DNA and the unique terminal structure. In *Proceedings of the National Academy of Sciences* 79 (22), pp. 6899–6902. DOI: 10.1073/pnas.79.22.6899.

Selitsky, Sara R.; Baran-Gale, Jeanette; Honda, Masao; Yamane, Daisuke; Masaki, Takahiro; Fannin, Emily E. et al. (2015): Small tRNA-derived RNAs are increased and more abundant than microRNAs in chronic hepatitis B and C. In *Scientific reports* 5, p. 7675. DOI: 10.1038/srep07675.

Shapiro, Robert; Vallee, Bert L. (1989): Site-directed mutagenesis of histidine-13 and histidine-114 of human angiogenin. Alanine derivatives inhibit angiogenin-induced angiogenesis. In *Biochemistry* 28 (18), pp. 7401–7408. DOI: 10.1021/bi00444a038.

Sharma, Upasna; Conine, Colin C.; Shea, Jeremy M.; Boskovic, Ana; Derr, Alan G.; Bing, Xin Y. et al. (2016): Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. In *Science (New York, N.Y.)* 351 (6271), pp. 391–396. DOI: 10.1126/science.aad6780.

Short, A. K.; Yeshurun, S.; Powell, R.; Perreau, V. M.; Fox, A.; Kim, J. H. et al. (2017): Exercise alters mouse sperm small noncoding RNAs and induces a transgenerational modification of male offspring conditioned fear and anxiety. In *Translational psychiatry* 7, e1114 EP -. DOI: 10.1038/tp.2017.82.

Skorupa, Alexandra; King, Matthew A.; Aparicio, Isabela M.; Dussmann, Heiko; Coughlan, Karen; Breen, Bridget et al. (2012): Motoneurons secrete angiogenin to induce RNA cleavage in astroglia. In *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32 (15), pp. 5024–5038. DOI: 10.1523/JNEUROSCI.6366-11.2012.

Sobala, Andrew; Hutvagner, Gyorgy (2013): Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. In *RNA biology* 10 (4), pp. 553–563. DOI: 10.4161/rna.24285.

Sun, Vera; Zhou, Wen B.; Nosrati, Mehdi; Majid, Shahana; Thummala, Suresh; Semir, David de et al. (2015): Antitumor activity of miR-1280 in melanoma by regulation of Src. In *Molecular Therapy* 23 (1), pp. 71–78.

Tello-Montoliu, A.; Patel, J. V.; Lip, G. Y. H. (2006): Angiogenin: a review of the pathophysiology and potential clinical applications. In *Journal of thrombosis and haemostasis : JTH* 4 (9), pp. 1864–1874. DOI: 10.1111/j.1538-7836.2006.01995.x.

Telonis, Aristeidis G.; Kirino, Yohei; Rigoutsos, Isidore (2015): Mitochondrial tRNA-lookalikes in nuclear chromosomes: could they be functional? In *RNA biology* 12 (4), pp. 375–380. DOI: 10.1080/15476286.2015.1017239.

Thiyagarajan, Nethaji; Ferguson, Ross; Subramanian, Vasanta; Acharya, K. Ravi (2012): Structural and molecular insights into the mechanism of action of human angiogenin-ALS variants in neurons. In *Nature communications* 3, p. 1121. DOI: 10.1038/ncomms2126.

Torres, Adrian Gabriel; Reina, Oscar; Stephan-Otto Attolini, Camille; Ribas de Pouplana, Lluís (2019): Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. In *Proceedings of the National Academy of Sciences of the United States of America* 116 (17), pp. 8451–8456. DOI: 10.1073/pnas.1821120116.

Tuorto, Francesca; Liebers, Reinhard; Musch, Tanja; Schaefer, Matthias; Hofmann, Sarah; Kellner, Stefanie et al. (2012): RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. In *Nature structural & molecular biology* 19 (9), p. 900. DOI: 10.1038/nsmb.2357.

Ullu, Elisabetta; Tschudi, Christian; Chakraborty, Tirtha (2004): RNA interference in protozoan parasites. In *Cellular microbiology* 6 (6), pp. 509–519. DOI: 10.1111/j.1462-5822.2004.00399.x.

Wang, Fengfei; Remke, Marc; Bhat, Kruttika; Wong, Eric T.; Zhou, Shuang; Ramaswamy, Vijay et al. (2015): A microRNA-1280/JAG2 network comprises a novel biological target in high-risk medulloblastoma. In *Oncotarget* 6 (5), pp. 2709–2724. DOI: 10.18632/oncotarget.2779.

Wang, Qingrong; Lee, Inhan; Ren, Junping; Ajay, Subramanian Shankar; Lee, Yong Sun; Bao, Xiaoyong (2013): Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. In *Molecular therapy : the journal of the American Society of Gene Therapy* 21 (2), pp. 368–379. DOI: 10.1038/mt.2012.237.

Xu, Li-Min; Li, Li-Qin; Li, Jing; Li, Hong-Wei; Shen, Qi-Bin; Ping, Jin-Liang et al. (2015): Upregulation of miR-1280 expression in non-small cell lung cancer tissues. In *Chinese medical journal* 128 (5), pp. 670–673. DOI: 10.4103/0366-6999.151672.

Yamasaki, Satoshi; Ivanov, Pavel; Hu, Guo-Fu; Anderson, Paul (2009): Angiogenin cleaves tRNA and promotes stress-induced translational repression. In *The Journal of cell biology* 185 (1), pp. 35–42. DOI: 10.1083/jcb.200811106.

Yeung, Man Lung; Bennasser, Yamina; Watashi, Koichi; Le, Shu-Yun; Houzet, Laurent; Jeang, Kuan-Teh (2009): Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. In *Nucleic acids research* 37 (19), pp. 6575–6586. DOI: 10.1093/nar/gkp707.

Zhang, Donglei; Tu, Shikui; Stubna, Michael; Wu, Wei-Sheng; Huang, Wei-Che; Weng, Zhiping; Lee, Heng-Chi (2018a): The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. In *Science (New York, N.Y.)* 359 (6375), pp. 587–592. DOI: 10.1126/science.aao2840.

Zhang, Yunfang; Zhang, Xudong; Shi, Junchao; Tuorto, Francesca; Li, Xin; Liu, Yusheng et al. (2018b): Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. In *Nature cell biology* 20 (5), p. 535. DOI: 10.1038/s41556-018-0087-2.

2. Expression profile and targeting rules of 5' tRNA halves

Julia Jehn[1], Jana Treml[1], Svenja Wulsch[1], Benjamin Ottum[1], Verena Erb[1], Charlotte Hewel[1], Roxana N. Kooijmans[2], Laura Wester[1], Isabel Fast[1], David Rosenkranz[1,3]

[1]Institute of Organismic and Molecular Evolution iOME, Anthropology, Johannes Gutenberg University Mainz, Anselm-Franz-von-Bentzel-Weg 7, 55099 Mainz, Germany.

[2]Primate Brain Bank, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands.

[3]Senckenberg Zentrum für Humangenetik, Weismüllerstraße 50, 60314 Frankfurt am Main, Germany

This chapter is publicly available as a preprint on bioRxiv under the title "5' tRNA halves are highly expressed in the primate hippocampus and sequence-specifically regulate gene expression" (doi:10.1101/774380) and is in preparation to be published. Numbering of figures and tables as well as the citation style have been adjusted.

2.1. Abstract

Fragments of mature tRNAs have long been considered as mere degradation products without physiological function. However, recent reports show that tRNA fragments (tRFs) play prominent roles in diverse cellular processes across a wide spectrum of species. Contrasting the situation in other small RNA pathways the mechanisms behind these effects appear more diverse, more complex and are generally less well understood. In addition, surprisingly little is known about the expression profiles of tRFs across different tissues and species. Here, we provide an initial overview of tRF expression in different species and tissues, revealing very high tRF-levels particularly in the primate hippocampus. We further modulated the regulation capacity of selected tRFs in human cells by transfecting synthetic tRF mimics ("overexpression") or antisense-RNAs ("inhibition") and identified differentially expressed transcripts based on RNAseq. We then used a novel k-mer mapping approach to dissect the underlying targeting rules, demonstrating that 5' tRNA halves (5' tRHs) silence genes in a sequence-specific manner, while the most efficient target sites align to the mid-region of the 5' tRH and are located within the CDS or 3' UTR of the target. This amends previous observations that tRFs guide Argonaut proteins to silence their targets via a miRNA-like 5' seed match and suggests a yet unknown mechanism of regulation. Finally, our data suggests that some 5' tRHs are also able to sequence-specifically stabilize mRNAs as upregulated mRNAs are also significantly enriched for 5' tRH target sites.

2.2. Introduction

tRNAs are well known for their conserved role in protein biosynthesis. However, mature tRNAs and their precursors give also rise to a class of small non-coding RNAs, the tRNA fragments (tRFs). While tRF1s (15-22 nt) are generated by clipping off the trailer sequence from the tRNA precursor molecule, 5' tRFs and 3' tRFs (18-35 nt) stem from the respective end of mature

tRNAs (70-90 nt). Primarily considered as degradation products, tRFs are now recognized as additional players in small RNA-mediated gene regulation that act in a variety of cellular processes across species from all domains of life.

tRFs were found to play a role in fundamental physiological processes such as proliferation (Lee et al. 2009) and protein translation control (Ivanov et al. 2011; Gebetsberger et al. 2017; Keam et al. 2017; Guzzi et al. 2018). Moreover, tRFs are implicated in defense mechanisms of *Escherichia coli* against bacteriophages (Levitz et al. 1990) and human cells against trypanosoma or viruses (Garcia-Silva et al. 2014; Deng et al. 2015). Additionally, tRFs were found to prime the reverse transcription or inhibit the promotion of retroviruses and retrotransposons (Ruggero et al. 2014; Yeung et al. 2009; Schorn et al. 2017; Martinez et al. 2017). Furthermore, tRFs are associated with several diseases such as cancer (Lee et al. 2009; Goodarzi et al. 2015; Huang et al. 2017) and amyotrophic lateral sclerosis (Greenway et al. 2006; Ivanov et al. 2014) and were lately revealed to act as transgenerational transmitters that induce metabolic disorders and addictive behavior in mice (Chen et al. 2016; Sharma et al. 2016; Short et al. 2017; Sarker et al. 2019). While some of the described effects seem to base on similar mechanisms, tRFs from the same class were shown to trigger completely converse functions in other cases (Jehn and Rosenkranz 2019), suggesting that the regulatory potential of tRFs is more complex than observed for the well-studied major small RNA classes.

So far, two superordinate concepts that aim explain how tRFs regulate gene expression have been identified. While some studies show that tRFs globally repress translation by inhibiting the assembly of the translation initiation machinery (Ivanov et al. 2011; Gebetsberger et al. 2017; Keam et al. 2017; Guzzi et al. 2018), others report a sequence-specific gene regulation (Haussecker et al. 2010; Deng et al. 2015; Kuscu et al. 2018; Luo et al. 2018). As tRFs were found to associate with Argonaut proteins, a miRNA-like gene regulation mechanism seems apparent (Maute et al. 2013; Kumar et al. 2014; Kuscu et al. 2018). Indeed, studies could show that tRF-mediated transcript silencing is dependent on a 5’ seed, which is complementary to target sites within the 3’ UTR (Haussecker et al. 2010; Kuscu et al. 2018). Contrasting this, other studies observed sequence-specific gene silencing effects, where complementarity of the 5’ seed was neglectable in favor for other tRF portions (Deng et al. 2015; Luo et al. 2018). Since Argonaut structure coerces the 5' end of small RNAs for target recognition (Boland et al. 2011), the results of these studies suggest that tRFs interact with other effector proteins or act protein-independent to regulate genes. Indications from recent studies where tRFs were shown to interact with different proteins depending on the differentiation state of the cell or the modification status of the tRF further erode the AGO-centric view (Krishna et al. 2019; Guzzi et al. 2018) illustrating that we still lack a sufficiently deep understanding of how tRFs regulate gene expression mechanistically.

Surprisingly, although it has been noted that tRFs are much more abundant in tissues than in cultured cells (Torres et al. 2019), their expression profile across tissues and species has not yet been investigated. We therefore initially analyzed several available and own small RNA sequencing datasets to provide a first overview on tRF expression profiles. Subsequently, we tested if 5’ tRHs regulate genes by modulating the level of 5’ tRHs-Glu-CTC and 5’ tRH-Gly-GCC

in cultured cells by transfecting either synthetic RNA mimics or inhibiting antisense RNAs. By RNA sequencing, *in silico* target predictions and a novel k-mer mapping-based approach, we then dissected the targeting rules of the respective 5' tRHs. Finally, applying the k-mer analysis on similar RNA sequencing datasets of other species, we examined whether the identified targeting patterns of individual tRFs are conserved across species.

2.3. Results and Discussion

2.3.1. 5' tRHs are highly abundant in the hippocampus of primates

To investigate the expression levels of tRFs across different tissues, we annotated available human small RNA sequencing datasets complemented by own small RNA sequencing data and found 5' tRHs to be predominantly expressed in the hippocampus (figure 2.1A). Here, 30% of all mapped reads were classified as tRNA-fragments with 82% of these tRF-annotated sequences being 5' tRHs (figure 2.1B). In comparison, the overall tRF level was only 13% in both the frontal cortex and the cerebellum (figure 2.1B and C). Amongst the predominant tRNA-derived sequences in the hippocampus were the 5' tRNA-halves 5' tRH-Glu-CTC (26%) and 5' tRH-Gly-GCC (10%). Notably, these 5' tRHs were also amongst the most abundant 5' tRHs in the analyzed small RNA sequencing libraries of other tissues (figure 2.1A). Both, 5' tRH-Glu-CTC and 5' tRH-Gly-GCC have been previously shown to play major roles in various cellular functions. They were amongst the sperm small RNAs that are involved in the epigenetic inheritance of paternal diet-induced metabolic disorders and addictive-like behavior in mice (Chen et al. 2016; Sharma et al. 2016; Zhang et al. 2018b; Sarker et al. 2019). Additionally, they were found amongst a group of 5' tRHs to be upregulated in cells upon infection by the Respiratory Syncytial Virus (Wang et al. 2013). 5' tRH-Gly-GCC was furthermore found to be part of a specific subset of 5' tRHs that is dynamically expressed during stem cell differentiation (Krishna et al. 2019), while 5' tRH-Glu-CTC was found to be highly expressed in human monocytes, where it triggers the transcriptional suppression of the surface glycoprotein CD1 (Zhang et al. 2016).

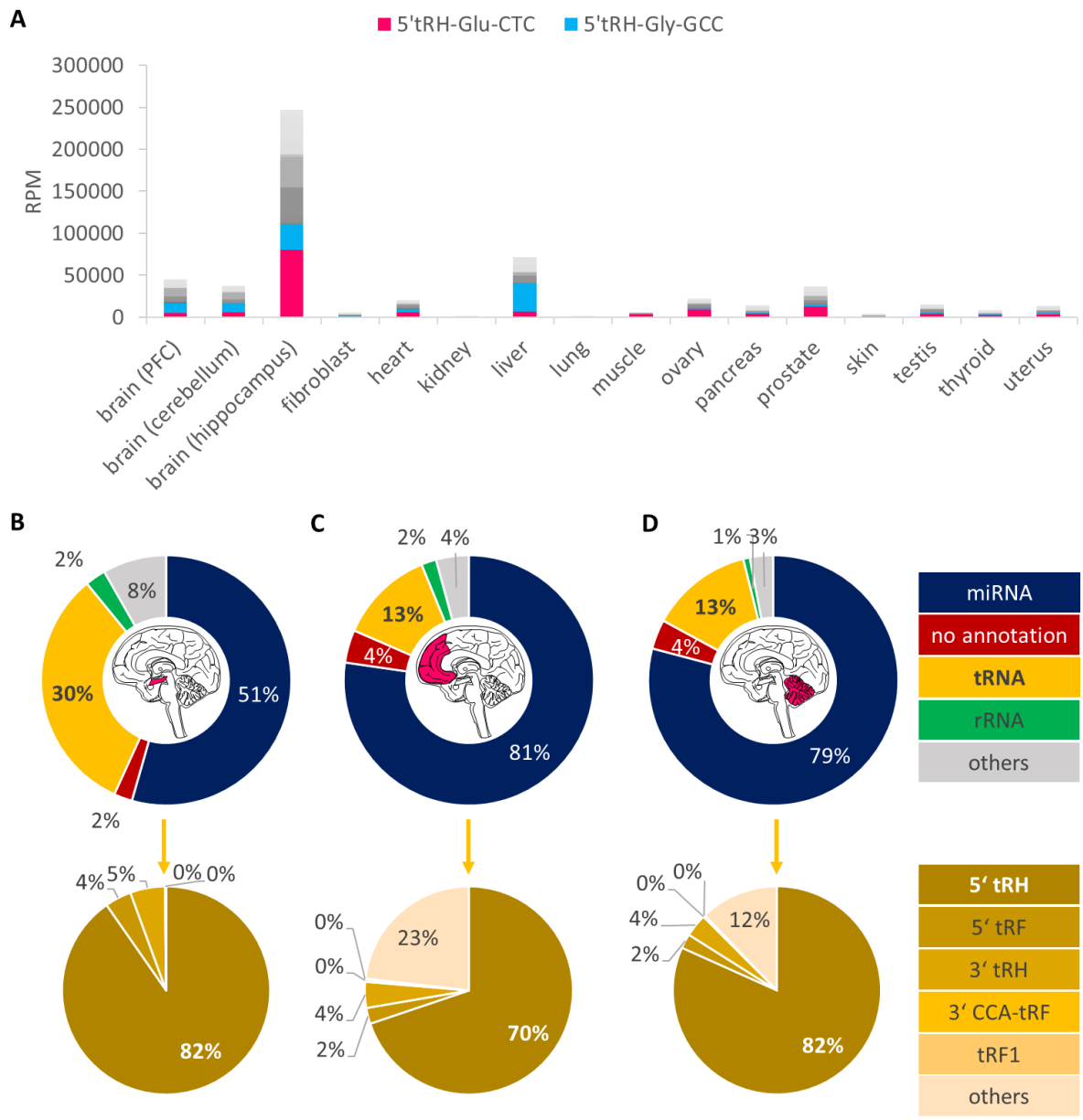


Figure 2.1: (A) Reads per million (RPM) of small RNAs annotated as 5’ tRHs in small RNA sequencing datasets of human tissues. The mean value of multiple datasets is shown. (B-D) Small RNA annotation of mapped reads from small RNA libraries of the human hippocampus, frontal cortex and cerebellum.

In order to investigate whether high 5’ tRH levels in the hippocampus are a common feature in mammals, we generated small RNA sequencing libraries of chimpanzee and macaque hippocampus samples, and additionally analyzed publicly available small RNA sequencing datasets from pig, rat and mouse hippocampus. Interestingly, we found even higher proportions of reads being annotated as tRFs in the hippocampal libraries of the two primates (42% in the chimpanzee / 51% in the macaque; figure 2.2) compared to the human hippocampus libraries (30%). Again, 5’ tRHs made up the majority of the tRF-annotated reads (82% / 75%, figure 2.2) and 5’ tRH-Gly-GCC (15% / 33%) and 5’ tRH-Glu-CTC (14% / 11%) were amongst the most abundantly expressed tRFs. In contrast, even though 5’ tRHs are also the

predominant tRF class in the analyzed sequencing libraries of the other mammals, only 10% of the pig and 3% of the rodent mapped reads were classified as tRFs (figure 2.2). This suggests that high levels of 5' tRHs in the hippocampus are a primate-specific trait, which raises the question if a conserved subset of 5' tRHs specifically fine-tunes hippocampal gene expression in primates.

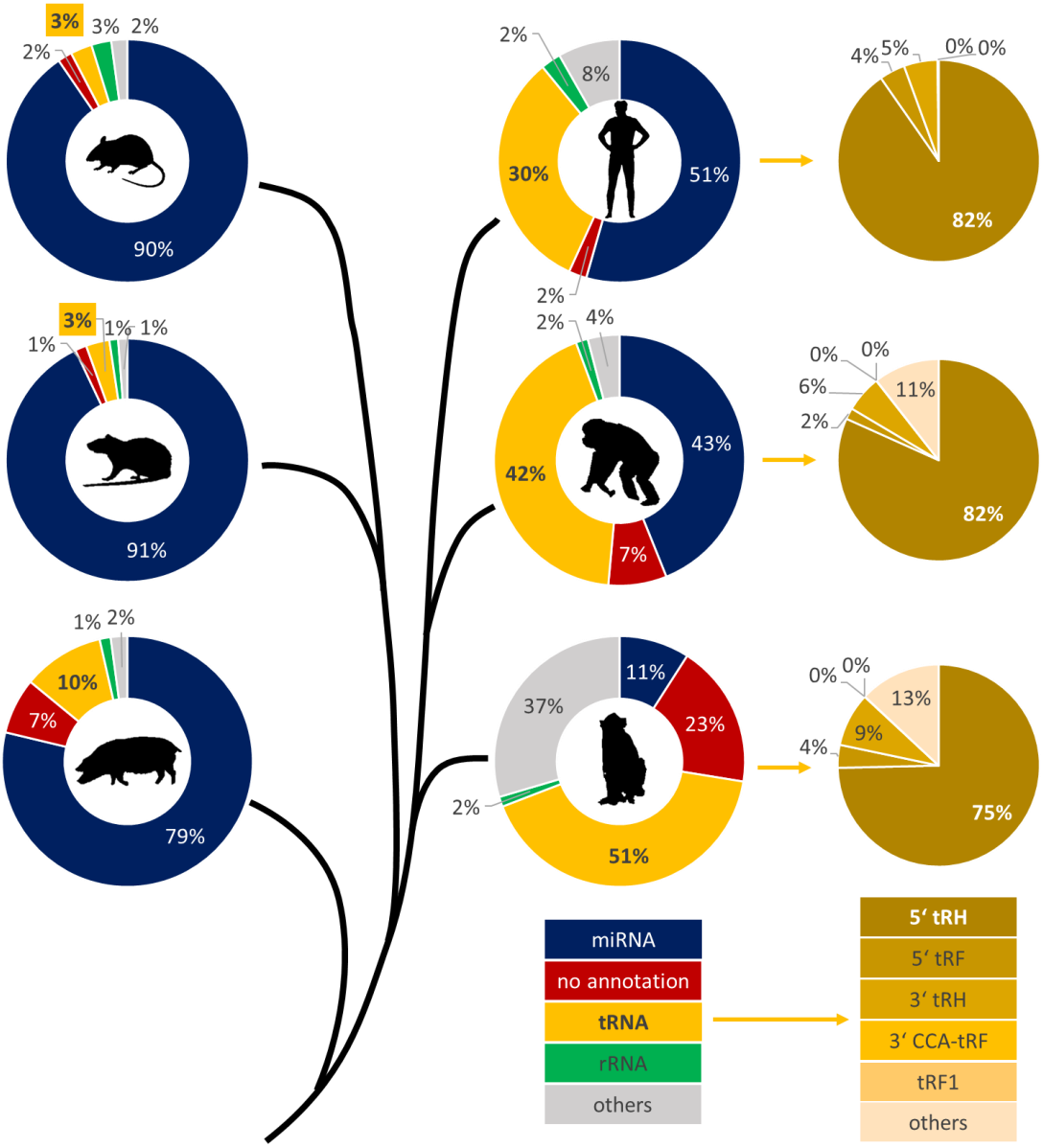


Figure 2.2: Phylogenetic tree and small RNA annotation of mapped reads from small RNA sequencing libraries of hippocampal samples from mouse, rat, pig, human, chimpanzee and macaque.

2.3.2. miRNA- and piRNA-like targeting rules scarcely identify targets of the 5' tRNA-halves Glu-CTC and Gly-GCC in HEK293T

In order to identify genes that are regulated by these 5' tRHs, we overexpressed 5' tRH-Glu-CTC and 5' tRH-Gly-GCC in HEK293T cells by transfecting 50 nM synthetic 5' tRH mimics. HEK293T cells were chosen as 5' tRHs are barely expressed in this cell line: from the 9% of all mapped reads that were assigned as tRNA-fragments, only 1% are 5' tRHs (S-Figure 2.1A). As validated by quantitative RT-PCR (qPCR), transfection of the RNA mimics successfully increased the number of 5' tRH copies in comparison to a non-target-siRNA transfection control (S-Figure 2.2A).

As a first test, we quantified the expression change of twenty abundantly expressed transcripts, which were predicted to represent targets of the respective tRFs according to miRNA (miRanda) or piRNA targeting rules, via qPCR. For piRNA-like target prediction, we developed a software named piRanha that by default applies targeting rules empirically verified by Zhang et al. (Zhang et al. 2018a). All ten tested potential targets of 5' tRH-Gly-GCC were lower expressed in the overexpression cells compared to the control cells. In contrast, neither the five miRanda- nor the five piRanha-predicted 5' tRH-Glu-CTC-target transcripts were differentially expressed upon overexpression of the 5' tRH-Glu-CTC (figure 2.3A), suggesting that target rules might differ between different tRFs.

To get a global overview of 5' tRH-regulated genes, we then sequenced the transcriptome of HEK293T 5' tRH-overexpression and control cells and performed differential expression analysis. Upon overexpression of the respective 5' tRH, more genes were significantly downregulated than upregulated (adjusted p-value < 0.01). However, most of the potential 5' tRH targets previously quantified by qPCR were not significantly differentially expressed (figure 2.3B).

When analyzing the cumulative distribution of all miRanda and piRanha predicted genes, only miRanda-targets of 5' tRH-Gly-GCC were enriched for downregulated genes over the general distribution (figure 2.3C and D). We therefore conclude that for 5' tRNA-halves, the targeting rules of piRNAs are not applicable, while the targeting rules of miRNAs cannot satisfyingly predict 5' tRH targets neither.

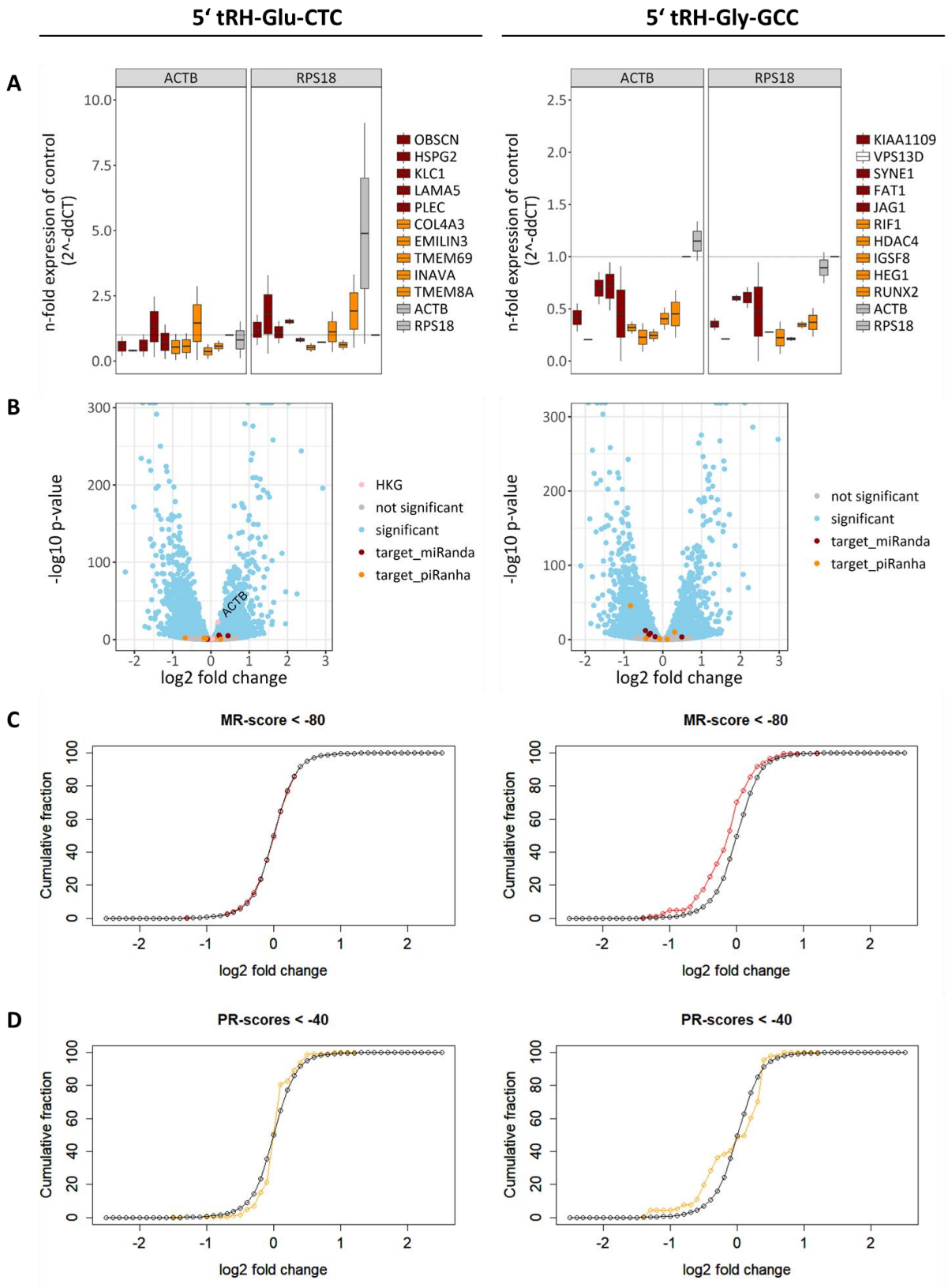


Figure 2.3: (A) n-fold expression of potential 5’ tRH targets in 5’ tRH overexpression HEK293T cells compared to control cells as quantified by qPCR. The selected transcripts were predicted by miRanda (red) and piRanha (orange) to be targets of the 5' tRNA-halves Glu-CTC (left) or Gly-GCC (right). The housekeeping genes β-actin and RPS18 were used as normalizers. Note that either β-actin or RPS18 expression changes upon transfection of the 5' tRNA-half Glu-CTC. (B) Volcano plot of differential expression analysis for protein-coding genes of 5’ tRH overexpression and control HEK293T cells (blue, adjusted p-value < 0.01). The 10 previously tested putative

target genes are highlighted in red (miRanda prediction) and orange (piRanda prediction). For the 5’ tRH-Glu-CTC analysis (left), the housekeeping genes β-actin and RPS18 are additionally highlighted in pink, showing that β-actin (ACTB) gets upregulated upon 5’ tRH-Glu-CTC overexpression, while the RPS18 expression is not affected. (C/D) Cumulative plots for log2-fold-change values of genes that were identified as potential targets of the 5' tRHs Glu-CTC or Gly-GCC by miRanda (red) or piRanha (orange) and of all genes (black). Only in 5’ tRH-Gly-GCC overexpression HEK293T cells miRanda-predicted genes are significantly more repressed.

2.3.3. Non-miRNA-like targeting rules for 5’ tRHs

To unravel the targeting rules of 5’ tRHs, we developed a k-mer mapping approach, where k-mers of the 5’ tRH sequences with all possible lengths ($k \geq 5$) and start positions within the tRH are mapped separately to the three major regions of each human transcript (5’ UTR, CDS and 3’ UTR). By calculating the fraction of “targeted” (k-mer alignment) or “not targeted” (no k-mer alignment) genes per k-mer that get significantly downregulated (adjusted p-value < 0.01) we identify the portion within the tRH that is most likely to be important for target recognition and thus silencing. Figure 2.4 visualizes the underlying principles of the analysis.

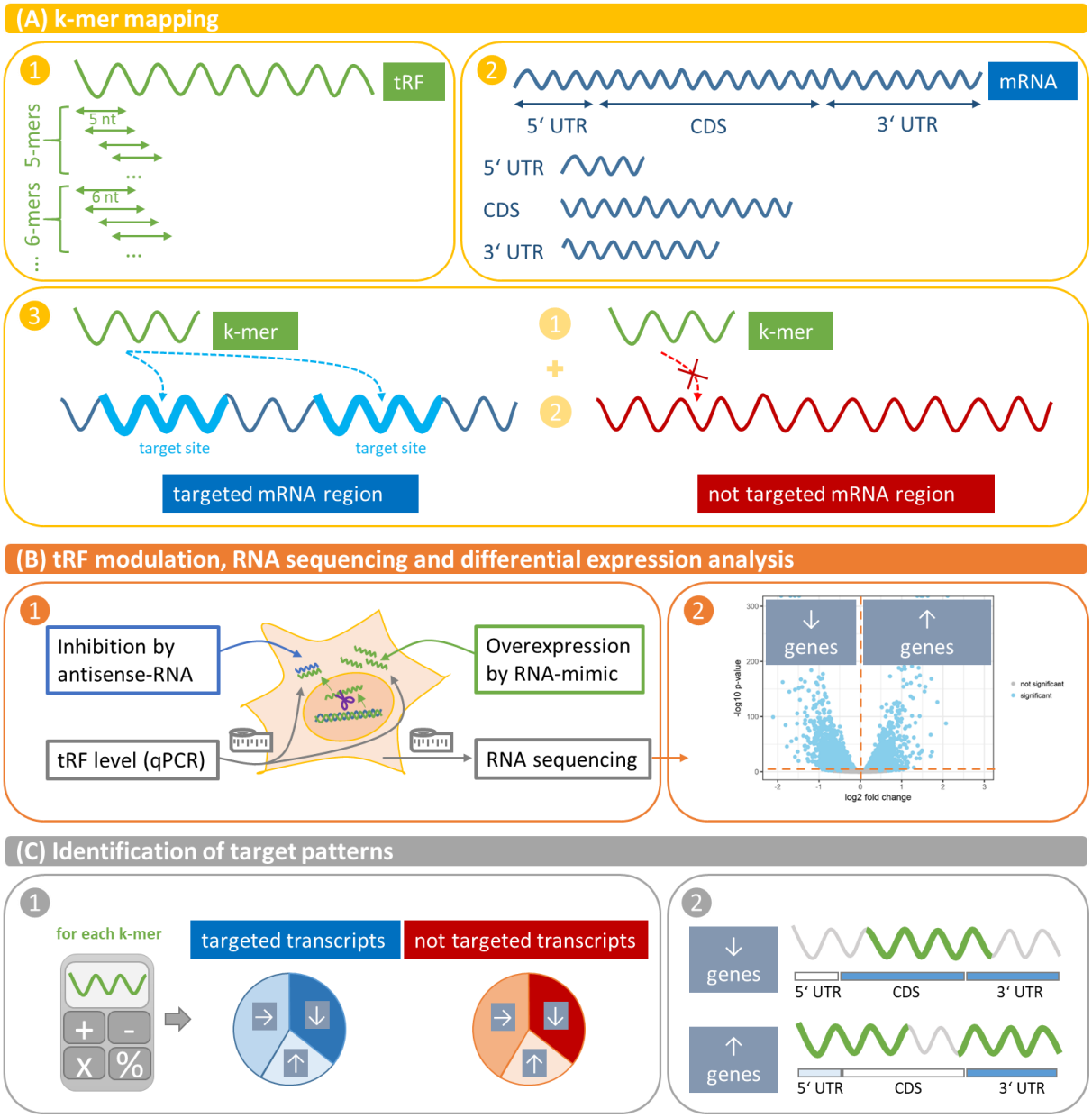


Figure 2.4: Graphical overview of identification of tRF targeting rules via a k-mer mapping approach.

Our analysis revealed that downregulation of transcripts with k-mer alignment (“targeted”) is significantly enriched over the number of downregulated transcripts without k-mer alignment (“not targeted”) particularly when k-mers of 5-10 nt length with start positions in the middle of the 5’ tRH align to the 3’ UTR or the CDS of the target (figure 2.5). Thus, unlike miRNAs that bind to the 3’ UTR of their targets via a 7 nt long seed at the 5’ end, 5’ tRNA-halves such as 5’ tRH-Glu-CTC and 5’ tRH-Gly-GCC downregulate transcripts mainly by binding with a 5-10 nt long stretch of their middle region to the 3’ UTR or the CDS of the target transcripts.

The resulting target pattern additionally suggests that 5’ tRH-Gly-GCC, but not 5’ tRH-Glu-CTC, is able to downregulate targets by binding with its miRNA-seed like 5’end to the transcript.

This is in line with, the different target regulation behavior of 5’ tRH-Glu-CTC and 5’ tRH-Gly-GCC as revealed by qPCR quantification of target transcripts predicted with miRanda.

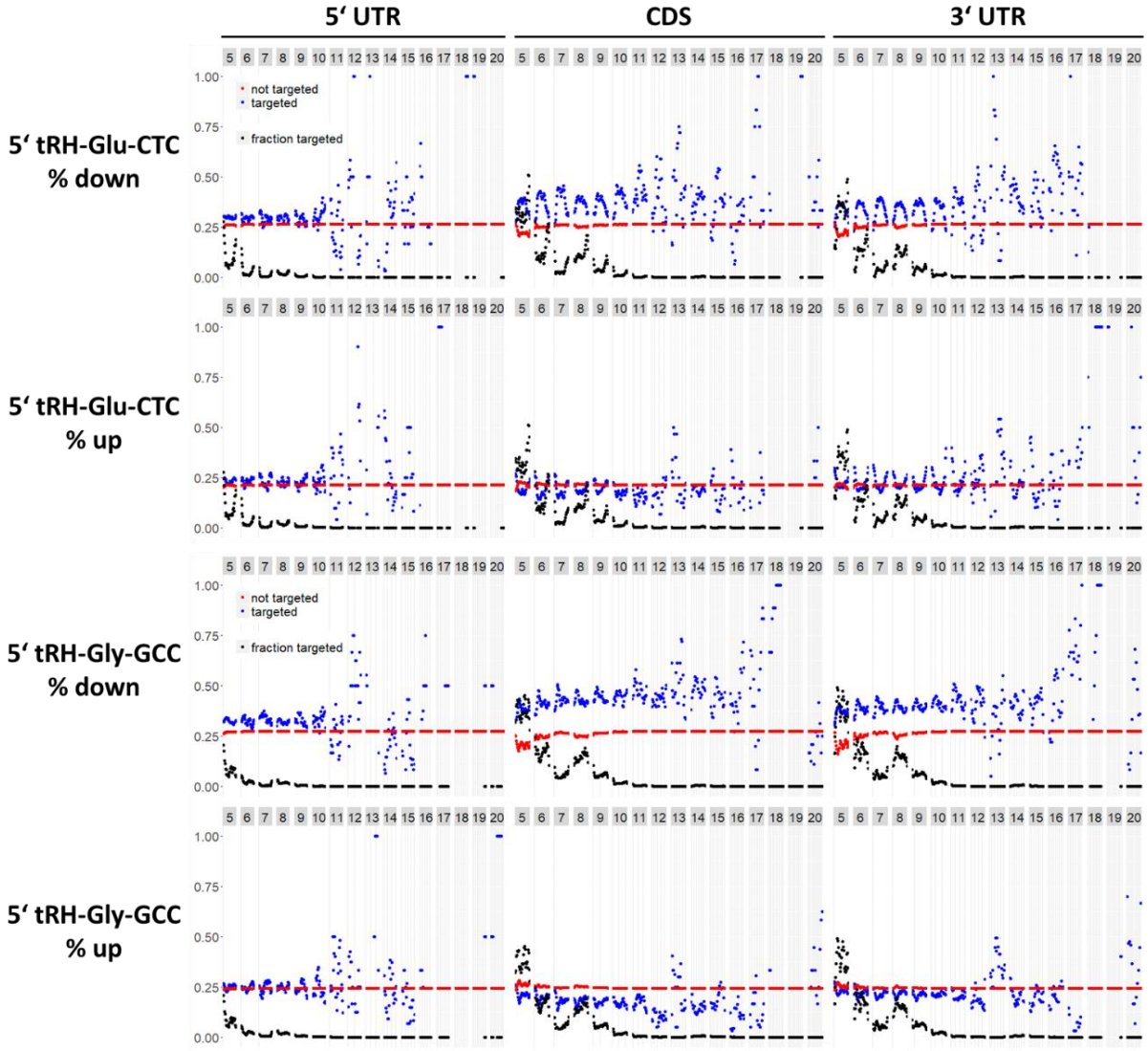


Figure 2.5: k-mer analysis to elucidate the targeting rules of 5’ tRH-mediated transcript regulation in HEK293T cells. The plots shows for each k-mer the percentage of transcripts with (blue) and without k-mer alignment (red) that are down/-upregulated. The fraction of genes targeted by the respective k-mer is displayed in black.

As a previous study found that 5’ tRH-Glu-CTC interacts with PIWIL4 to transcriptionally silence the surface glycoprotein CD1 in human monocytes (Zhang et al. 2016), we wanted to check whether we find indications for transcriptional silencing by 5’ tRH-Glu-CTC and 5’ tRH Gly-GCC in our datasets. We therefore restricted our k-mer analysis to alignments that span target gene exon-junctions. If the corresponding 5’ tRHs suppresses their target genes at the transcriptional level, we would expect a decreasing fraction of transcripts that have an exon-junction spanning alignment and that are downregulated. In contrast, the fraction of transcripts that have a non-exon-junction spanning alignment and that are downregulated

should not change. As we did not observe such an effect, we conclude that the studied 5' tRHs silence their targets at the post-transcriptional level in HEK293T cells (S-Figure 2.3).

Interestingly, mRNA quantification by qPCR and RNA sequencing revealed that the housekeeping gene *ACTB* gets upregulated upon 5' tRH-Glu-CTC transfection (figure 2.3A and B). Unexpectedly, *ACTB* has several potential binding sites for 5' tRH-Glu-CTC. Hence, we were curious to check whether upregulated genes might be enriched for 5' tRHs target sites as well, which would indicate that 5' tRH targeting might also have a protective effect. Indeed, our analysis revealed that transcripts that have target sites in their 3' UTR for the 5' and 3' ends of 5' tRH-Glu-CTC are more likely to be upregulated than those that do not have a target site, while a corresponding upregulating effect was not detectable for 5' tRH-Gly-GCC (figure 2.5).

Concluding from the above outlined target pattern analyses we assume that tRNA-fragments, even from the same tRF-series, may recognize their targets via different parts of the tRF instead of a dominating 5' seed match (figure 2.5). This demonstrates that a Argonaut-dependent regulation mechanism as proposed for tRFs previously (Haussecker et al. 2010; Kuscu et al. 2018) cannot fully explain the observed changes in gene expression. However, our results are in line with other studies identifying miRNA-untypical targeting (Deng et al. 2015; Luo et al. 2018). The mechanism seems to act at the post-transcriptional level, as transcripts with exon-junction spanning k-mer alignments were similarly often downregulated compared to transcripts with intra-exonic k-mer alignment (S-Figure 2.3). Our analysis additionally suggests that some tRFs (e.g. 5' tRH-Glu-CTC) trigger not only the downregulation of target genes, but may also stabilize target transcripts (figure 2.5). This is a rather unexpected finding, since a stabilizing effect of small RNAs is rarely seen in eukaryotes, but rather a trait of prokaryotic small RNA pathways (Fröhlich and Vogel 2009). However, a recent study showed that sequence-specific binding of a 3' tRF to the mRNAs of ribosomal proteins enhances the translation of the target transcript (Kim et al. 2017). It was suggested that structural changes induced by this interaction allow for higher transcription rates. Similar alterations in the secondary structure could also lead to a stabilization of the transcript.

2.3.4. Identification of genuine 5' tRH targets by antisense-inhibition of 5' tRH-Glu-CTC suggests a role of 5' tRHs in neurogenesis

Noteworthiliy, drawing conclusions regarding tRF targets only from overexpression experiments has several shortcomings. First, our synthetic RNA mimic does not carry post-transcriptional modifications as endogenous tRFs do. Thus, we might observe an artificial regulation behavior that does not reflect the physiological situation. Second, genes that were found to be differentially expressed upon tRF overexpression in cells that typically express these tRFs at very low levels must not necessarily be genuine targets of the modulated tRF under natural conditions. To circumvent these distorting effects and gain support for our conclusion, we additionally performed an inverse experiment where we inhibited the regulation capacity of 5' tRH-Glu-CTC by transfecting an antisense RNA.

We chose to perform the experiment with HepG2 cells, in order to exclude cell line specific regulation effects. HepG2 cells have a similar overall tRF level compared to HEK293T cells (6% of the total reads compared to 9% in HEK293T cells; S-Figure 2.1B), but express a higher proportion of 5' tRHs according to small RNA sequencing data (17% of the reads assigned to tRNAs compared to 1% in HEK293T cells; S-Figure 2.1B). Like in HEK293T cells, 5' tRH-Glu-CTC and 5' tRH-Gly-GCC are the most abundant 5' tRHs (6% and 3% of the reads assigned to tRNAs) in the analyzed library. As quantified by qPCR, transfection of the antisense-RNA decreased the number of 5' tRH-Glu-CTC copies in HepG2 cells by about 70% (S-Figure 2.2B). Even though the level of 5' tRH-Glu-CTC was decreased, only a small amount of genes was significantly differentially expressed compared to the control state (818 downregulated and 717 upregulated genes, adjusted p-value < 0.01; figure 2.6A). While the majority of significantly differentially expressed genes of the antisense-inhibited HepG2 was likewise regulated as in the overexpression HEK293T cells (figure 2.6B grey areas), only few genes were inversely regulated (figure 2.6B rosy and lutescent area) and we assume these genes to represent the genuine targets of 5' tRH-Glu-CTC. In the following we refer to the 34 genes that might get downregulated by 5' tRH-Glu-CTC as "perish targets" (figure 2.6B rosy area) and name the subset of 50 genes that get upregulated possibly due to stabilizing effects exerted by 5' tRH-Glu-CTC "shelter targets" (figure 2.6B lutescent area).

In order to characterize 5' tRH-targeting further, we subsetted the three major transcript regions (5' UTR, CDS, 3' UTR) of the potential perish and shelter targets and computed the thermodynamics for RNA-duplex formation with the 5' tRH-Glu-CTC. Considering the free energy needed to open intrinsic secondary structures, the free energy of the interaction and the so-called "dangling end" energies that non-interacting sequence stretches are causing, we identified the energetic optimal region of interaction. For the CDS and the 3' UTR of both the potential perish targets and the potential shelter targets, these optimal binding sites were enriched for k-mer alignments with a length of around 20 bp (figure 2.6C and D) and alignments preferentially started around the sixth position from the 5' end of the tRH. Thus, most thermodynamically favored interactions involved big stretches of the middle part of 5' tRH-Glu-CTC. This is in line with the target pattern identified by the k-mer mapping approach in case of the potential perish targets, but not with the target pattern for the potential shelter targets. As stabilizing effects probably involve other protein interactors than silencing effects, this objection may be a result of the stabilizing interactor exposing only the ends of the tRH for target recognition.

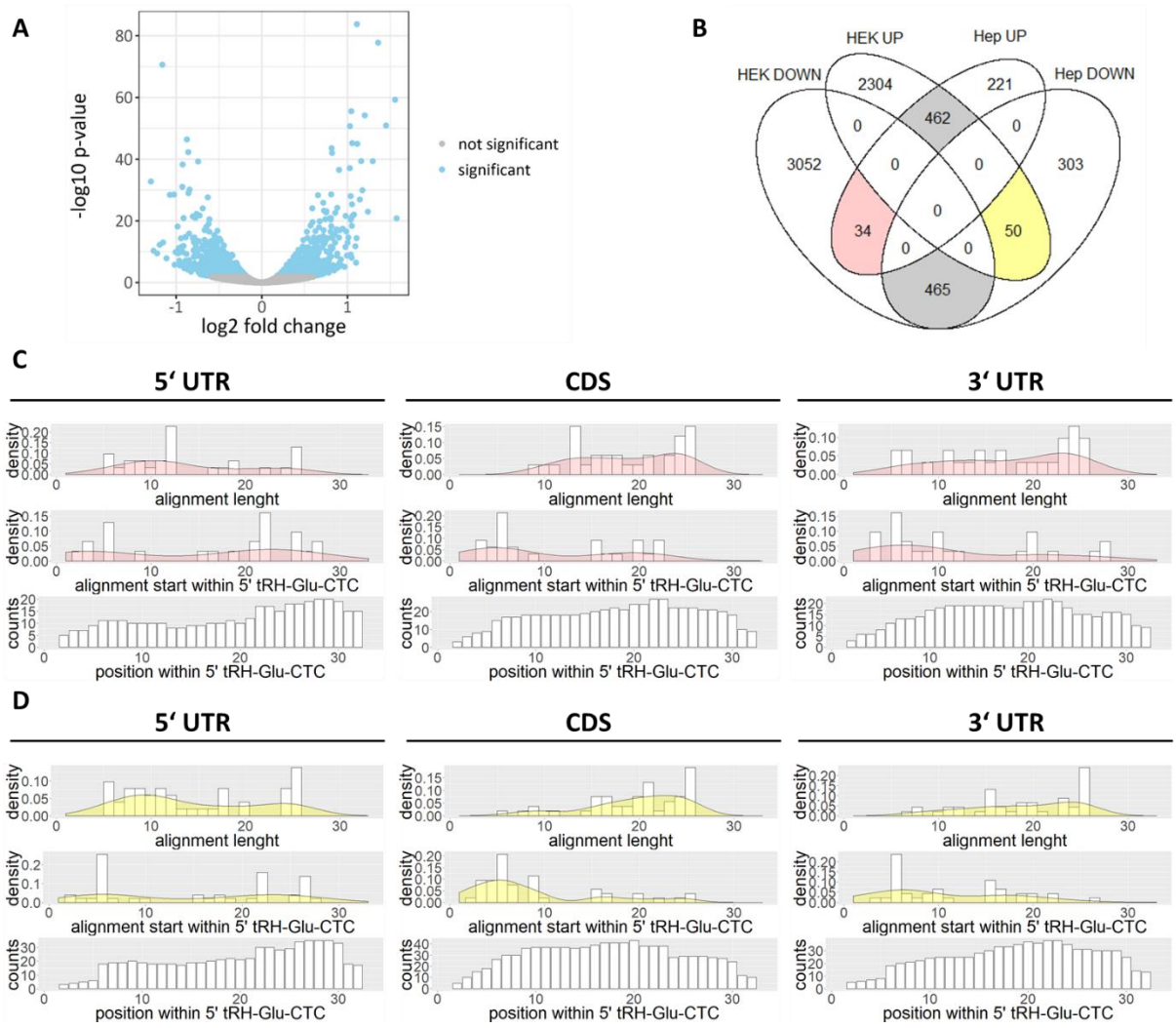


Figure 2.6: (A) Volcano plot of differential expression analysis for protein-coding genes of 5’ tRH antisense-inhibition and control HepG2 cells (blue, adjusted p-value < 0.01). (B) Venn diagram of significantly differentially expressed genes. Genes that are likewise regulated in HEK as in HepG2 cells are highlighted in grey. Genes that are inversely regulated in HEK overexpression and HepG2 inhibition cells are highlighted rosy (“potential targets”) and lutescent (“potential anti-targets”). (C/D) Analysis of thermodynamically favored alignments for the major transcript regions of “potential targets” or “potential anti-targets” with 5’ tRH-Glu-CTC.

As we found 5’ tRHs like 5’ tRH-Glu-CTC to be extremely abundant in small RNA libraries of primate hippocampal tissues and previous studies suggested a role in targeting genes involved in neural processes (Sarker et al. 2019; Krishna et al. 2019), we were interested whether the identified potential targets of 5’ tRH-Glu-CTC are implicated in neuronal processes. Indeed, we found 20% of the potential shelter targets (transcripts might be stabilized), but only 5% of the potential perish targets (transcripts might be degraded) to have an assigned neuronal function. In comparison only 7% and 8% of all expressed genes in HEK293T and HepG2 cells were assigned the gene ontology term “neurogenesis” (GO:0022008). Hence, there is a statistic correlation between genes that are potential targets of 5’ tRH-Glu-CTC and neurogenesis ($\alpha<0.001$ for shelter targets and $\alpha<0.01$ for perish targets; chi-squared test). While genes like *MDK*, *VEGFA* and *EVL* play a role in the regulation of axon outgrowth

(Kurosawa et al. 2001; He et al. 2015; Drees and Gertler 2008), genes like *NOTCH1* and *NP1L1* are involved in neuronal differentiation (Patten et al. 2006; Qiao et al. 2018). Interestingly, decapitation studies with the planarian *Dugesia japonica* showed that 5' tRH-Gly-GCC is not only upregulated in regenerating animals, but is furthermore required for proper head regeneration (Z Cao, D Rosenkranz, S Wu, H Liu, Q Pang, B Liu, B Zhao; manuscript in preparation). Taking into consideration that planarians share many CNS genes with humans (Mineta et al. 2003), and that the hippocampus is one of the few brain regions known to have high rates of adult neurogenesis (Eriksson et al. 1998; Boldrini et al. 2018), it is tempting to speculate that 5' tRHs may have critical functions in human neurogenesis as well.

2.3.5. Targeting pattern of 5' tRHs/tRFs are unique for each fragment, but seem to be conserved across species

In order to gain further support for our assumptions, we checked whether the identified 5' tRH target patterns are conserved among different species. We therefore analyzed published RNA sequencing data from Drosophila S2 cells that were transfected with a 20 nt long 5' tRF-Glu-CTC, which exhibits high sequence homology with the human 5' tRNA-half (S-Figure 2.4). As is the case for the human 5' tRNA-half, the fly 5' tRNA-fragment of the tRNA-Glu-CTC showed the strongest downregulating effect on targets when its ~7 nt long middle part binds to the 3' UTR or CDS of the target transcript (figure 2.7). However, unlike the human 5' tRNA-half, the both ends of the fly 5' tRF did not seem to contribute to target upregulation (data not shown), which is possibly due to its shorter sequence which lacks the corresponding nucleotides. Given the similar target regulation pattern, we suggest that 5' tRNA-fragments can regulate their targets via conserved mechanisms across different species, while the sequence stretch being most important for target regulation can vary for fragments from different parental tRNAs.

To confirm this assumption, we analyzed additional published RNA sequencing datasets from fly S2 cells that were transfected with the 20 nt long 5' tRNA-fragments Asp-GTC and Lys-TTT (figure 2.7). These analyses support our hypothesis, that 5' tRNA-fragments from different parental tRNAs act via different regions to regulate their targets on a sequence-complementary basis. What they have in common, is that 3' UTR and CDS targeting leads to the biggest regulatory effect.

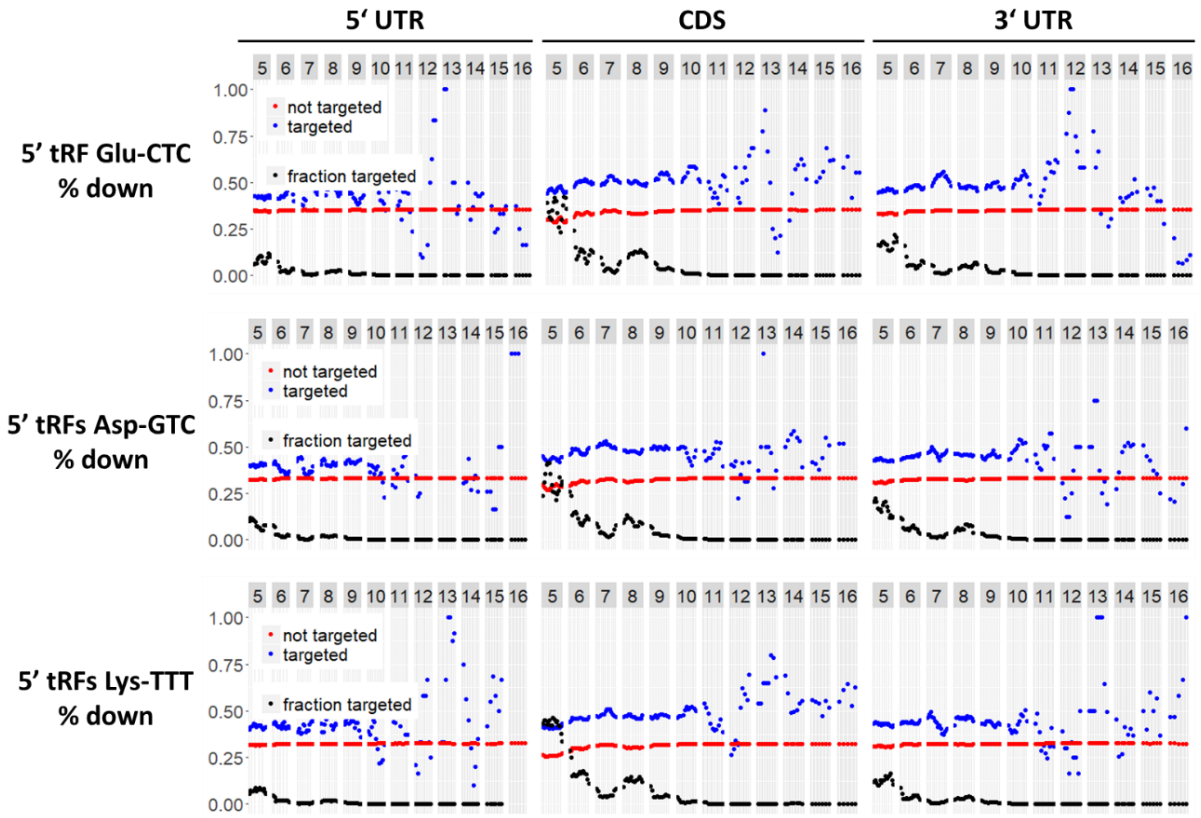


Figure 2.7: Target pattern analysis of published RNA sequencing data of fly S2 cells, where the 20 nt long 5' tRF Glu-CTC, 5' tRFs Asp-GTC or Lys-TTT was overexpressed by tRF mimic transfection (Luo et al. 2018). Plotted is the percentage of genes with (blue) or without (red) k-mer alignment that get downregulated.

2.3.6. Concluding remarks

Despite the rise of NGS techniques, exploring the mechanisms by which small RNAs target genes remains challenging. Even for miRNAs, where gene regulation mechanisms are well studied, correct target prediction is difficult since the *in vivo* accessibility of potential target sites is difficult to assess. RNA binding proteins may not only occupy the putative binding site, but may also change the secondary structure of the target when binding elsewhere in the transcript. Predicting and identifying targets of rather unexplored small RNAs like tRFs is the more complicated, as it is unknown if and to which extend mismatches, wobble base pairs and bulges are tolerated. Aggravatingly is that target interactions can be surprisingly variable (Backofen and Hess 2010).

Using a k-mer mapping approach, we sought to identify the target patterns of 5' tRHs. Our analysis suggested that 5' tRHs silence genes, which have complementary binding sites for long stretches of the tRNA half that not necessarily need to include the miRNA-typical 5' region. This finding strongly suggests that Argonaut proteins are not necessarily indispensable as effector proteins for tRF-dependent gene regulation as it had been suggested by other studies (Haussecker et al. 2010; Kuscu et al. 2018). Whether 5' tRHs independently regulate targets, or do so in association with other effector proteins than Argonauts remains to be investigated. As it was shown that specific subsets of tRFs bind to certain proteins (Saikia

et al. 2014; Goodarzi et al. 2015; Krishna et al. 2019; Ivanov et al. 2011), it is likely that 5' tRHs might interact with different proteins to regulate distinct targets. The tRNA 3' processing endoribonuclease RNase Z might be such an effector protein, as it was shown to cleave transcripts, which form RNA hybrids with 5' tRHs that have similar secondary structures like pre-tRNAs (Elbarbary et al. 2009). Identifying more proteins involved in tRF-mediated gene regulation and elucidating the underlying mechanism will greatly enhance our understanding of gene regulation.

Analyzing the expression profile of tRFs across tissues and species, we found 5' tRHs to be particularly high expressed in the hippocampus of primates, while their expression was rather low in the hippocampus of the pig, the rat and the mouse. In the hippocampus of all three tested primate species (human, chimpanzee and macaque), 5' tRH-Glu-CTC and 5' tRH-Gly-GCC were amongst the most abundant 5' tRHs suggesting that they have a conserved role in the primate hippocampus. As we find transcripts that are presumably stabilized by 5' tRH-Glu-CTC targeting to be substantially enriched for a function in neuronal processes such as axon outgrowth and neuronal differentiation, while transcripts that are presumably degraded upon 5' tRH-Glu-CTC targeting are rather depleted of a function in neurogenesis, we hypothesize that 5' tRHs play a role in fine-tuning primate neurogenesis. Alternatively, since retrotransposons are highly active in hippocampal neurons (Upton et al. 2015) and as 5' tRH-Gly-GCC was shown to downregulate transcripts that are driven by the LTR-retrotransposon MERVL in the developing mouse embryo (Sharma et al. 2016), it is also possible that hippocampal 5' tRH expression serves the purpose of regulating transposition-related events.

2.4. Materials and Methods

2.4.1. Small RNA sequencing, data processing and annotation

Small RNA sequencing data of various human tissues was downloaded from the SRA Database (for accession numbers see S-Table 2.1) and quality checked by FastQC (version 0.11.2). Adapter and quality trimming was performed using BBDuk (version 36.77; ktrim=r overwrite=true k=20 mink=9 ziplevel=2 hdist=1 qtrim=rl trimq=10 minlen=15 maxlen=34; for Encode data additionally: forcetrimleft=6) before the reads were FastQC checked again and mapped to the human genome (version GCA_000001405.27_GRCh38.p12) using Bowtie 2 (version 2.3.0). Based on these map-files small RNA annotation was performed with unitas (version 1.7.3; Gebert et al. 2017). Using the custom Perl script annotationtable2RPM.pl the RPM values were calculated for the respective tRF species.

Total RNA from adult normal male human hippocampal tissue was obtained from AMS Biotechnology (Cat. Nr: R1234052-10). A small RNA library was prepared as described in Gebert et al. 2015. In brief, small RNAs (15-40 nt) were extracted from a denaturing polyacrylamide gel. Subsequently, a 3' adapter (5'-rAppCTGTAGGCACCATCAATddC-3') and a 5' adapter (5'- GACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3') were ligated to the small RNAs. Following cDNA synthesis with the RT-Primer 5'-

ATTGATGGTGCCTACAG-3', the cDNA transcripts were PCR amplified using the forward primer 5'-ACATGGACTGAAGGAGTAGA-3' and the index-containing reverse primer 5'-ggctcATTGATGGTGCCTACAG-3'. The generated library was high throughput sequenced in parallel with six other indexed samples by GENterprise (Mainz) on an Illumina HiSeq 2000 system. After converting the fastq file to fasta formate using the NGS toolbox Perl script *fastq2fasta*, the 5' adapter sequence was clipped off the 120 nt long reads using the NGS toolbox Perl script *clip* (-m 5.AGTAGAAA). As reads may contain the reverse complementary sequence, the reverse complementary variant of the 5' adapter sequences were also clipped off (-m TTTCTACT.3) and the remaining sequences were transcribed to the original sequence direction using the NGS toolbox Perl script *rev-comp*. Both outputs were then concatenated using the NGS toolbox Perl script *concatenate*. To extract only the reads with the right index and to remove the 3' adapter sequence the NGS toolbox Perl script *clip* (-m CTGTA.GAGCC.3) was applied. Subsequent analysis was performed as described below.

Total RNA from tissue samples of human adult brain regions (hippocampus, cortex, cerebellum) was obtained from the BioChain Institute (Newark, CA, USA). Total RNA of hippocampal tissue samples of a female and a male chimpanzee brain was obtained from the National Chimpanzee Brain Resource (www.chimpanzeebrain.org, USA). A macaque brain provided by the Primate Brain Bank (The Netherlands Institute for Neuroscience, Amsterdam, Netherlands) was dissected and hippocampal tissue was homogenized in TRIzol™ (Thermo Fisher). Total RNA was isolated according to the manufacturers protocol. The total RNA was sent to LC Sciences (Houston, TX, USA) for small RNA library preparation and small RNA sequencing. The 3' adapter sequences were clipped off using the NGS toolbox (version 2.1; http://www.smallrnagroup.uni-mainz.de/) Perl script *clip* (-m TGGAATTC.3).

Small RNA sequencing data of hippocampal tissue from three male adult Wistar rats (Study: PRJEB24026, Run accessions: ERR2226477, ERR2226482 and ERR2226487) was downloaded from the European Nucleotide Archive. Subsequent analysis was performed as described below.

Small RNA sequencing data of hippocampal tissue from the pig (SRR3105507 and SRR3105508) and the mouse (SRR5144167, SRR5144168 and SRR5144169) was downloaded via the SRA toolkit tool *fastq-dump* (version 2.8.2). The 3' adapter sequences were clipped off using the NGS toolbox (version 2.1) Perl script *clip* (-m TGGAATTC.3). Subsequent analysis was performed as described below.

Total RNA of HEK293T cells was extracted with TRI Reagent™ (Thermo Fisher) 24 h after transfection with 10 nM Silencer™ Select Negative Control No. 1 siRNA (4390843; Thermo Fisher) using Lipofectamine™ RNAiMAX (Thermo Fisher). The RNA was sent to BGI (Hongkong) for small RNA library preparation and 50 bp single-end sequencing. The reads were delivered adapter-trimmed. Subsequent analysis was performed as described below.

Small RNA sequencing data of HepG2 cells (SRR6823987) was downloaded via the SRA toolkit tool *fastq-dump* (version 2.8.2). 3' adapter sequences were clipped off using the NGS toolbox

(version 2.1) Perl script *clip* (-m AGATCGGA.3). Subsequent analysis was performed as described below.

Adapter-trimmed data was first quality checked by FastQC (version 0.11.2), then converted to the fasta-format, length filtered (15-40 nt), collapsed to non-identical reads and depleted for low complexity reads using the NGS toolbox Perl scripts *length-filter*, *collapse* and *duster* (version 2.1; http://www.smallrnagroup.uni-mainz.de/). The remaining sequences were mapped to the respective genome (versions GCA_000001405.27_GRCh38.p12, GCA_000001515.5_Pan_tro_3.0, GCA_000772875.3_Mmul_8.0.1, GCA_000003025.6_Sscrofa11.1, GCA_000001895.4_Rnor_6.0 and GCA_000001635.8_GRCm38.p6) using the Perl script *sRNAmapper* (version 1.0.5; -a best) that employs SeqMap (Jiang and Wong 2008) as mapping tool. The map-files were used for small RNA annotation with unitas (version 1.6.1; Gebert et al. 2017).

2.4.2. Transfection of tRF-mimics and tRF antisense 2'-OMe-RNAs

HEK293T (2.5E4 cells/well) and HepG2 (1E5 cells/well) cells were seeded in 24-well plates and cultured in 1x GlutaMAX™-I DMEM supplemented with 10% FBS (Thermo Fisher). The next day the cells were transfected with 50 nM tRF-mimics or tRF antisense 2' OMe-RNAs (biomers) using Lipofectamine™ RNAiMAX (Thermo Fisher) according to the manufacturer's protocol. Sequences of the transfected RNAs are available in S-Table 2.2. As control, cells were transfected with Silencer™ Select Negative Control No. 1 siRNA (Thermo Fisher). 48 hours after transfection, the RNA was isolated according to the TRI Reagent™ protocol (Thermo Fisher).

2.4.3. RT-PCR quantification of tRFs

In order to measure the tRF level after transfection, the 15-40 nt small RNA fraction was eluted from a denaturing polyacrylamide gel (see Gebert et al. 2015) and polyadenylated using the A-Plus Poly(A) Polymerase Tailing Kit (Cellscript). After ethanol precipitation the poly-A tailed RNA was reversely transcribed with the SuperScript™ IV reverse transcriptase (Thermo Fisher) using the RT-primer 5'-CGAATTCTAGAGCTCGAGGCAGGCGACATGT25VN-3'. For qPCR 1 µL of cDNA was mixed 0.5 µL 10 µM sequence-specific forward primer, 0.5 µL 10 µM RT-primer-specific reverse primer, 3 µL water and 5 µL 2x QuantiFast® SYBR® Green PCR Master Mix (Qiagen). Technical duplicates of this reaction mix were then analyzed on a Corbett Rotor-Gene 6000 real-time PCR cycler. Finally, the copy numbers of the respective tRFs were quantified by standard curves of the individual primer pair amplicons. As normalizers the miRNAs miR25, miR532 and miR99a were used. The boxplots were created with R using the R packages *ggplot2* and *Cairo* (version 3.4.3). qPCR primer sequences are available in S-Table 2.3.

2.4.4. RT-PCR quantification of miRanda/piRanha predicted targets

Potential target transcripts of the tRFs were predicted using the algorithm miRanda (version 3.3a) that bases on miRNA targeting rules (Enright et al. 2003) and a self-developed software named piRanha (version 0.0.0) that bases on piRNA targeting rules (Zhang et al. 2018a). The reference transcriptome was downloaded from the Ensembl database (release 94). The custom Perl scripts MRscript.pl and PRscript.pl were applied to extract and sort the transcript IDs from the output files by the respective miRanda and piRanha score, which is the sum of binding energies of all alignments with the tRF seed. It is assumed that the lower the miRanda or piRanha score (i.e. free energy of the alignments) of a transcript, the more likely it is a target of the respective tRF. For further analysis, only transcripts that have a TPM value above 0.2 in HEK293 cells were considered. Therefore the RNA sequencing datasets SRR629569 and SRR629570 were downloaded with NCBI's fastq-dump (-I --split-files --gzip) and uploaded to the Galaxy server (usegalaxy.org), where they were mapped to the human genome (Galaxy hg38) using RNA STAR (Galaxy Tool Version 2.6.0b-1). Transcript wise counting was performed with featureCounts (Galaxy Tool Version 1.6.0.6) on basis of an Ensemble GTF-file (Homo_sapiens.GRCh38.90), which had been converted to UCSC coordinates using the File Chameleon tool of Ensembl. The generated count tables and gene length files were used to calculate the mean TPM values and select the expressed transcripts with R using the R packages *biomaRt* and *stats* (version 3.4.3). For each tRF and algorithm, the five transcripts with the lowest miRanda or piRanha scores were chosen for RT-PCR quantification. Primers with the length of ~20 nt were designed to be either exon-junction spanning or to include intronic regions that are bigger than 700 bp. Furthermore, only primers that exclusively amplify the same amplicon from the different splicing isoforms that are potential targets were taken into account. For each primer pair a test PCR with cDNA from untransfected HEK293T cells was performed to check the amplicon quality and length on an agarose gel. In order to compare the relative copy number of the selected potential tRF targets in tRF-mimic and control transfected cells, the respective total RNA was extracted with TRI Reagent™ (Thermo Fisher), reversely transcribed and quantified as described for the poly-A tailed RNA above. The housekeeping genes *ACTB* and *RPS18* were used as normalizers to calculate the relative expression by means of the delta-delta-C_T method. qPCR primer sequences are available in S-Table 2.3. The boxplots were created with R using the R packages *ggplot2* and *Cairo* (version 3.4.3).

2.4.5. RNA sequencing, data processing and differential expression analysis

Total RNA isolated from tRF-mimic, tRF-antisense and control transfected cells was send for library construction and paired-end sequencing to BGI (Hong Kong). On average 35 million paired-end reads were obtained. Using the online platform Galaxy (usegalaxy.org) the reads were first mapped to the human genome (Galaxy hg38) by RNA STAR (Galaxy Tool Version 2.6.0b-1). Afterwards, gene wise counting was performed with featureCounts (Galaxy Tool Version 1.6.0.6) on basis of the above mentioned GTF-file. Based on the generated count tables, DESeq2 (Galaxy Tool Version 2.11.40.2) was used to identify differentially expressed

genes (adjusted p-value < 0.01). Analysis of the *Drosophila* datasets (Study: PRJNA378597, Run accessions: SRR6930617, SRR6930619, SRR6930621; Luo et al. 2018) was likewise conducted using the organism-specific files. Volcano plots based on the DESeq2 result file and cumulative distribution plots based on the DESeq2 result file with thresholds of -80 for miRanda scores and -40 for piRanha scores were created with R using the R packages *ggplot2*, *ggrepel* and *Cairo* (version 3.4.3).

2.4.6. Identification of targeting patterns using a k-mer mapping approach

tRF sequences with length n were split into all possible k-mers with k=5..n. All k-mers were mapped individually to the 5' UTR, the CDS and the 3' UTR of the transcripts of the corresponding organism in reverse complementarity to identify putative target sites. Only the longest transcript per annotated gene was taken into account. The transcriptomes and respective genome annotation files were downloaded from Ensembl database (release 94). Splitting the transcripts into 5' UTR, CDS and 3' UTR was performed using the custom Perl scripts select+split_dmel.pl and select+split_hsap.pl. The following numbers of total mismatches (mm, including insertions/deletions) within a k-mer alignment were allowed: For k≥6 1 mm, for k≥12 2 mm, for k≥18 3 mm, for k≥22 4 mm, for k≥26 5 mm. The numbers of allowed insertions/deletions (indel) within a k-mer alignment were: For k≥12 1 indel, for k≥18 2 indels, for k≥22 3 indels. Mismatches were not allowed in the first two or last two position of the alignment. Nested k-mer alignments, i.e. alignments that completely overlapped with larger k-mer alignments, were ignored. k-mer mapping and filtering was performed using the custom Perl script map_kmers.pl.

Gene expression values (fpkm) were calculated based on the featureCounts count tables and genes with an average expression ratio below 1 fpkm were depleted from the DESeq2 result file for further analysis using the R script DESeq2-Analysis-for-get_targets.R. The custom Perl script get_targets_for_DESeq2.pl was then used to count for each tRF and its k-mers defined by start position and length the number of significantly up-/downregulated genes (adjusted p-value < 0.01) that have a corresponding k-mer alignment, and the number of significantly up-/downregulated genes that do not have a corresponding k-mer alignment. The R script get_targets_visualization.R was used to visualize for each transcript region the fraction of targeted as well as not targeted genes that are up-/downregulated. For the individual plots, the script calculates the average values of a 5 nt sliding window for the start position with sliding window increment of 1 nt.

To check whether the transfected tRFs act on a transcriptional or post-transcriptional level, the adapted Perl script select+split4exonjunction_hsap_EJ.pl was used to first split the transcripts into 5' UTR, CDS and 3' UTR and then mask all nucleotides, but eight nucleotides at the ultimate exon end (EE), the ultimate exon start (ES) or symmetrically situated around the exon-junction (EJ). Subsequently, 5-mers of the respective tRF sequence was mapped to the masked sequence files using the adapted Perl script map_kmers_EJ.pl. Further analysis and visualization was performed as described above.

2.4.7. Analysis of potential targets regarding thermodynamically favored alignments with 5' tRH-Glu-CTC and GO term annotation

Based on the DESeq2 result tables, a Venn diagram of significantly differentially expressed genes (adjusted p-value<0.01) was generated using the R packages *VennDiagram* and *polyclip* (version 3.4.3). Genes that were inversely regulated in the HEK overexpression and the HepG2 inhibition cells were assigned to the groups "potential perish targets" (HEK: log2FC<0; HepG2: log2FC>0) and "potential shelter targets" (HEK: log2FC>0; HepG2: log2FC<0).

Input files for the program RNAup were generated using the custom Perl script RNAup_input.pl, which prints for each potential target gene the sequence of each transcript region (5' UTR, CDS, 3' UTR) together with the tRF sequence as fasta-format. RNAup (version 2.4.13) from the ViennaRNA Package (Lorenz et al. 2011) was executed for each input file with the parameters -b -d2 --noLP -c 'S'. Using the custom Perl script merge_RNAupOutput.pl the RNAup output information was merged per transcript region. Using the R script visualize_RNAup.R the respective merged RNAup output files were visualized as a histogram displaying the alignment length, a histogram displaying the alignment start within the tRF and a bar plot displaying the alignment count per position within the tRF.

The list of human protein-coding genes with the gene ontology term "neurogenesis" (GO:0022008) was downloaded from AmiGO 2 (http://amigo.geneontology.org/amigo). Corresponding Ensembl gene identifier for the UniProt identifiers were retrieved from UniProt (https://www.uniprot.org/uploadlists/). To evaluate a potential statistical correlation between the perish or shelter targets of 5' tRH-Glu-CTC and the GO term "neurogenesis" the chi-squared test was applied.

2.4.8. Code availability and data deposition

All above mentioned custom Perl and R scripts are freely available at GitHub (github.com/jjehn/tRH-targeting). Sequencing datasets are accessible at NCBI's sequence read archive (SRA) under the accession numbers SRR10091207 (1st sRNAseq human hippocampus), SRR10091206 (2nd sRNAseq human hippocampus), SRR10091205 (sRNAseq human cortex), SRR10091204 (sRNAseq human cerebellum), SRR10092006 (sRNAseq female chimpanzee hippocampus), SRR10092005 (sRNAseq male chimpanzee hippocampus), SRR10092004 (sRNAseq macaque hippocampus), SRR10082984 (sRNAseq HEK293T cells), and SRR10085693 to SRR10085704 (RNAseq of the HEK293T and HepG2 experiments).

2.5. Author Contributions

DR and **JJ** designed the study. VE and CH analyzed the published small RNA sequencing datasets of human tissues. LW and IF generated and analyzed the small RNA libraries of the

human hippocampus. RNK dissected the macaque brain and homogenized the hippocampal tissue sample. IF isolated the total RNA from the macaque sample and coordinated the procurement and sequencing of all primate brain samples. **JJ** analyzed the small RNA sequencing datasets of the cell lines, the primate brain samples and the pig, rat and mouse hippocampus. **JJ**, SW, and BO performed the laboratory experiments. **JJ**, JT, and DR analyzed the RNA sequencing data. **JJ** and DR wrote the manuscript.

2.6. Declarations and Acknowledgements

The authors declare to have no conflict of interest.

We thank the Wolfrum and Strand lab for kindly providing HEK and HepG2 cells. We further thank René Ketting, Susanne Strand, Daniel Gebert, Christine Barbara Kiefer, and Lena Mazzariello for helpful comments and discussion. Chet C. Sherwood and Cheryl Stimpson from the National Chimpanzee Brain Resource (www.chimpanzeebrain.org, USA) are acknowledged for their great assistance in preparing chimpanzee RNA samples. The chimpanzee samples were supported by the NIH grant NS092988. This work was supported by the International PhD Program (IPP) coordinated by the Institute of Molecular Biology IMB, Mainz, Germany, funded by the Boehringer Ingelheim Foundation.

2.7. References

Backofen, Rolf; Hess, Wolfgang R. (2010): Computational prediction of sRNAs and their targets in bacteria. In *RNA biology* 7 (1), pp. 33–42. DOI: 10.4161/rna.7.1.10655.

Boland, Andreas; Huntzinger, Eric; Schmidt, Steffen; Izaurralde, Elisa; Weichenrieder, Oliver (2011): Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (26), pp. 10466–10471. DOI: 10.1073/pnas.1103946108.

Boldrini, Maura; Fulmore, Camille A.; Tartt, Alexandria N.; Simeon, Laika R.; Pavlova, Ina; Poposka, Verica et al. (2018): Human Hippocampal Neurogenesis Persists throughout Aging. In *Cell stem cell* 22 (4), 589-599.e5. DOI: 10.1016/j.stem.2018.03.015.

Chen, Qi; Yan, Menghong; Cao, Zhonghong; Li, Xin; Zhang, Yunfang; Shi, Junchao et al. (2016): Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. In *Science (New York, N.Y.)* 351 (6271), pp. 397–400. DOI: 10.1126/science.aad7977.

Deng, Junfang; Ptashkin, Ryan N.; Chen, Yu; Cheng, Zhi; Liu, Guangliang; Phan, Thien et al. (2015): Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism. In *Molecular therapy : the journal of the American Society of Gene Therapy* 23 (10), pp. 1622–1629. DOI: 10.1038/mt.2015.124.

Drees, Frauke; Gertler, Frank B. (2008): Ena/VASP: proteins at the tip of the nervous system. In *Current opinion in neurobiology* 18 (1), pp. 53–59. DOI: 10.1016/j.conb.2008.05.007.

Elbarbary, Reyad A.; Takaku, Hiroaki; Uchiumi, Naoto; Tamiya, Hiroko; Abe, Mayumi; Takahashi, Masayuki et al. (2009): Modulation of gene expression by human cytosolic tRNase Z(L) through 5'-half-tRNA. In *PloS one* 4 (6), e5908. DOI: 10.1371/journal.pone.0005908.

Enright, Anton J.; John, Bino; Gaul, Ulrike; Tuschl, Thomas; Sander, Chris; Marks, Debora S. (2003): MicroRNA targets in Drosophila. In *Genome biology* 5 (1), R1. DOI: 10.1186/gb-2003-5-1-r1.

Eriksson, P. S.; Perfilieva, E.; Björk-Eriksson, T.; Alborn, A. M.; Nordborg, C.; Peterson, D. A.; Gage, F. H. (1998): Neurogenesis in the adult human hippocampus. In *Nature medicine* 4 (11), pp. 1313–1317. DOI: 10.1038/3305.

Fröhlich, Kathrin S.; Vogel, Jörg (2009): Activation of gene expression by small RNA. In *Current opinion in microbiology* 12 (6), pp. 674–682. DOI: 10.1016/j.mib.2009.09.009.

Garcia-Silva, Maria R.; Cabrera-Cabrera, Florencia; das Neves, Roberta Ferreira Cura; Souto-Padrón, Thaís; Souza, Wanderley de; Cayota, Alfonso (2014): Gene expression changes induced by Trypanosoma cruzi shed microvesicles in mammalian host cells: relevance of tRNA-derived halves. In *BioMed research international* 2014, p. 305239. DOI: 10.1155/2014/305239.

Gebert, Daniel; Hewel, Charlotte; Rosenkranz, David (2017): unitas: the universal tool for annotation of small RNAs. In *BMC genomics* 18 (1), p. 644. DOI: 10.1186/s12864-017-4031-9.

Gebert, Daniel; Ketting, René F.; Zischler, Hans; Rosenkranz, David (2015): piRNAs from Pig Testis Provide Evidence for a Conserved Role of the Piwi Pathway in Post-Transcriptional Gene Regulation in Mammals. In *PloS one* 10 (5), e0124860. DOI: 10.1371/journal.pone.0124860.

Gebetsberger, Jennifer; Wyss, Leander; Mleczko, Anna M.; Reuther, Julia; Polacek, Norbert (2017): A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. In *RNA biology* 14 (10), pp. 1364–1373. DOI: 10.1080/15476286.2016.1257470.

Goodarzi, Hani; Liu, Xuhang; Nguyen, Hoang C. B.; Zhang, Steven; Fish, Lisa; Tavazoie, Sohail F. (2015): Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. In *Cell* 161 (4), pp. 790–802. DOI: 10.1016/j.cell.2015.02.053.

Greenway, Matthew J.; Andersen, Peter M.; Russ, Carsten; Ennis, Sean; Cashman, Susan; Donaghy, Colette et al. (2006): ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. In *Nature genetics* 38 (4), p. 411. DOI: 10.1038/ng1742.

Guzzi, Nicola; Cieśla, Maciej; Ngoc, Phuong Cao Thi; Lang, Stefan; Arora, Sonali; Dimitriou, Marios et al. (2018): Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. In *Cell* 173 (5), 1204-1216.e26. DOI: 10.1016/j.cell.2018.03.008.

Haussecker, Dirk; Huang, Yong; Lau, Ashley; Parameswaran, Poornima; Fire, Andrew Z.; Kay, Mark A. (2010): Human tRNA-derived small RNAs in the global regulation of RNA silencing. In *RNA (New York, N.Y.)* 16 (4), pp. 673–695. DOI: 10.1261/rna.2000810.

He, Weiwei; Bai, Ge; Zhou, Huihao; Wei, Na; White, Nicholas M.; Lauer, Janelle et al. (2015): CMT2D neuropathy is linked to the neomorphic binding activity of glycyl-tRNA synthetase. In *Nature* 526 (7575), pp. 710–714. DOI: 10.1038/nature15510.

Huang, Bingqing; Yang, Huipeng; Cheng, Xixi; Wang, Dan; Fu, Shuyu; Shen, Wencui et al. (2017): tRF/miR-1280 Suppresses Stem Cell-like Cells and Metastasis in Colorectal Cancer. In *Cancer research* 77 (12), pp. 3194–3206. DOI: 10.1158/0008-5472.CAN-16-3146.

Ivanov, Pavel; Emara, Mohamed M.; Villen, Judit; Gygi, Steven P.; Anderson, Paul (2011): Angiogenin-induced tRNA fragments inhibit translation initiation. In *Molecular cell* 43 (4), pp. 613–623. DOI: 10.1016/j.molcel.2011.06.022.

Ivanov, Pavel; O'Day, Elizabeth; Emara, Mohamed M.; Wagner, Gerhard; Lieberman, Judy; Anderson, Paul (2014): G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (51), pp. 18201–18206. DOI: 10.1073/pnas.1407361111.

Jehn, Julia; Rosenkranz, David (2019): tRNA-Derived Small RNAs: The Good, the Bad and the Ugly. In *Med One* 4 (4).

Jiang, Hui; Wong, Wing Hung (2008): SeqMap: mapping massive amount of oligonucleotides to the genome. In *Bioinformatics (Oxford, England)* 24 (20), pp. 2395–2396. DOI: 10.1093/bioinformatics/btn429.

Keam, Simon P.; Sobala, Andrew; Have, Sara ten; Hutvagner, Gyorgy (2017): tRNA-Derived RNA Fragments Associate with Human Multisynthetase Complex (MSC) and Modulate Ribosomal Protein Translation. In *Journal of proteome research* 16 (2), pp. 413–420. DOI: 10.1021/acs.jproteome.6b00267.

Kim, Hak Kyun; Fuchs, Gabriele; Wang, Shengchun; Wei, Wei; Zhang, Yue; Park, Hyesuk et al. (2017): A transfer-RNA-derived small RNA regulates ribosome biogenesis. In *Nature* 552 (7683), p. 57. DOI: 10.1038/nature25005.

Krishna, Srikar; Yim, Daniel Gr; Lakshmanan, Vairavan; Tirumalai, Varsha; Koh, Judice Ly; Park, Jung Eun et al. (2019): Dynamic expression of tRNA-derived small RNAs define cellular states. In *EMBO reports* 20 (7), e47789. DOI: 10.15252/embr.201947789.

Kumar, Pankaj; Anaya, Jordan; Mudunuri, Suresh B.; Dutta, Anindya (2014): Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. In *BMC Biol* 12 (1), pp. 12–78. DOI: 10.1186/PREACCEPT-5867533061403216.

Kurosawa, N.; Chen, G. Y.; Kadomatsu, K.; Ikematsu, S.; Sakuma, S.; Muramatsu, T. (2001): Glypican-2 binds to midkine: the role of glypican-2 in neuronal cell adhesion and neurite outgrowth. In *Glycoconjugate journal* 18 (6), pp. 499–507.

Kuscu, Canan; Kumar, Pankaj; Kiran, Manjari; Su, Zhangli; Malik, Asrar; Dutta, Anindya (2018): tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. In *RNA (New York, N.Y.)* 24 (8), pp. 1093–1105. DOI: 10.1261/rna.066126.118.

Lee, Yong Sun; Shibata, Yoshiyuki; Malhotra, Ankit; Dutta, Anindya (2009): A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). In *Genes & Development* 23 (22), pp. 2639–2649. DOI: 10.1101/gad.1837609.

Levitz, Ruth; Chapman, Daphne; Amitsur, Michal; Green, Robert; Snyder, Larry; Kaufmann, Gabriel (1990): The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. In *The EMBO journal* 9 (5), pp. 1383–1389.

Lorenz, Ronny; Bernhart, Stephan H.; Höner Zu Siederdissen, Christian; Tafer, Hakim; Flamm, Christoph; Stadler, Peter F.; Hofacker, Ivo L. (2011): ViennaRNA Package 2.0. In *Algorithms for molecular biology : AMB* 6, p. 26. DOI: 10.1186/1748-7188-6-26.

Luo, Shiqi; He, Feng; Luo, Junjie; Dou, Shengqian; Wang, Yirong; Guo, Annan; Lu, Jian (2018): Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response. In *Nucleic acids research* 46 (10), pp. 5250–5268. DOI: 10.1093/nar/gky189.

Martinez, German; Choudury, Sarah G.; Slotkin, R. Keith (2017): tRNA-derived small RNAs target transposable element transcripts. In *Nucleic acids research* 45 (9), pp. 5142–5152. DOI: 10.1093/nar/gkx103.

Maute, Roy L.; Schneider, Christof; Sumazin, Pavel; Holmes, Antony; Califano, Andrea; Basso, Katia; Dalla-Favera, Riccardo (2013): tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (4), pp. 1404–1409. DOI: 10.1073/pnas.1206761110.

Mineta, Katsuhiko; Nakazawa, Masumi; Cebria, Francesc; Ikeo, Kazuho; Agata, Kiyokazu; Gojobori, Takashi (2003): Origin and evolutionary process of the CNS elucidated by comparative genomics analysis of planarian ESTs. In *Proceedings of the National Academy of Sciences of the United States of America* 100 (13), pp. 7666–7671. DOI: 10.1073/pnas.1332513100.

Patten, Brooke A.; Sardi, S. Pablo; Koirala, Samir; Nakafuku, Masato; Corfas, Gabriel (2006): Notch1 signaling regulates radial glia differentiation through multiple transcriptional mechanisms. In *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26 (12), pp. 3102–3108. DOI: 10.1523/JNEUROSCI.4829-05.2006.

Qiao, Huimin; Li, Yanxin; Feng, Chao; Duo, Shuguang; Ji, Fen; Jiao, Jianwei (2018): Nap1l1 Controls Embryonic Neural Progenitor Cell Proliferation and Differentiation in the Developing Brain. In *Cell reports* 22 (9), pp. 2279–2293. DOI: 10.1016/j.celrep.2018.02.019.

Ruggero, Katia; Guffanti, Alessandro; Corradin, Alberto; Sharma, Varun Kumar; Bellis, Gianluca de; Corti, Giorgio et al. (2014): Small noncoding RNAs in cells transformed by human T-cell leukemia virus type 1: a role for a tRNA fragment as a primer for reverse transcriptase. In *Journal of virology* 88 (7), pp. 3612–3622. DOI: 10.1128/JVI.02823-13.

Saikia, Mridusmita; Jobava, Raul; Parisien, Marc; Putnam, Andrea; Krokowski, Dawid; Gao, Xing-Huang et al. (2014): Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress. In *Molecular and cellular biology* 34 (13), pp. 2450–2463. DOI: 10.1128/MCB.00136-14.

Sarker, Gitalee; Sun, Wenfei; Rosenkranz, David; Pelczar, Pawel; Opitz, Lennart; Efthymiou, Vissarion et al. (2019): Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. In *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1820810116.

Schorn, Andrea J.; Gutbrod, Michael J.; LeBlanc, Chantal; Martienssen, Rob (2017): LTR-Retrotransposon Control by tRNA-Derived Small RNAs. In *Cell* 170 (1), 61-71.e11. DOI: 10.1016/j.cell.2017.06.013.

Sharma, Upasna; Conine, Colin C.; Shea, Jeremy M.; Boskovic, Ana; Derr, Alan G.; Bing, Xin Y. et al. (2016): Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. In *Science (New York, N.Y.)* 351 (6271), pp. 391–396. DOI: 10.1126/science.aad6780.

Short, A. K.; Yeshurun, S.; Powell, R.; Perreau, V. M.; Fox, A.; Kim, J. H. et al. (2017): Exercise alters mouse sperm small noncoding RNAs and induces a transgenerational modification of male offspring conditioned fear and anxiety. In *Translational psychiatry* 7, e1114 EP -. DOI: 10.1038/tp.2017.82.

Torres, Adrian Gabriel; Reina, Oscar; Stephan-Otto Attolini, Camille; Ribas de Pouplana, Lluís (2019): Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. In *Proceedings of the National Academy of Sciences of the United States of America* 116 (17), pp. 8451–8456. DOI: 10.1073/pnas.1821120116.

Upton, Kyle R.; Gerhardt, Daniel J.; Jesuadian, J. Samuel; Richardson, Sandra R.; Sánchez-Luque, Francisco J.; Bodea, Gabriela O. et al. (2015): Ubiquitous L1 mosaicism in hippocampal neurons. In *Cell* 161 (2), pp. 228–239. DOI: 10.1016/j.cell.2015.03.026.

Wang, Qingrong; Lee, Inhan; Ren, Junping; Ajay, Subramanian Shankar; Lee, Yong Sun; Bao, Xiaoyong (2013): Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. In *Molecular therapy : the journal of the American Society of Gene Therapy* 21 (2), pp. 368–379. DOI: 10.1038/mt.2012.237.

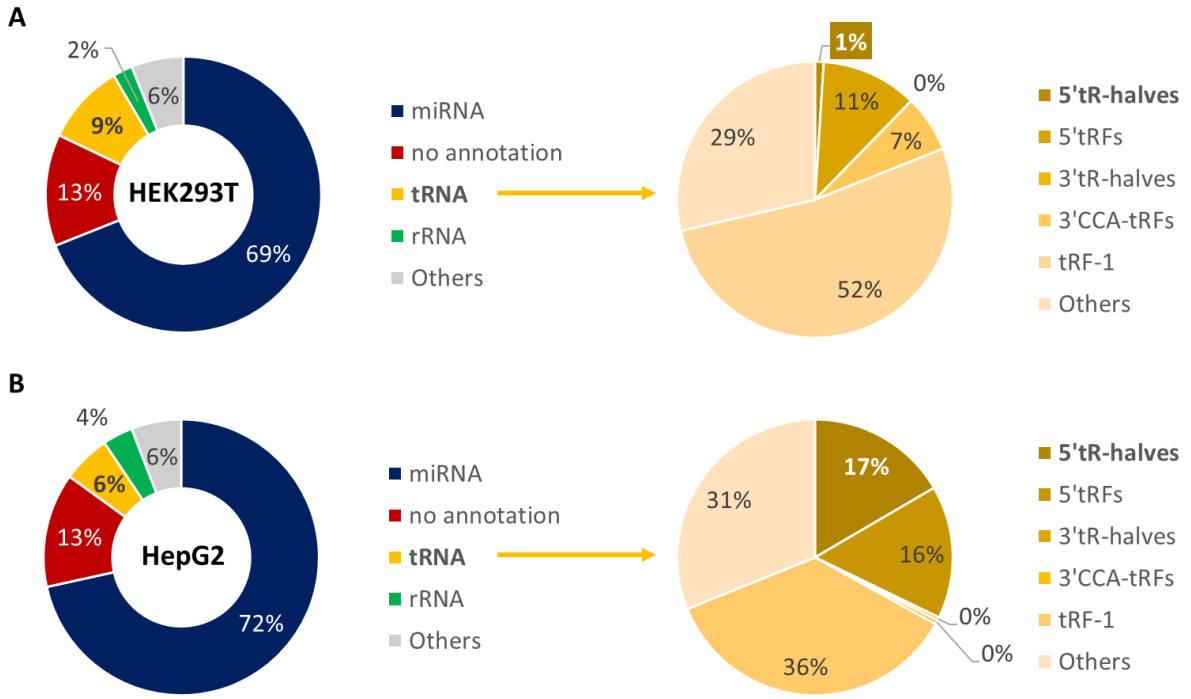
Yeung, Man Lung; Bennasser, Yamina; Watashi, Koichi; Le, Shu-Yun; Houzet, Laurent; Jeang, Kuan-Teh (2009): Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. In *Nucleic acids research* 37 (19), pp. 6575–6586. DOI: 10.1093/nar/gkp707.

Zhang, Donglei; Tu, Shikui; Stubna, Michael; Wu, Wei-Sheng; Huang, Wei-Che; Weng, Zhiping; Lee, Heng-Chi (2018a): The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. In *Science (New York, N.Y.)* 359 (6375), pp. 587–592. DOI: 10.1126/science.aao2840.

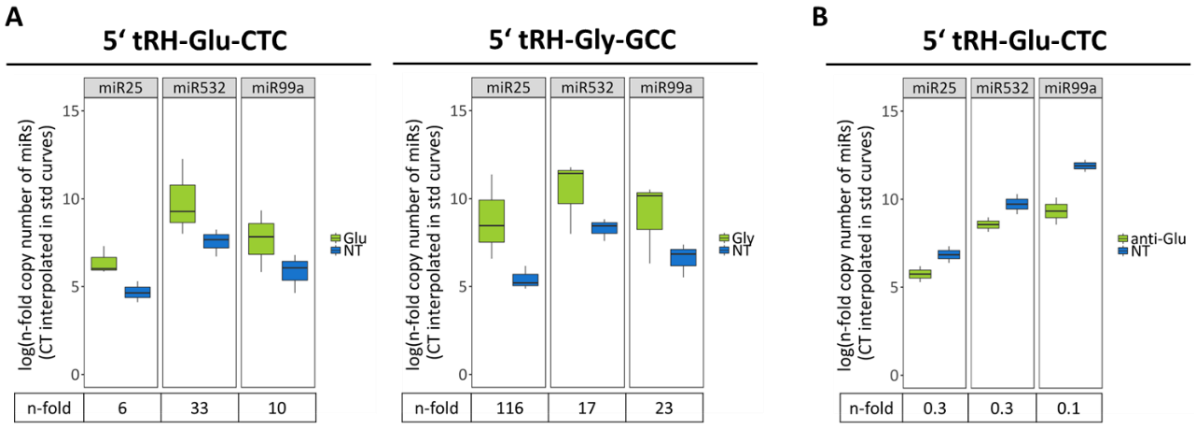
Zhang, Xue; He, Xin; Liu, Chao; Liu, Jun; Hu, Qifei; Pan, Ting et al. (2016): IL-4 Inhibits the Biogenesis of an Epigenetically Suppressive PIWI-Interacting RNA To Upregulate CD1a Molecules on Monocytes/Dendritic Cells. In *Journal of immunology (Baltimore, Md. : 1950)* 196 (4), pp. 1591–1603. DOI: 10.4049/jimmunol.1500805.

Zhang, Yunfang; Zhang, Xudong; Shi, Junchao; Tuorto, Francesca; Li, Xin; Liu, Yusheng et al. (2018b): Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. In *Nature cell biology* 20 (5), p. 535. DOI: 10.1038/s41556-018-0087-2.

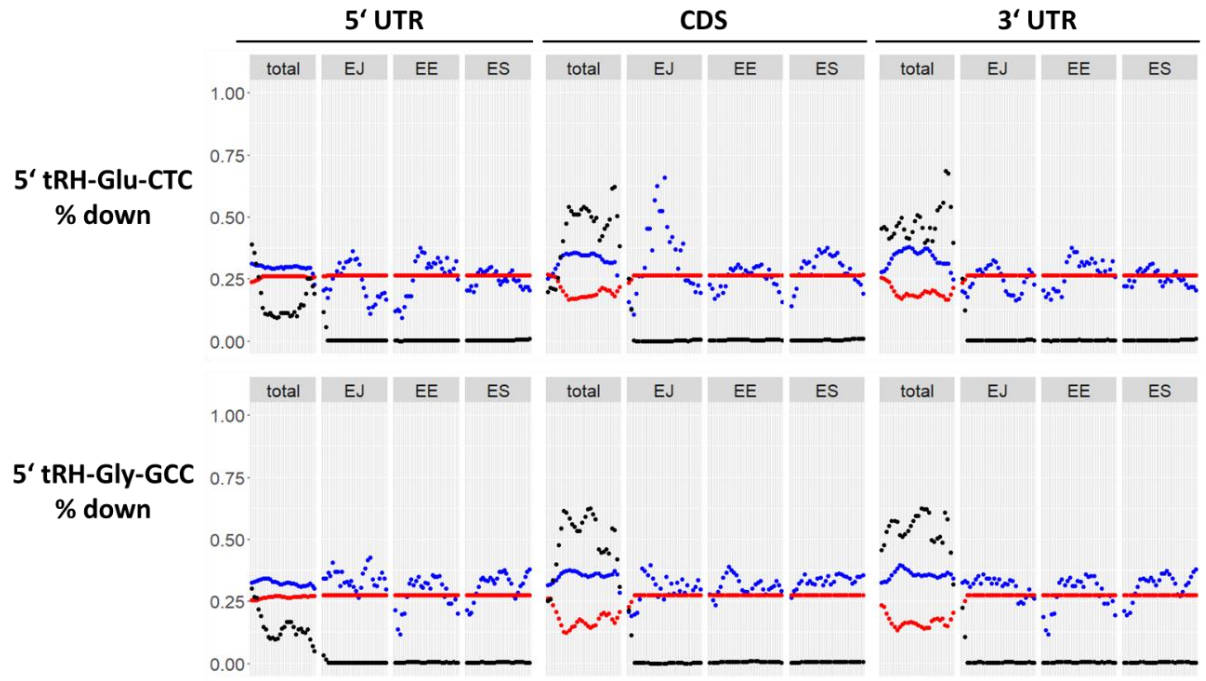
2.8. Supplement



S-Figure 2.1: Small RNA annotation of mapped reads from small RNA libraries of (A) HEK293T cells and (B) HepG2 cells.



S-Figure 2.2: (A) qPCR quantification of the 5' tRNA-halves Glu-CTC (left) and Gly-GCC (right) in HEK293T cells that were transfected with the synthetic 5’ tRH-mimics (green) or a control non-target siRNA (blue). The three miRNAs miR25, miR532 and miR99a were used as normalizers. The given n-fold change is the ratio between the relative 5’ tRH expression in the overexpression and the control cells. (B) qPCR quantification of the 5' tRNA-half Glu-CTC in HepG2 cells that were transfected with an antisense RNA targeting this 5’ tRH (green) or a control non-target siRNA (blue). The three miRNAs miR25, miR532 and miR99a were used as normalizers. The given n fold change is the ratio between the relative 5’ tRH-Glu-CTC expression in the antisense transfected and the control cells.



S-Figure 2.3: Analysis of 5-mer mapping spanning (EJ) or being adjacent to (EE and ES) exon junctions. Displayed is the percentage of transcripts with (blue) or without (red) 5-mer alignment that are downregulated in HEK293T cells upon 5' tRH mimic transfection.

```
human 5' tRh-Glu-CTC     TCCCTGGTGGTCTAGTGGTTAGGATTCGGCGCT
fly   5' tRF-Glu-CTC     T-----ATTGTCTAGTGGTTAGGAT--------
                         *      * ****************
```

S-Figure 2.4: Sequence alignment between human and fly 5' tRF Glu-CTC.

S-Table 2.1: SRA Run Accession numbers and respective tissue of analyzed small RNA sequencing datasets.

Tissue	SRA Run Accession#
brain (PFC)	SRR1635903
brain (cerebellum)	SRR1635904
brain (cerebellum)	SRR553573
fibroblast	SRR4235732
fibroblast	SRR4235731
heart	SRR4421857
heart	SRR4422132
heart	SRR553574
heart	SRR4422133
kidney	SRR1635906
kidney	SRR553575
liver	SRR1273998
liver	SRR1273999

Tissue	SRA Run Accession#
liver	SRR1274001
liver	SRR4422422
liver	SRR4422421
lung	SRR1240787
lung	SRR1240796
lung	SRR1240797
muscle	SRR1820680
muscle	SRR1820682
muscle	SRR1820684
muscle	SRR1820686
muscle	SRR1820688
muscle	SRR1820690
ovary	SRR4422259
ovary	SRR4422260
pancreas	SRR4421754
pancreas	SRR4422182
pancreas	SRR4422183
pancreas	SRR4421755
prostate	SRR4421462
prostate	SRR4421463
prostate	SRR4421707
prostate	SRR4421708
skin	SRR4421488
skin	SRR4421489
skin	SRR4422314
skin	SRR4422315
testis	SRR4422669
testis	SRR4422670
testis	SRR553576
testis	SRR4422476
testis	SRR4422477
thyroid	SRR4421964
thyroid	SRR4421965
uterus	SRR4421536
uterus	SRR4421537
uterus	SRR4421943
uterus	SRR4421944

S-Table 2.2: Sequences of transfected RNAs.

RNA name	RNA type	Sequence from 5’ to 3’
5’ tRH-Glu-CTC	RNA	UCCCUGGUGGUCUAGUGGUUAGGAUUCGGCGCU
5’ tRH-Gly-GCC	RNA	GCAUUGGUGGUUCAGUGGUAGAAUUCUCGCCU
anti-5' tRH-Glu-CTC	2’-OMe-RNA	AGCGCCGAAUCCUAACCACUAGACCACCAGGGA

S-Table 2.3: Sequences of qPCR primers.

Primer name	Sequence from 5’ to 3’	cDNA type
5’ tRH-Glu-CTC-fwd	TCCCTGGTGGTCTAGTGGTTAG	Polyadenylated small RNAs
5’ tRH-Gly-GCC-fwd	GCATTGGTGGTTCAGTGGTAG	
miR25-fwd	CATTGCACTTGTCTCGGTCTG	
miR-532-fwd	CATGCCTTGAGTGTAGGACCGT	
miR99a-fwd	AACCCGTAGATCCGATCTTGT	
PCR-against-RT-PolyT-rev	CGAATTCTAGAGCTCGAGGCAGG	
KIAA1109_fwd	ATCATTTTTCGGTGGTGGAA	Polyadenylated transcripts
KIAA1109_rev	CAACTCTTGAAGGCGTCCAT	
VPS13D_fwd	CTTACAGGGCAGCATTGGGA	
VPS13D_rev	CTGCCTGGAAACGCTGAGTA	
SYNE1_fwd	TTTGGAGGCCTGGATAGTGG	
SYNE1_rev	AGATCAGAGACCAATGGCGG	
FAT1_fwd	CGAGGCATTTGATCCAGATT	
FAT1_rev	TCGGTCTAGCTTCCTTGACG	
JAG1_fwd	CGATGAATGTGCCAGCAACC	
JAG1_rev	CCTTCAGGTGTGTCGTTGGA	
RIF1_fwd	TTCTGGAATGCCACTTTTGC	
RIF1_rev	CCACTGGATTCCTCCATCAT	
HDAC4_fwd	GCAGCACATGGTCTTACTGG	
HDAC4_rev	CTGGAACTGCTGCTTGTGTT	
IGSF8_fwd	GAAGGTGGCATCCAGAACAT	
IGSF8_rev	GGTACACTGTGCCTCCTGCT	
HEG1_fwd	AGGAGCGGCTCTTCAAGTAG	
HEG1_rev	TGGATGGCAGGTGAAGACTT	
RUNX2_fwd	CTGTGGTTACTGTCATGGCG	
RUNX2_rev	AGGTAGCTACTTGGGGAGGA	
OBSCN_fwd	AGGGCCGAAAATACATCCTG	
OBSCN_rev	GACCACATCATACTTCTGGC	
HSPG2_fwd	ACTTCATCTCCTTCGGCCTC	
HSPG2_rev	TTCCTCGTTCAGATCCAGGC	
KLC1_fwd	TGTTCACAAACAGAGGGTGG	
KLC1_rev	GCTGCTGTCGTTTTCCACAA	
LAMA5_fwd	CAGTACTGTGACATCTGCAC	
LAMA5_rev	GGCAAACTTGATGAGGACGT	

Primer name	Sequence from 5’ to 3’	cDNA type
PLEC_fwd	GAGAAGGTCTTGGCCCTACC	Polyadenylated transcripts
PLEC_rev	ACCAGGCTGATGGTCTTGAG	
COL4A3_fwd	TGGCCAGAAAGGATTCACAG	
COL4A3_rev	GTACACCGACAAGTCCGTAA	
EMILIN3_fwd	CCTCCCGCTACAGTCTCTAC	
EMILIN3_rev	CCCATCTACACTGCCGGTAT	
TMEM69_fwd	GCCTAGGAACCAATTAGCGC	
TMEM69_rev	CTTCTGGATGAAGCGAAGCA	
INAVA_fwd	GTGTCCGAGGAGCTCAAGT	
INAVA_rev	GATCTCCGCAGCACACAAAC	
TMEM8A_fwd	CTGTGCATCCTCAGCTACGA	
TMEM8A_rev	TGGAGGCCATGATCACGAAG	
ACTB_fwd	CGAGCACAGAGCCTCGCCTTT	
ACTB_rev	CATGCCCACCATCACGCCCTGG	
RPS18_fwd	GCGGCGGAAAATAGCCTTTG	
RPS18_rev	GGATCTTGTACTGGCGTGGA	

3. PIWI genes and piRNAs are ubiquitously expressed in somatic tissues of mollusks

Julia Jehn[1*], Daniel Gebert[1*], Frank Pipilescu[1*], Sarah Stern[1], Julian S. T. Kiefer[1], Charlotte Hewel[1], David Rosenkranz[1]

[1]Institute of Organismic and Molecular Evolution, Anthropology, Johannes Gutenberg University Mainz, Anselm-Franz-von-Bentzel-Weg 7, 55099 Mainz, Germany

[*]These authors contributed equally to this work

This chapter was published as a Research Article in Communications Biology under the title "PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation" (Jehn et al., Communications Biology 2018 1:137). Numbering of figures and tables as well as the citation style have been adjusted.

3.1. Abstract

PIWI proteins and PIWI-interacting RNAs (piRNAs) suppress transposon activity in animals, thus protecting their genomes from detrimental insertion mutagenesis. Here, we reveal that *PIWI* genes and piRNAs are ubiquitously expressed in mollusks, similar to the situation in arthropods. We describe lineage specific adaptations of transposon composition in piRNA clusters in the great pond snail and the pacific oyster, likely reflecting differential transposon activity in gastropods and bivalves. We further show that different piRNA clusters with unique transposon composition are dynamically expressed during oyster development. Finally, bioinformatics analyses suggest that different populations of piRNAs presumably bound to different *PIWI* paralogs participate in homotypic and heterotypic ping-pong amplification loops in a tissue- and sex specific manner. Together with recent findings from other animal species, our results support the idea that somatic piRNA expression represents the ancestral state in metazoans.

3.2. Introduction

In virtually all animals, PIWI proteins protect germ cells from the steady threat of mobile genetic elements, so-called transposons (Thomson and Lin 2009; Iwasaki et al. 2015). Based on sequence complementarity to their target transcripts, 23-31 nt non-coding RNAs, termed PIWI-interacting (pi-) RNAs, function as guide molecules for PIWI proteins that slice matching targets through their endonuclease activity. Besides post-transcriptional transposon control, PIWI proteins and piRNAs can trigger the establishment of repressive epigenetic DNA or chromatin modifications, thus inducing efficient transposon silencing on the transcriptional level (Reuter et al. 2011; Di Giacomo et al. 2013; Pezic et al. 2014; Manakov et al. 2015).

Analyses of piRNA pathways in representatives of many animal taxa have unveiled a great diversity of lineage specific adaptations, challenging the universal validity of insights obtained from model organisms (Grimson et al. 2008; Houwing et al. 2008; Das et al. 2008; Li et al. 2013; Lim et al. 2014; Ha et al. 2014; Hirano et al. 2014; Gebert et al. 2015; Roovers et al. 2015; Rosenkranz et al. 2015b; Madison-Villar et al. 2016; Praher et al. 2017; Lewis et al. 2018). For a long time, PIWI proteins and piRNAs were thought to be dispensable for female germ cell development in mammals until it became clear that the model organisms mouse and rat represent an exception from the mammalian rule in that they employ an oocyte specific Dicer isoform for transposon control instead of Piwil3 which is expressed in the bovine and human female germline (Roovers et al. 2015; Flemr et al. 2013). Similarly, evidence for a gene regulatory role of piRNAs (Gebert et al. 2015; Zhang et al. 2015; Russell et al. 2017; Rouget et al. 2010; Gou et al. 2014; Watanabe and Lin 2014; Barckmann et al. 2015; Rojas-Ríos et al. 2017) and their widespread somatic expression in many animals (Lewis et al. 2018; Palakodeti et al. 2008; Perrat et al. 2013; Nandi et al. 2016; Jones et al. 2016; Teixeira et al. 2017; Ross et al. 2014; Juliano et al. 2014; Funayama et al. 2010) have eroded the dogma that the piRNA pathway is restricted to the germline, being exclusively responsible for silencing of transposons. Indeed, it has been shown that piRNAs are essential for regeneration and stem cell maintenance in the flatworm *Schmidtea mediterranea* (Palakodeti et al. 2008), provide an adaptive immunity against virus infections in *Aedes aegypti* (Miesen et al. 2015), are responsible for sex determination in *Bombyx mori* (Kiuchi et al. 2014) and memory-related synaptic plasticity in *Aplysia californica* (Rajasethupathy et al. 2012).

Despite the likely more than seventy thousand living molluskan species (Rosenberg 2014) there exist only a few functional descriptions of PIWI proteins or piRNAs for this taxon based on experiments in the sea slug *Aplysia californica* (Rajasethupathy et al. 2012), the Farrer's scallop *Chlamys farreri* (Ma et al. 2017) and in the dog whelk *Nucella lapillus* (Waldron et al. 2018). Importantly, Waldron and coworkers recently showed that piRNA-like small RNAs matching virus and transposon sequences are somatically expressed in *Nucella lapillus*. However, the available data does not allow to draw any conclusions on whether this represents a conserved or lineage-specific feature of the PIWI/piRNA system within mollusks. In order to further elucidate the evolution of the PIWI/piRNA system in mollusks, we have reconstructed the evolution of *PIWI* genes in this phylum based on 11 sequenced genomes showing that *Piwil1* and *Piwil2* are conserved in mollusks. We perform quantitative real-time PCR experiments to analyze the expression patterns of the identified *PIWI* paralogs across a representative set of tissues from the great pond snail *Lymnaea stagnalis* (*L. stagnalis*) and the pacific oyster *Crassostrea gigas* (*C. gigas*). We apply high-throughput sequencing of small RNAs from *L. stagnalis* to verify the presence of piRNAs in germline and muscle tissue. We further reanalyze published small RNA sequence data from *C. gigas* to characterize the dynamic expression of piRNAs from distinct piRNA clusters during oyster development. Finally, we use bioinformatics approaches to show that different piRNA populations and *PIWI* paralogs participate in the ping-pong amplification loop in a tissue- and sex specific manner.

3.3. Results

3.3.1. The molluskan PIWI gene repertoire

Many *PIWI* gene tree reconstructions have been published in the past years, however they do not provide a coherent picture regarding the evolution of *PIWI* genes in early bilaterians. Thus, we first wanted to characterize the PIWI protein equipment of sequenced mollusks to infer the ancestral molluskan state and subsequent evolution of *PIWI* paralogs within the molluskan clade. To this end, we used available PIWI protein sequence data from six molluskan species (*Biomphalaria glabrata*, *Aplysia californica*, *Crassostrea gigas*, *Crassostrea virginica*, *Mizuhopecten yessoensis*, *Octopus bimaculoides*) and further manually annotated *PIWI* genes based on five publicly available but not yet (sufficiently) annotated genomes (*Lymnaea stagnalis*, *Radix auricularia*, *Lottia gigantea*, *Bathymodiolus platifrons*, *Pinctada martensii*). We found that the PIWI family members *Piwil1* and *Piwil2* are conserved in mollusks and are orthologous to *Piwil1* and *Piwil2* in vertebrates, suggesting a duplication event in an early bilaterian ancestor prior to the split of protostomes and deuterostomes. According to our results and in consistency with a number of previously published gene trees, *Drosophila* AGO3 shares a common ancestral gene with *Piwil2* clade members (Praher et al. 2017; Zhou et al. 2007; Schurko et al. 2009; Lewis et al. 2016). However, the insect-specific *PIWI* genes *Piwi* and *Aubergine*, the latter one resulting from a duplication event in dipteran flies (Lewis et al. 2016; Kerner et al. 2011), do not group with the *Piwil1* clade (figure 3.1A). It is worth mentioning in this context that different rates of sequence evolution, selective regimes and gene turnover for Argonaute subfamilies make it difficult to infer their ancient evolutionary history, which is mirrored by numerous published but contradicting *PIWI* gene trees, none of which correctly mirrors the phylogenetic relationship of the included species. Consequently, the presented gene tree reconstruction aims to provide a reliable reconstruction of molluskan *PIWI* gene evolution while the deeper topology should be considered with caution.

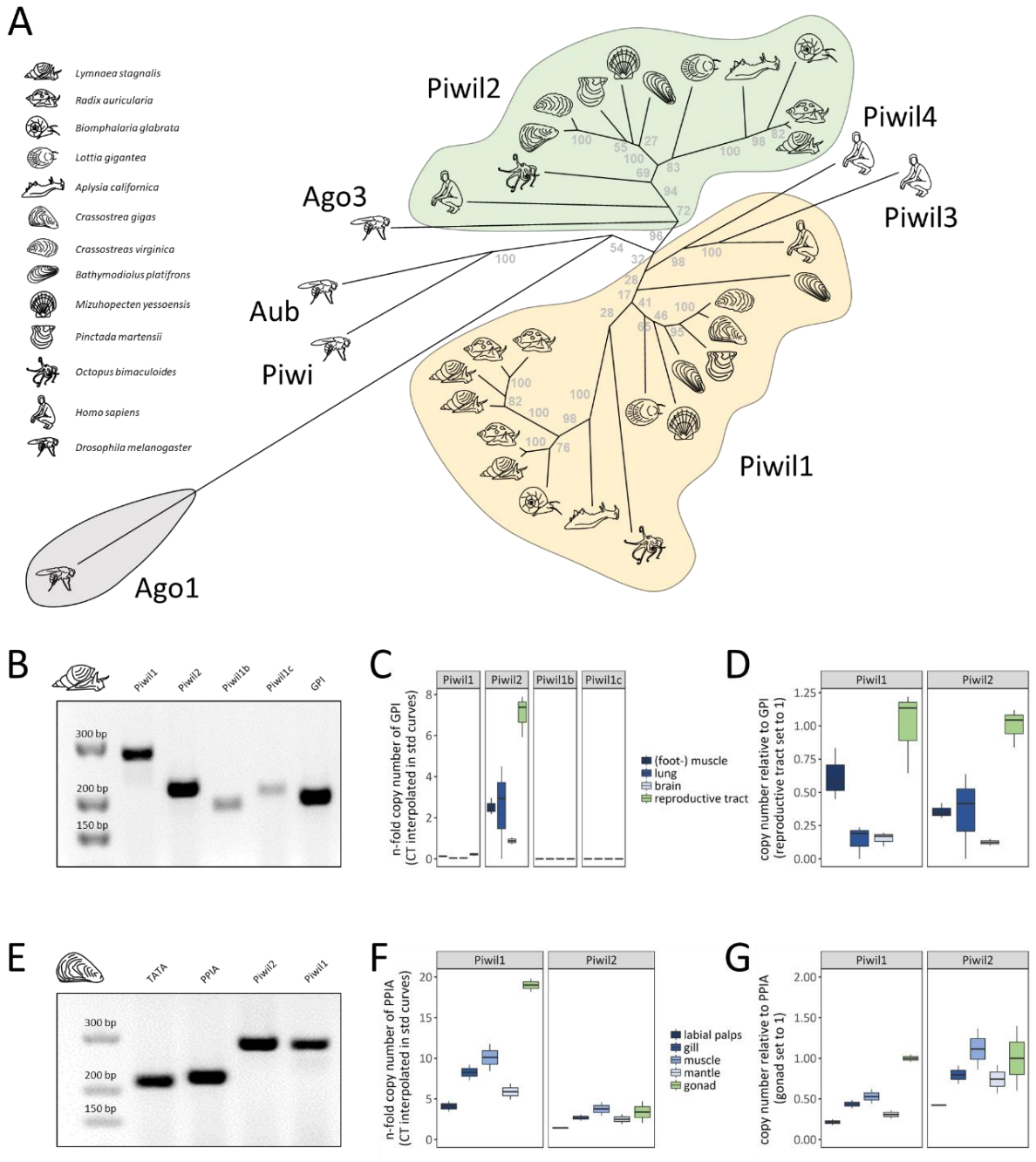


Figure 3.1: Evolution and expression of PIWI genes in mollusks. (A) PIWI gene tree reconstruction of molluskan PIWI genes. (B) Control PCR with PIWI paralog specific primers and L. stagnalis cDNA from the reproductive tract. The complete gel is shown in electronic supplemental figure 1g (C) RT-qPCR results for PIWI paralog expression in different tissues of L. stagnalis, measured as n-fold expression of the housekeeping gene GPI. Center line indicates median, box limits represent the 50th percentile, whiskers show the upper and lower extremes. (D) PIWI paralog expression in different tissues of L. stagnalis, normalized by the expression of the housekeeping gene GPI, values from reproductive tract set to 1. Center line indicates median, box limits represent the 50th percentile, whiskers show the upper and lower extremes. (E) Control PCR with PIWI paralog specific primers and C. gigas cDNA from the adductor muscle. The complete gel is shown in electronic supplemental figure 1h (F) RT-qPCR results for PIWI paralog expression in different tissues of C. gigas, measured as n-fold expression of the housekeeping gene PPIA. Center line indicates median, box limits represent the 50th percentile, whiskers show the upper and lower extremes. (G) PIWI paralog expression in different tissues of C. gigas, normalized by the expression of the housekeeping gene PPIA, values from male gonad set to 1. Center line indicates median, box limits represent the 50th percentile, whiskers show the upper and lower extremes.

While we did not observe further gene duplication events within the molluskan *Piwil2* clade, several duplication events are present in the *Piwil1* clade resulting in two *Piwil1* paralogs in *Bathymodiolus platifrons* and even three *Piwil1* paralogs in *Lymnaea stagnalis* and *Radix auricularia*. Generally, *PIWI* gene duplication events are in line with the previously described erratic evolution of *PIWI* family genes in arthropods (Lewis et al. 2018; Lewis et al. 2016; Kerner et al. 2011; Dowling et al. 2016). Noteworthily, it was also a successive duplication of *Piwil1* on the eutherian lineage that gave rise to *Piwil3* (with subsequent loss on the murine lineage) and *Piwil4* (Sasaki et al. 2003; Murchison et al. 2008) (figure 3.1A).

3.3.2. Expression of PIWI genes in *L. stagnalis* and *C. gigas*

To investigate the expression of *PIWI* genes in mollusks we chose two representative species, the pacific oyster *Crassostrea gigas* (*C. gigas*, Bivalvia) showing no *Piwil1* duplication, and the great pond snail *Lymnaea stagnalis* (*L. stagnalis*, Gastropoda), featuring three predicted *Piwil1* paralogs (figure 3.1A). We performed quantitative real-time PCR (qPCR) for each *PIWI* paralog on a representative set of tissues from both species.

For the great pond snail *L. stagnalis* we measured *PIWI* expression on the mRNA level in the hermaphroditic reproductive tract, comprising both male and female gametes, foot muscle, lung and brain. Relevant expression was detectable for *Piwil1* and particularly *Piwil2*, while the *Piwil1* duplicates *Piwil1b* and *Piwil1c* were only expressed at very low levels (figure 3.1B+C and electronic supplemental figure 1) suggesting a spatiotemporal sub-functionalization. As expected, we observed the highest expression of *Piwil1* and *Piwil2* in the reproductive tract. However, both genes were significantly expressed in the other analyzed tissues as well, reaching 62%, 21% and 15% of germline expression for *Piwil1* in muscle, lung and brain respectively, and 36%, 53% and 12% of germline expression for *Piwil2* in muscle, lung and brain, respectively (figure 3.1D).

For the dioecious pacific oyster *C. gigas*, *PIWI* mRNA expression was measured in the male gonad, labial palps, gill, adductor muscle and mantle. We detected significant expression of *Piwil1* and *Piwil2* across all analyzed tissues, particularly in gonadal tissue (figure 3.1E+F), confirming data on *Piwil1* expression in the Hong Kong Oyster *Crassostrea honkongensis* (Tong et al. 2015). In relation to gonadal expression, *Piwil1* and *Piwil2* were expressed in levels ranging from 21% (*Piwil1* in labial palps) to 111% (*Piwil2* in adductor muscle, figure 3.1G). The observed expression patterns suggest that a functional PIWI machinery acting in the soma and the germline is conserved in mollusks. Considering the somatic expression of PIWI proteins and piRNAs in many arthropod species (Lewis et al. 2018), it is parsimonious to assume that somatic PIWI/piRNA expression represents the ancestral state that was established in an early protostomian ancestor.

3.3.3. piRNAs in *L. stagnalis* muscle and reproductive tract

In order to characterize molluskan piRNAs, we sequenced small RNA transcriptomes from *L. stagnalis* extracted from the hermaphroditic reproductive tract and (foot-) muscle, since muscle tissue was found to exhibit the highest somatic *PIWI* expression in both *L. stagnalis* and *C. gigas*. Importantly, we want to clarify that we will use the term piRNA *bona fide*, without formal evidence for physical interaction with PIWI proteins but based on the evidence provided in the following.

The sequence read length profiles for both tissues show a maximum for 21 nt RNAs, with a considerable amount of 22 nt RNAs being present in the muscle, but not in the reproductive tract. We further observed a smaller fraction of RNAs in the range of 24-29 nt in both samples (figure 3.2A). Annotation of sRNA sequences with unitas (Gebert et al. 2017) revealed a similar proportion of different sRNA classes in each tissue type, with miRNAs accounting for 47% and 53% of reads in the reproductive tract and muscle, respectively (figure 3.2B, electronic supplemental table 1). Interestingly, we found a substantial difference in the abundance of tRNA fragments (tRFs). In both samples, 21 nt RNAs derived from the 3' end of tRNAs (3' tRFs, particularly from tRNA-Gly-TCC) constitute the vast majority of tRNA fragments. However, the share of 3' tRFs in the reproductive tract is considerably higher compared to muscle (17% and 10%, respectively, electronic supplemental table 1). Recently, 3' tRFs were found to silence Long Terminal Repeat (LTR) retrotransposons in mouse stem cells by targeting their functionally essential and highly conserved primer binding sites (Schorn et al. 2017). The remarkable amount of 3' tRFs in the analyzed samples supports the idea proposed by Schorn and coworkers who assume that this mechanism could be highly conserved across different species, providing an innate immunity against LTR propagation.

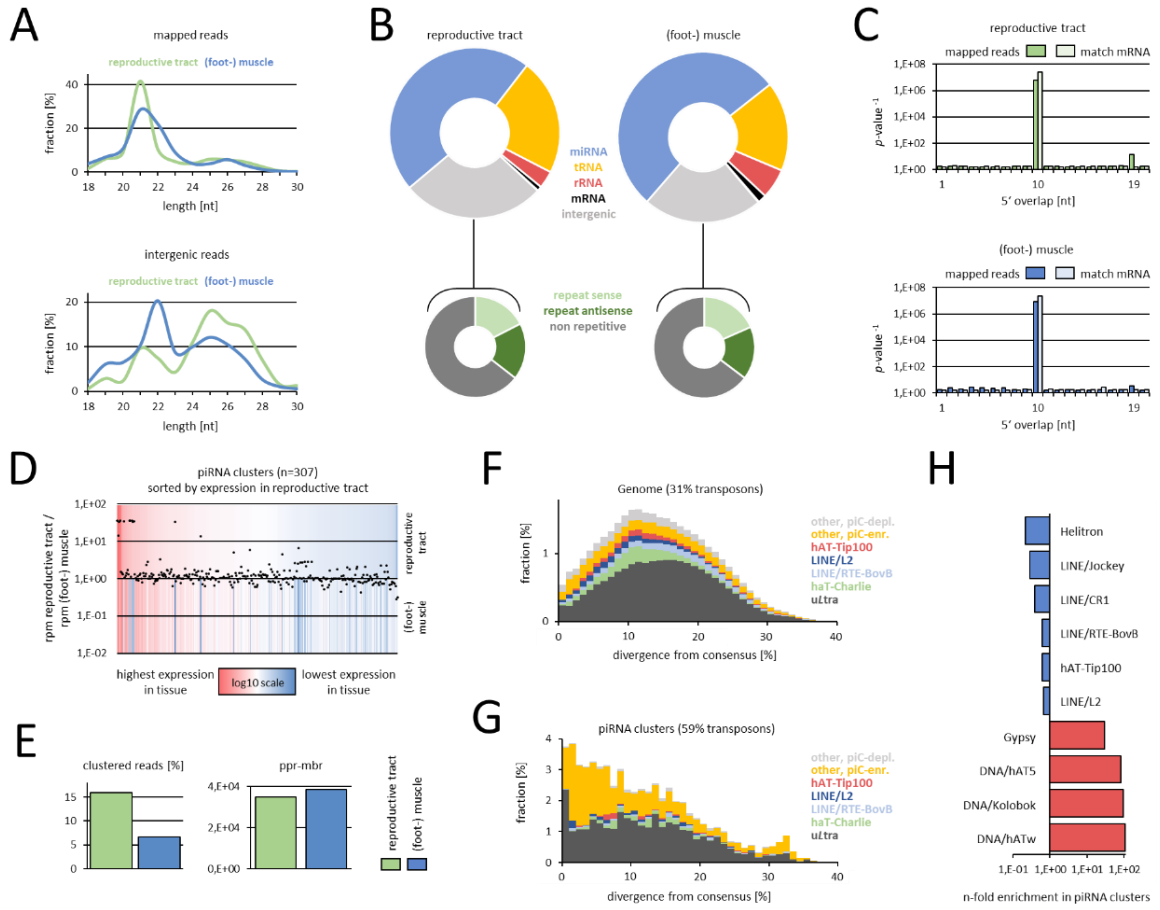


Figure 3.2: Characterization of small RNAs from L. stagnalis (foot-) muscle and reproductive tract. (A) Sequence read length distribution of mapped (top) and unannotated (intergenic) reads (bottom). (B) Results from small RNA annotation with unitas (top) and transposon content of intergenic reads (bottom). (C) Ping-pong signature. P-values are deduced from the corresponding Z-scores. P-values for all reads and reads that match mRNA are shown. (D) Differential expression of 307 predicted piRNA clusters. Colors refer to expression relative to highest/lowest expression within one tissue. Dots indicate n-fold expression of a given cluster in reproductive tract relative to muscle. (E) Amount of clustered reads and ping-pong reads per million bootstrapped reads (ppr-mbr). (F) Representation of transposons in the genome of L. stagnalis, plotted by divergence [%] from transposon consensus. (G) Representation of transposons within piRNA clusters of L. stagnalis, plotted by divergence [%] from transposon consensus. (H) Prominent transposons that are enriched or depleted in L. stagnalis piRNA clusters.

Focusing on putative piRNAs, we analyzed the fraction of sequence reads that did not match to any other class of non-coding RNA nor mRNA. This dark matter of intergenic sRNAs comprises 27% and 23% of sequence reads in the reproductive tract and in muscle, respectively, and is enriched for transposon sequences, suggesting a role in transposon control (figure 3.2B). Analyses of their sequence read length distribution revealed a prominent class of 22 nt molecules in muscle and to a lesser extend in the reproductive tract, suggesting that transposon defense in *L. stagnalis* involves 22 nt siRNAs in addition to piRNAs (figure 3.2A). To verify the presence of piRNAs, we checked for the so-called ping-pong signature (bias for 10 bp 5' overlap of mapped sequence reads), which is a hallmark of secondary piRNA

biogenesis and requires the catalytic activity - and thus expression - of PIWI proteins (Czech and Hannon 2016). Remarkably, we detected a significant ping-pong signature in both, the reproductive tract and muscle (figure 3.2C), suggesting active PIWI/piRNA-dependent transposon silencing in the germline and in the soma. In addition, a ping-pong signature can also be observed for sequence reads that match protein coding genes, indicating piRNA-dependent gene regulation (figure 3.2C).

Next, we used proTRAC (Rosenkranz and Zischler 2012) to identify 308 piRNA producing loci in the reproductive tract, and 246 piRNA producing loci in muscle tissue. Merging of independently annotated contiguous (<10 kb distance) or overlapping piRNA producing loci revealed a total of 307 distinct piRNA clusters in *L. stagnalis*, covering 0.27% of the genome (figure 3.2D, electronic supplemental data 1). More precisely, all piRNA producing loci identified in muscle tissue correspond to predicted piRNA clusters based on piRNAs from the reproductive tract, which illustrates that piRNAs in muscle originate from the same set of piRNA clusters compared to the reproductive tract. Nonetheless, there exist 12 clusters whose expression is 14- to 36-fold higher in the reproductive tract compared to muscle tissue, while no clusters show muscle-specific expression to a comparable extent. We found that 15.9% of sequence reads from the reproductive tract map to piRNA clusters, while only 6.7% of sequence reads from muscle do so, indicating rather moderate production of primary piRNAs in the soma compared to the germline (figure 3.2E). Besides the presence of primary piRNAs, we found that the number of piRNAs that participate in ping-pong-amplification (measured as ping-pong reads per million bootstrapped reads, ppr-mbr) is slightly higher in muscle (~39k ppr-mbr) compared to the situation in the reproductive tract (~35k ppr-mbr), suggesting higher amounts of secondary piRNAs and emphasizing the functional importance of somatic PIWI/piRNA expression (figure 3.2E). In line with the transposon-suppressive role of piRNAs, the identified piRNA clusters show a 2-fold enrichment for transposon sequences compared to the whole genome situation (59% and 31%, respectively, figure 3.2F+G), whereas only 1.7% of piRNA cluster sequence represents protein coding sequence. Interestingly, the transposon composition in piRNA clusters does not at all reflect the transposon landscape of the genome. Instead, piRNA clusters are enriched for Gypsy retrotransposons and particularly DNA transposons such as Kolobok, hAT5 or hATw showing up to 108-fold enrichment in piRNA clusters (figure 3.2G+H). This non-random distribution suggests a selective regime that favors insertion events of transposons with low divergence from their consensus sequence, likely representing evolutionary young and active elements.

3.3.4. Ubiquitous and dynamic expression of piRNAs in *C. gigas*

Based on our observation that *PIWI* genes and piRNAs are expressed in the soma and the germline of *L. stagnalis*, we reanalyzed previously published small RNA datasets from *C. gigas* that were used to investigate the dynamic expression of miRNAs during oyster development without further examination of a putative piRNA fraction (Xu et al. 2014) (NCBI Sequence Read Archive Project ID SRP007591). We annotated *C. gigas* sRNAs from the male and female

gonad, different developmental stages ranging from the egg to juvenile, and a representative set of somatic tissues from adult animals (electronic supplemental table 2). In all datasets, particularly in gonadal tissues, eggs and early embryo stages but also in hemolymph we detected a large amount of sequence reads that did not match to any known ncRNA class but was instead enriched for transposon sequences. The transposon-matching sub-fraction itself was enriched for antisense sequences (electronic supplemental table 2). Analogous to the procedure applied for the *L. stagnalis* datasets, we verified the presence of primary and secondary piRNAs by analyzing the ping-pong signature of each dataset. Remarkably, we detected a significant ping-pong signature across all analyzed datasets (figure 3.3A, electronic supplemental figure 2), but also found that the number of ping-pong reads (measured as ppr-mbr) differs considerably depending on the tissue and developmental stage (figure 3.3A, electronic supplemental figure 3). Noteworthily, as is the case with *L. stagnalis*, a ping-pong signature is also detectable when taking only those reads into account that match protein coding sequences, suggesting a relevant and conserved role of the PIWI/piRNA pathway in post-transcriptional regulation of protein coding genes in gonads, egg, blastula, digestive gland and hemolymph (electronic supplemental table 3). We further used sequences without ncRNA annotation to predict piRNA clusters with proTRAC (electronic supplemental data 2) and checked whether we can observe a differential expression of specific piRNA clusters in time and space (figure 3.3A).

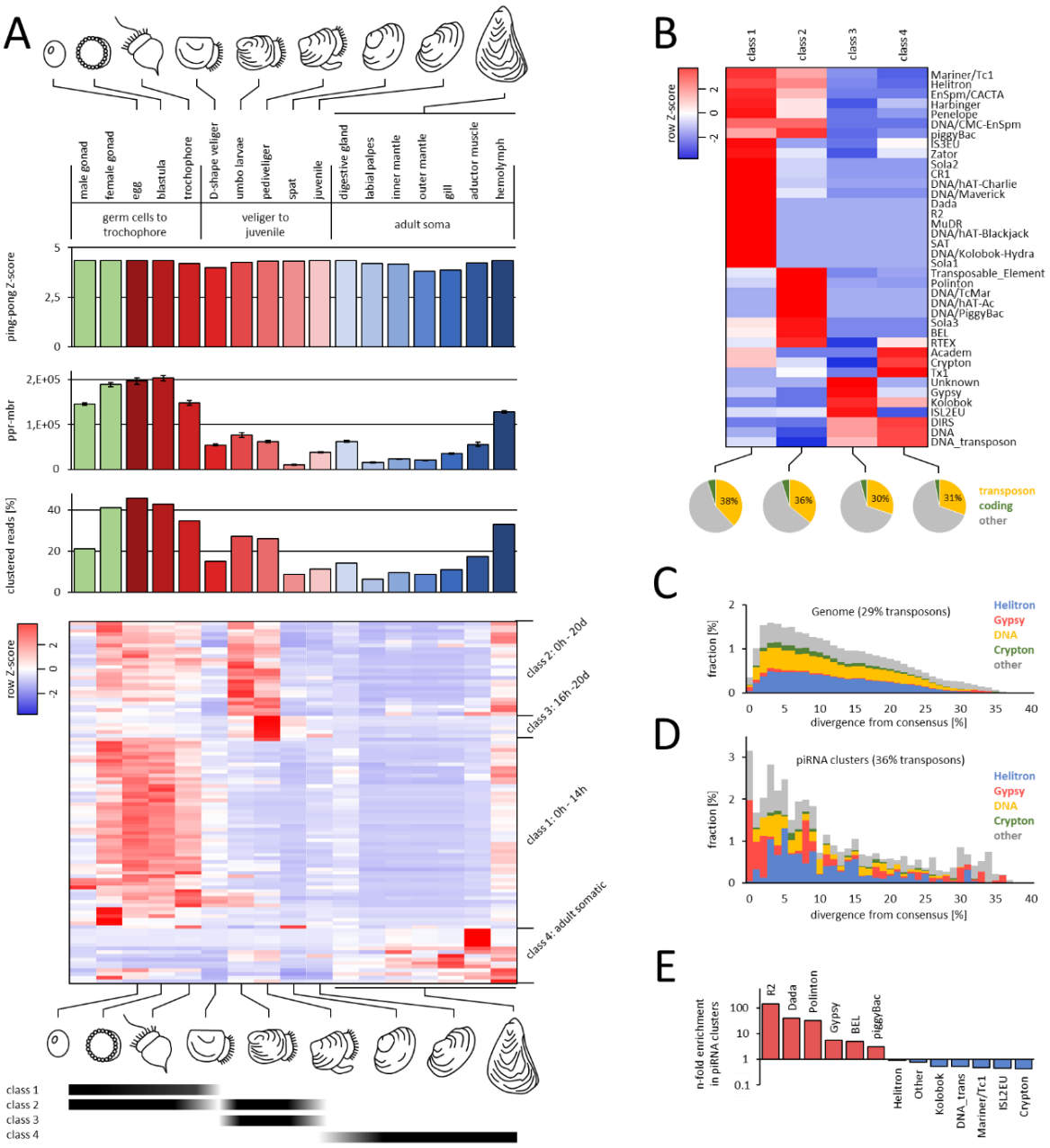


Figure 3.3: Characterization of small RNAs and piRNA clusters from different C. gigas samples. (A) Sequence reads without annotation produce a significant ping-pong signature (top row of bars, only Z-scores for 10 bp 5' overlap are shown). The number of ping-pong reads per million bootstrapped reads (middle row of bars), and the number of clustered reads (bottom row of bars) differs considerably across the samples. Heatmap shows the differential expression of the top 100 piRNA clusters in terms of maximum rpm coverage. Different classes of piRNA clusters are expressed during oyster development and in adult somatic tissues (bottom). (B) Transposon composition of piRNA clusters belonging to four different classes. (C) Representation of transposons in the genome of C. gigas, plotted by divergence [%] from transposon consensus. (D) Representation of transposons within piRNA clusters of C. gigas, plotted by divergence [%] from transposon consensus. (E) Prominent transposons that are enriched or depleted in C. gigas piRNA clusters. Error bars indicate standard deviation.

In contrast to the situation in *L. stagnalis*, we found that different genomic loci are responsible for production of primary piRNAs in the germline and in the soma, but also during different

developmental stages, which is similar to the situation in the sea anemone *Nematostella vectensis* (Praher et al. 2017) and the German cockroach *Blatella germanica* (Llonga et al. 2018). A clustering approach based on average linkage (Babicki et al. 2016) revealed four distinct groups of piRNA clusters which we named class 1-4 piRNA clusters (figure 3.3A). Class 1 piRNA clusters are active in the adult germline (male and female) and in the early embryo until the D-shaped veliger stage where larvae are approximately 14 hours old. The same applies to class 2 piRNA clusters, however, following the D-shape veliger stage, class 1 piRNA clusters become inactive, while class 2 piRNA clusters remain active and class 3 piRNA clusters start piRNA production. Both, class 2 and class 3 piRNA cluster activity is measurable until the juvenile stage, where oysters are approximately 20 days old. In somatic tissues of adult oysters, class 4 piRNA clusters represent the main source of primary piRNAs (figure 3.3A, bottom). Interestingly, all four classes of piRNA clusters are active in hemocytes, which also feature the highest amount of clustered reads, and ping-pong reads compared to other somatic tissues. This might reflect the presence of stem cells within the hemocyte cell population, which are subject to complex differentiation processes (Fisher 1986; Lau et al. 2017).

Interestingly, the four classes of piRNA clusters differ considerably regarding the overall transposon content as well as the specific transposon composition (figure 3.3B-D). Class 1 and class 2 piRNA clusters are generally enriched for transposon sequences showing 38% and 36% transposon derived sequences, respectively, compared to a genomic transposon content of 29%. The surprisingly high accumulation of young (as deduced from the divergence from their consensus) Gypsy elements in piRNA clusters, suggests a strong selection for Gypsy element insertions, probably as a consequence of Gypsy activity in *C. gigas*. Noteworthily, the accumulation of young transposons in molluskan piRNA clusters sharply contrasts the situation in *Drosophila* and human, where older transposons are more abundant in piRNA producing loci (Senti et al. 2015; Gainetdinov et al. 2017) Considering transposons that are generally enriched in piRNA clusters, we found that R2 retrotransposons (149-fold enrichment in piRNA clusters) and Dada DNA transposons (40-fold enrichment in piRNA clusters) are most abundant in class 1 piRNA clusters (figure 3.3E). In contrast, Polinton DNA transposons (32-fold enrichment in piRNA clusters) and BEL retrotransposons (5-fold enrichment in piRNA clusters) are most abundant in class 2 piRNA clusters. Different from class 1 and class 2 piRNA clusters, class 3 and class 4 piRNA clusters display only slight transposon enrichment (30% and 31%, respectively). Noteworthily, high copy number Gypsy retrotransposons (5-fold enrichment in piRNA clusters) are most abundant in class 3 piRNA clusters, while Academ, Crypton and Tx1 transposons are most abundant in class 4 piRNA clusters.

The fact that different piRNA clusters are expressed in the germline (class 1 and class 2) and in adult somatic tissues (class 4) of *C. gigas* contrasts with the situation in *L. stagnalis*, where identical piRNA producing loci are active in the germline and in the soma. Moreover, we can observe considerable differences in the transposon composition of piRNA clusters in the two species, which likely reflect a divergent transposon activity in gastropods and bivalves, resulting in varying selective constraints on the different phylogenetic lineages.

3.3.5. Homotypic and heterotypic ping-pong amplification

The ping-pong amplification loop describes a process that is responsible for the post-transcriptional silencing of transposable elements (Czech and Hannon 2016). In *Drosophila* and mouse, this process typically involves two PIWI paralogs (heterotypic ping-pong), one loaded with antisense piRNAs targeting transposon transcripts, and the other loaded with sense piRNAs targeting piRNA cluster transcripts, which contain transposon sequences in antisense orientation (Brennecke et al. 2007; Aravin et al. 2008). Likely for steric reasons, premature piRNAs loaded onto the different PIWI paralogs are more or less rigorously trimmed at their 3' ends. This is why piRNA populations bound to different PIWI paralogs not only differ regarding the amount of sense- and antisense-transposon sequences, but also in their sequence length profiles (Czech and Hannon 2016; Aravin et al. 2007; Kawaoka et al. 2011). In addition to the heterotypic ping-pong amplification, homotypic ping-pong has been shown to occur in *qin* mutant flies (Aub:Aub, Zhang et al. 2011), and wildtype prenatal mouse testis (Miwi2:Miwi2, Mili:Mili, Aravin et al. 2008).

Since the typical molluskan genome encodes two ubiquitously expressed PIWI paralogs, *Piwil1* and *Piwil2*, we asked whether we can provide evidence for the participation of distinct piRNA populations and PIWI paralogs in the ping-pong cycle. We conducted a bioinformatics approach under the premise that Piwil1- and Piwil2-bound piRNAs exhibit different length profiles, which is the case for the corresponding mouse homologs Piwil1 (Miwi) that preferentially binds 29/30 nt piRNAs, and Piwil2 (Mili) which preferentially binds 26/27 nt piRNAs (Vourekas et al. 2012). A similar, yet not equally pronounced, difference between Piwil1 (Ziwi) and Piwil2 (Zili) -bound piRNAs also exists in zebrafish, suggesting the evolutionary conservation of this pattern (Houwing et al. 2008). We analyzed pairs of mapped *C. giga*s and *L. stagnalis* sequence reads that showed a 10 bp 5' overlap (ping-pong pairs), with respect to the sequence length of each ping-pong partner (figure 3.4, electronic supplemental figure 4). In the female gonad of *C. gigas*, most ping-pong pairs combine piRNAs with a length of 25 nt and 29 nt (figure 3.4A), suggesting heterotypic Piwil1-Piwil2-dependent ping-pong amplification as depicted in figure 3.4B. In support of this, 29 nt piRNAs, presumably bound to Piwil1, are heavily biased for a 5' uridine (a hallmark of primary piRNAs), whereas 25 nt piRNAs, presumably bound to Piwil2, show a stronger bias for an adenine at position 10 (typical for secondary piRNAs). In contrast, ping-pong pairs in *C. gigas* muscle predominantly combine two 29 nt piRNAs, suggesting homotypic, Piwil1-dependent ping-pong amplification (figure 3.4B). Generally, the observed patterns of ping-pong pairs are very diverse across the different samples, for instance displaying heterotypic ping-pong in the digestive gland and homotypic Piwil2-dependent ping-pong in hemolymph cells (electronic supplemental figure 4).

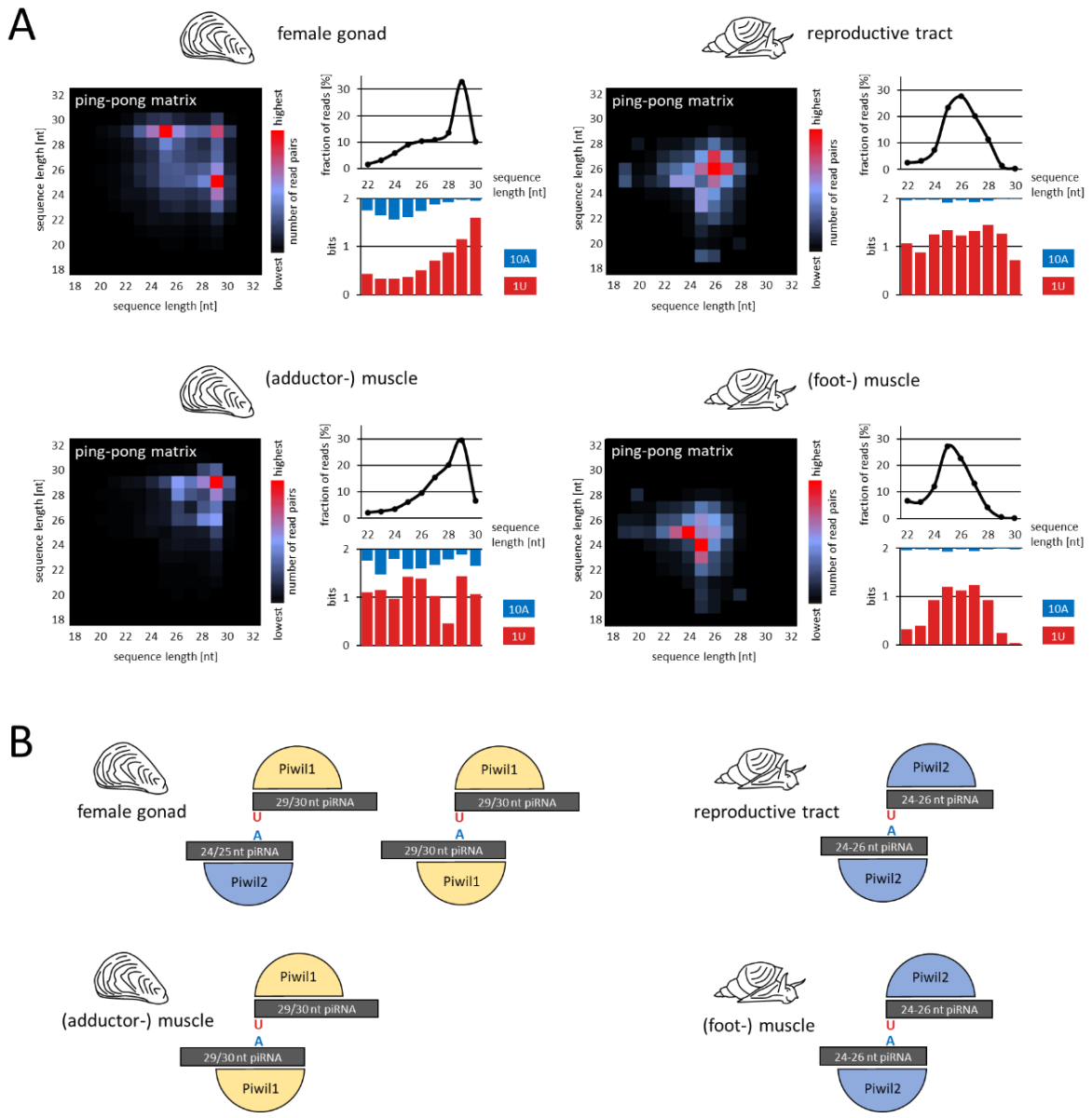


Figure 3.4: Analysis of piRNAs that participate in the ping-pong amplification loop. (A) Ping-pong matrices illustrate frequent length-combinations of ping-pong pairs (sequences with 10 bp 5' overlap). Sequence read length distribution and 1U/10A bias [bits] for ping-pong sequences are shown. (B) Proposed model of ping-pong amplification in the germline and muscle of C. gigas and L. stagnalis.

Since the expression of *Piwil1* compared to *Piwil2* is considerably lower in *L. stagnalis*, we were curious to check whether the corresponding ping-pong pairs might reflect this fact. Indeed, 26/26 nt pairs (homotypic, Piwil2-dependent ping-pong) represent the majority of ping-pong pairs in the reproductive tract (figure 3.4A). In addition, homotypic Piwil2-dependent ping-pong amplification with 24/25 nt ping-pong pairs is also dominant in the *L. stagnalis* muscle (figure 3.4B). However, we also observed differences in ping-pong patterns that do not correlate with the measured mRNA levels of *Piwil1* and *Piwil2*. For example, our data suggests homotypic Piwil2-dependent ping-pong amplification in the oyster gill but homotypic Piwil1-dependent ping-pong amplification in the oyster muscle (electronic

supplemental figure 4), while both tissues display a very similar expression of both *PIWI* paralogs on the mRNA level (figure 3.1F). Thus, we assume that factors other than mere PIWI expression critically influence characteristics of the ping-pong amplification loop.

Moreover, we clearly cannot rule out the possibility that binding preferences of PIWI paralogs have changed on the molluskan lineage and are different from those observed in fly, fish and mouse. This could mean that length profiles of piRNAs associated to each of the molluskan PIWI paralogs might be exactly reciprocal compared to our presumption. One could even speculate that both PIWI paralogs may bind the whole range of piRNAs, which is not possible to disprove without performing corresponding co-Immunoprecipitation experiments. However, based on the presence of piRNA populations with different length profiles (figure 3.2A), their representation in ping-pong pairs together with the differences in their amount of 1U and 10A reads (figure 3.4A), we believe that the above made interpretations are a reasonable and parsimonious interpretation of the data at hand, yet not the only possible one.

3.4. Discussion

Our results reveal that mollusks utilize the PIWI/piRNA pathway as a defense against transposable elements in the germline and in the soma, which corresponds to the situation in arthropods and therefore suggests somatic PIWI/piRNA expression to represent a plesiomorphic protostomian character state. In fact, available data from deeper branching metazoans such as poriferans and cnidarians supports the view that this system was established in the soma even long before the split of protostomes and deuterostomes (Grimson et al. 2008; Praher et al. 2017; Waldron et al. 2018). In addition, based on the observation that a substantial fraction of arthropod and mollusk piRNAs targets messenger RNAs producing the generic ping-pong signature, it seems likely that the last common ancestor of arthropods and mollusks applied the PIWI/piRNA pathway also for post-transcriptional regulation of protein coding genes. Recently, the Xenacoelomorpha phylum, a group of marine worms that were previously thought to belong to the Platyhelminthes clade, was found to represent the sister group of Nephrozoa which comprise protostomes and deuterostomes (Cannon et al. 2016; Rouse et al. 2016). Presently, piRNAs for this outgroup are not characterized but having such data would doubtlessly provide valuable insights and allow to draw conclusions regarding the function of the PIWI/piRNA system in the last common ancestor of all bilaterians, particularly with respect to an ancestral gene-regulatory role. Especially with regard to the latter, functional studies in non-model organisms are urgently needed since the pure bioinformatical evidence for piRNA-dependent processing of protein coding genes does not give any information on its factual biological relevance this process might have in different species. In vertebrates, somatic PIWI/piRNA expression appears to have faded away and reports on somatically expressed piRNAs in mammals are often considered with skepticism for good reasons (Tosar et al. 2018). However, remnants of the former somatic expression might have outlasted to fulfill special functions in specific cells and/or in narrowly defined timespans of development or cell differentiation in the one or the

other clade. Our results indicate that studying the PIWI/piRNA pathway in organisms outside of the main experimental models of *Drosophila* and mouse is necessary to fully understand its evolution and functions.

3.5. Materials and Methods

3.5.1. PIWI gene annotation and tree reconstruction

In order to reconstruct the phylogenetic relations of mollusk Piwi proteins, we first searched for *PIWI* genes in species with an available genome sequence that lack proper annotation (*Lymnaea stagnalis*, *Radix auricularia*, *Lottia gigantea*, *Bathymodiolus platifrons*, *Pinctada martensii*). To this end, we scanned the relevant genomes for sequences that are homologous to annotated *PIWI* paralogs of the pacific oyster (EKC35279 and EKC29295) by aligning translated DNA sequences using tblastx (v2.7.1+, Camacho et al. 2009). Neighboring hits with a distance smaller than 10 kb were grouped as exons of distinct gene loci. Only groups containing the overall best hits for a given locus were retained. Finally, the predicted gene sequences were checked for presence of PIWI and PAZ domains using NCBI conserved domain database (Marchler-Bauer et al. 2015). Similarly, for *PIWI* expression analysis by qPCR in the pond snail, we identified the housekeeping gene GPI (glucose-6-phosphate isomerase) by comparison with the human ortholog (ARJ36701).

The predicted and annotated PIWI protein sequences of the 11 available molluskan species together with PIWI paralogs of human (Piwil1-4) and fly (Ago3, Piwi, Aub), as well as fly argonaute Ago1 were aligned using MUSCLE (v.3.8.31, Edgar 2004). Subsequently, the resulting protein alignment was curated with Gblocks (v.0.91b), allowing smaller final blocks with gap positions and less strict flanking positions. Using ModelGenerator (v.0.85, Keane et al. 2006) we determined LG+G+F (Le and Gascuel 2008) to be the best-fitting model of substitution for our data. The curated alignment (electronic supplemental data 3) was then used for phylogenetic tree reconstruction with PhyML (v3.1, Guindon et al. 2009) applying approximate likelihood-ratio test (SH-like) and LG substitution model, including empirical gamma distribution (G) and character frequencies (F). Support values were generated by bootstrap with 100 replicates.

3.5.2. qPCR

Experiments were performed on commercially available *C. gigas* animals from the western French Atlantic coast (Ile d'Oleron) and captured wild living *L. stagnalis* animals from South-western Germany (Heppenheim). To estimate the expression of the *Piwil* homologs in several tissues of *L. stagnalis* and *C. gigas* we performed qPCR with cDNA synthesized from the total RNA fraction of these tissues. Total RNA was isolated with TriReagent and the polyadenylated transcriptome was reversely transcribed with SuperScript IV using the RT-primer 5'-CGAATTCTAGAGCTCGAGGCAGGCGACATGT$_{25}$VN-3'. Primers amplifying ~ 200 bp long products

of the respective *Piwil* homologs and housekeeping genes were designed with the NCBI tool primer-BLAST on basis of the *L. stagnalis* genome assembly GCA_900036025.1 v1.0 and the *C. gigas* genome assembly GCA_000297895.1 oyster_v9. To prevent amplification of residual genomic DNA, primers were designed to be exon-junction spanning or to span at least several intronic regions. The respective biological replicates were analyzed as technical duplicates on a Corbett Rotor-Gene 6000 real-time PCR cycler and the copy numbers of the genes of interest were quantified by standard curves of the individual primer pair amplicons. For each cDNA sample the calculated *Piwil* copy numbers were relativized by the calculated copy numbers of the housekeeping genes to calibrate for variabilities in sample preparation. These n-fold expression values were finally used to calculate the mean and standard deviation of the replicates. For an improved visualization, the n-fold expression values of each *Piwi* homolog are additionally displayed as a percentage of the respective gonad value.

3.5.3. Small RNA extraction and sequencing

We extracted total RNA from *L. stagnalis* reproductive tract (incl. ovotestis, oviduct, spermatheca, spermiduct, prostate, uterus, vagina, vas deferens) and foot muscle, and total RNA from *C. gigas* adductor muscle and gonadal tissue with TriReagent according to the manufacturer's instructions. For each species we sampled two different individuals per tissue. The small RNA fractions of each obtained total RNA sample were sequenced at BGI, Hong Kong, on a BGISEQ-500 unit. Small RNA sequence datasets for *L. stagnalis* and *C. gigas* are deposited at NCBI's Sequence Read Archive (SRA) and can be accessed under the SRA project IDs SRP130729 and SRP130745. We further used previously published small RNA sequence data from *C. gigas* (Xu et al. 2014) to analyze piRNA expression and characteristics with respect to different developmental stages.

3.5.4. Repeat annotation

We performed *de novo* prediction of repetitive elements in the genome of *L. stagnalis* with RepeatScout (v. 1.0.5, Price et al. 2005). Predicted repetitive elements were classified with RepeatClassifier which is part of the RepeatModeler (v. 1.0.11) package. Transposons that failed to be classified based on known transposons from other species are referred to as unclassified *Lymnaea*-specific transposons (u*L*tra). The resulting repeat sequences, as well as a complete collection of currently available molluskan repeat sequences from RepBase (Bao et al. 2015) were used as reference sequences for repeat masking of the *L. stagnalis* and *C. gigas* genomes with RepeatMasker (v. 4.0.7) using the cross_match search engine and the option -s for most sensitive masking. Annotated repeats in the RepeatMasker output were analyzed with respect to transposon families and divergence from their consensus sequence using the Perl script TE_landscape.pl. Analysis was conducted with the entire repeat dataset as well as with repeats localized in predicted piRNA clusters. TE_landscape.pl is freely available at https://sourceforge.net/projects/protrac/files/tools/.

3.5.5. Gene annotation

We performed *de novo* gene annotation of the *L. stagnalis* genome assembly gLs_1.0 (Davison et al. 2016) using the MAKER genome annotation pipeline (v.2.31.8) in order to identify sRNAs that match protein-coding sequences (Cantarel et al. 2008). Initially, we masked the *L. stagnalis* genome with WindowMasker (Morgulis et al. 2006) using default settings including the duster option to mask low complexity regions. Then, we used available molluskan cDNA data from Ensembl database (release 92) and available mRNA and protein data from *L. stagnalis* deposited at NCBI (Effective April 25, 2018) as input for MAKER. MAKER output files for separate scaffolds were merged using the Perl script mergeMAKERoutput.pl which is freely available at https://sourceforge.net/projects/protrac/files/tools/. The complete genome annotation in GFF3 format and a corresponding mRNA sequence file in FASTA format are available as electronic supplemental data 4 and electronic supplemental data 5.

3.5.6. Processing and annotation of small RNA sequence data

Small RNA sequence datasets were collapsed to non-identical sequences, retaining information on sequence read counts using the Perl script *collapse*. Sequences >36nt were rejected using the Perl script *length-filter*. Finally, low complexity sequences were filtered using the Perl script *duster* with default parameters. All Perl scripts mentioned are part of the NGS toolbox (Rosenkranz et al. 2015a).

We then applied a customized mapping strategy of the remaining small RNA sequence reads based on the consideration that our datasets presumably contain considerable amounts of transposon-derived piRNAs as well as post-transcriptionally edited (e.g. A-to-I) or tailed miRNAs and piRNAs. Genomic mapping was performed with SeqMap (Jiang and Wong 2008) using the option /output_all_matches and allowing up to three mismatches. The obtained alignments were further filtered using the Perl script seqmap_filter.pl that is freely available at https://sourceforge.net/projects/protrac/files/tools/. For the final alignments we allowed up to two non-template 3' nucleotides and up to one internal mismatch. For each sequence, we only considered the best alignments in terms of mismatch counts, but did not reject alignments with equal quality in case of multiple mapping sequences. Sequences that did not produce at least one valid alignment to the reference genome were rejected.

To improve small RNA sequence annotation, we performed *de novo* tRNA, rRNA and miRNA prediction based on the available reference genome assemblies gLs_1.0 (*L. stagnalis*) and GCA_000297895.1 oyster_v9 (*C. gigas*). tRNA annotation was performed with a local copy of tRNAscan (v.1.3.1, Lowe and Chan 2016). Only tRNAs with less than 5% N's were taken for further analysis. rRNA sequences were predicted using a local copy of RNAmmer (v.1.2, Lagesen et al. 2007) and hmmer (v.2.2g, Johnson et al. 2010). Both tools were run with default parameters. We pooled small RNA sequence reads from different replicates and tissues for

each species separately to perform miRNA *de novo* prediction with ShortStack (v.3.8.4, Axtell 2013) using default parameters. The predicted tRNA, rRNA and miRNA precursor sequences, as well as previously published miRNA precursor sequences (Xu et al. 2014; Zhou et al. 2014; Zhao et al. 2016), were used as additional reference sequences for small non-coding RNA annotation with unitas (v.1.4.6, Gebert et al. 2017) which was run with the option -riborase. For *L. stagnalis*, we also included predicted cDNA data based on MAKER annotation (see above). sRNA sequences that did not match to any ncRNA or mRNA of *C. gigas* or *L. stagnalis* were blasted against NCBI nucleotide collection (nr) to search for possible contaminants of parasitic species. Sequences that produced better alignments to genomes of species that possibly parasitized the sampled individuals (*Dicrocoelium*, *Legionella*, *Panagrellus*, *Thelazia, Trichobilharzia*) were considered as contaminants and not used for downstream analyses.

3.5.7. piRNA cluster identification

Sequences that did not produce a match to known non-coding RNAs were considered as putative piRNAs and were used for prediction of piRNA clusters with proTRAC (v. 2.4.0, Rosenkranz and Zischler 2012) applying default settings. piRNA clusters were predicted for each dataset and species separately. The resulting piRNA cluster predictions for each species were condensed, merging clusters with less than 10 kb distance from each other using the Perl script merge_clusters which is freely available at https://sourceforge.net/projects/protrac/files/tools/. To preclude false positive annotation of e.g. tRNA or rRNA genes as piRNA clusters, we validated predicted piRNA clusters by analyzing sRNA reads that mapped to them with respect to their relation to mRNA or other ncRNA classes (electronic supplemental figure 5A). To further check whether piRNA cluster calling may under- or overestimate the number of primary piRNAs in our datasets, we performed an arithmetical approach to estimate the fraction of genuine primary piRNAs based on the fraction of 5' U reads in annotated and non-annotated reads with 24-29 nt length which yields results very close to the number of clustered reads (electronic supplemental methods, electronic supplemental figure 5B). We calculated the sequence read coverage [rpm] for each of the resulting piRNA clusters per dataset. For *C. gigas* piRNA clusters, a heat map for the top 100 piRNA clusters in terms of maximum rpm coverage (accounting for 64% of summed rpm values) was constructed with Heatmapper (Babicki et al. 2016) applying Pearson distance and average linkage clustering. Finally, predicted piRNA clusters were analyzed with respect to their repeat and gene content using the Perl script piC_content.pl which is freely available at https://sourceforge.net/projects/protrac/files/tools/.

3.5.8. Ping-pong quantification

In order to compare ping-pong signatures across multiple datasets with different sequencing depth, we constructed a software tool, PPmeter (v.0.4), that creates bootstrap pseudo-replicates from original datasets and subsequently analyzes the ping-pong signature and

number of ping-pong sequence reads of each pseudo-replicate (default: 100 pseudo-replicates each comprising one million sequence reads). The obtained parameters 'ping-pong score per million bootstrapped reads' (pps-mbr) and 'ping-pong reads per million bootstrapped reads' (ppr-mbr) can be used for quantification and direct comparison of ping-pong activity in different small RNA datasets. The software is freely available at http://www.smallRNAgroup.uni-mainz.de/software.html and https://sourceforge.net/projects/protrac/files/tools/.

3.5.9. Data availability

Sequence data have been uploaded to NCBI's Sequence Read Archive and can be accessed via the accessions SRP130729 and SRP130745.

3.5.10. Code availability

Source code of software that has been written for data processing and analysis is freely available at https://sourceforge.net/projects/protrac/files/tools/.

3.6. Author Contributions

These authors contributed equally: **JJ**, DG, FP.

JJ and JSTK performed total RNA extraction and qPCR experiments. **JJ** analyzed qPCR data and prepared the corresponding figures. DG identified PIWI paralogs in sequenced but unannotated molluskan genomes. DG performed PIWI gene tree reconstruction and prepared the corresponding figure. FP was responsible for farming and dissection of *L. stagnalis* animals. FP and SS performed RNA extraction for subsequent sRNA sequencing. CH, FP, and SS performed de novo miRNA annotation based on the obtained sRNA data. DG performed bioinformatics analysis of piRNA clusters and prepared the corresponding figures. DR analyzed sRNA data and developed Perl scripts for data analysis. DG, **JJ**, and DR wrote the manuscript. JSTK, FP, and CH provided valuable input for corrections and improvements of the manuscript.

3.7. Declarations and Acknowledgements

Figures and tables supporting this paper have been uploaded as electronic supplementary material (https://doi.org/10.1038/s42003-018-0141-4).

The authors declare no competing interests.

We thank Sacha Heerschop, Julia Schumacher, Isabel Fast, and Hans Zischler for helpful comments and discussion. We also thank Mark Helm group for kindly providing chemicals.

This work was supported by the International PhD Program (IPP) coordinated by the Institute of Molecular Biology IMB, Mainz, Germany, funded by the Boehringer Ingelheim Foundation.

3.8. References

Aravin, Alexei A.; Hannon, Gregory J.; Brennecke, Julius (2007): The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. In *Science (New York, N.Y.)* 318 (5851), pp. 761–764. DOI: 10.1126/science.1146484.

Aravin, Alexei A.; Sachidanandam, Ravi; Bourc'his, Deborah; Schaefer, Christopher; Pezic, Dubravka; Toth, Katalin Fejes et al. (2008): A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. In *Molecular cell* 31 (6), pp. 785–799. DOI: 10.1016/j.molcel.2008.09.003.

Axtell, Michael J. (2013): ShortStack: comprehensive annotation and quantification of small RNA genes. In *RNA (New York, N.Y.)* 19 (6), pp. 740–751. DOI: 10.1261/rna.035279.112.

Babicki, Sasha; Arndt, David; Marcu, Ana; Liang, Yongjie; Grant, Jason R.; Maciejewski, Adam; Wishart, David S. (2016): Heatmapper: web-enabled heat mapping for all. In *Nucleic acids research* 44 (W1), W147-53. DOI: 10.1093/nar/gkw419.

Bao, Weidong; Kojima, Kenji K.; Kohany, Oleksiy (2015): Repbase Update, a database of repetitive elements in eukaryotic genomes. In *Mobile DNA* 6 (1), p. 462. DOI: 10.1186/s13100-015-0041-9.

Barckmann, Bridlin; Pierson, Stéphanie; Dufourt, Jérémy; Papin, Catherine; Armenise, Claudia; Port, Fillip et al. (2015): Aubergine iCLIP reveals piRNA-dependent decay of mRNAs involved in germ cell development in the early embryo. In *Cell reports* 12 (7), pp. 1205–1216.

Brennecke, Julius; Aravin, Alexei A.; Stark, Alexander; Dus, Monica; Kellis, Manolis; Sachidanandam, Ravi; Hannon, Gregory J. (2007): Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. In *Cell* 128 (6), pp. 1089–1103. DOI: 10.1016/j.cell.2007.01.043.

Camacho, Christiam; Coulouris, George; Avagyan, Vahram; Ma, Ning; Papadopoulos, Jason; Bealer, Kevin; Madden, Thomas L. (2009): BLAST+: architecture and applications. In *BMC bioinformatics* 10 (1), p. 421. DOI: 10.1186/1471-2105-10-421.

Cannon, Johanna Taylor; Vellutini, Bruno Cossermelli; Smith, Julian; Ronquist, Fredrik; Jondelius, Ulf; Hejnol, Andreas (2016): Xenacoelomorpha is the sister group to Nephrozoa. In *Nature* 530 (7588), pp. 89–93. DOI: 10.1038/nature16520.

Cantarel, Brandi L.; Korf, Ian; Robb, Sofia M. C.; Parra, Genis; Ross, Eric; Moore, Barry et al. (2008): MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. In *Genome research* 18 (1), pp. 188–196. DOI: 10.1101/gr.6743907.

Czech, Benjamin; Hannon, Gregory J. (2016): One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. In *Trends in biochemical sciences* 41 (4), pp. 324–337. DOI: 10.1016/j.tibs.2015.12.008.

Das, Partha P.; Bagijn, Marloes P.; Goldstein, Leonard D.; Woolford, Julie R.; Lehrbach, Nicolas J.; Sapetschnig, Alexandra et al. (2008): Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. In *Molecular cell* 31 (1), pp. 79–90. DOI: 10.1016/j.molcel.2008.06.003.

Davison, Angus; McDowell, Gary S.; Holden, Jennifer M.; Johnson, Harriet F.; Koutsovoulos, Georgios D.; Liu, M. Maureen et al. (2016): Formin Is Associated with Left-Right Asymmetry in the Pond Snail and the Frog. In *Current biology : CB* 26 (5), pp. 654–660. DOI: 10.1016/j.cub.2015.12.071.

Di Giacomo, Monica; Comazzetto, Stefano; Saini, Harpreet; Fazio, Serena de; Carrieri, Claudia; Morgan, Marcos et al. (2013): Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. In *Molecular cell* 50 (4), pp. 601–608. DOI: 10.1016/j.molcel.2013.04.026.

Dowling, Daniel; Pauli, Thomas; Donath, Alexander; Meusemann, Karen; Podsiadlowski, Lars; Petersen, Malte et al. (2016): Phylogenetic Origin and Diversification of RNAi Pathway Genes in Insects. In *Genome biology and evolution* 8 (12), pp. 3784–3793. DOI: 10.1093/gbe/evw281.

Edgar, Robert C. (2004): MUSCLE: a multiple sequence alignment method with reduced time and space complexity. In *BMC bioinformatics* 5, p. 113. DOI: 10.1186/1471-2105-5-113.

Fisher, W. S. (1986): Structure and functions of oyster hemocytes in Immunity in Invertebrates (ed. Brehélin, M.) Proceedings in Life Sciences: Springer.

Flemr, Matyas; Malik, Radek; Franke, Vedran; Nejepinska, Jana; Sedlacek, Radislav; Vlahovicek, Kristian; Svoboda, Petr (2013): A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. In *Cell* 155 (4), pp. 807–816. DOI: 10.1016/j.cell.2013.10.001.

Funayama, Noriko; Nakatsukasa, Mikiko; Mohri, Kurato; Masuda, Yoshiki; Agata, Kiyokazu (2010): Piwi expression in archeocytes and choanocytes in demosponges: insights into the stem cell system in demosponges. In *Evolution & development* 12 (3), pp. 275–287. DOI: 10.1111/j.1525-142X.2010.00413.x.

Gainetdinov, Ildar; Skvortsova, Yulia; Kondratieva, Sofia; Funikov, Sergey; Azhikina, Tatyana (2017): Two modes of targeting transposable elements by piRNA pathway in human testis. In *RNA (New York, N.Y.)* 23 (11), pp. 1614–1625. DOI: 10.1261/rna.060939.117.

Gebert, Daniel; Hewel, Charlotte; Rosenkranz, David (2017): unitas: the universal tool for annotation of small RNAs. In *BMC genomics* 18 (1), p. 71. DOI: 10.1186/s12864-017-4031-9.

Gebert, Daniel; Ketting, René F.; Zischler, Hans; Rosenkranz, David (2015): piRNAs from Pig Testis Provide Evidence for a Conserved Role of the Piwi Pathway in Post-Transcriptional Gene Regulation in Mammals. In *PLoS ONE* 10 (5), e0124860. DOI: 10.1371/journal.pone.0124860.

Gou, Lan-Tao; Dai, Peng; Yang, Jian-Hua; Xue, Yuanchao; Hu, Yun-Ping; Zhou, Yu et al. (2014): Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. In *Cell research* 24 (6), pp. 680–700. DOI: 10.1038/cr.2014.41.

Grimson, Andrew; Srivastava, Mansi; Fahey, Bryony; Woodcroft, Ben J.; Chiang, H. Rosaria; King, Nicole et al. (2008): Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. In *Nature* 455 (7217), pp. 1193–1197. DOI: 10.1038/nature07415.

Guindon, Stéphane; Delsuc, Frédéric; Dufayard, Jean-François; Gascuel, Olivier (2009): Estimating maximum likelihood phylogenies with PhyML. In *Methods in molecular biology (Clifton, N.J.)* 537, pp. 113–137. DOI: 10.1007/978-1-59745-251-9_6.

Ha, Hongseok; Song, Jimin; Wang, Shuoguo; Kapusta, Aurélie; Feschotte, Cédric; Chen, Kevin C.; Xing, Jinchuan (2014): A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. In *BMC genomics* 15 (1), p. 545. DOI: 10.1186/1471-2164-15-545.

Hirano, Takamasa; Iwasaki, Yuka W.; Lin, Zachary Yu-Ching; Imamura, Masanori; Seki, Naomi M.; Sasaki, Erika et al. (2014): Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate. In *RNA (New York, N.Y.)* 20 (8), pp. 1223–1237. DOI: 10.1261/rna.045310.114.

Houwing, Saskia; Berezikov, Eugene; Ketting, René F. (2008): Zili is required for germ cell differentiation and meiosis in zebrafish. In *The EMBO journal* 27 (20), pp. 2702–2711. DOI: 10.1038/emboj.2008.204.

Iwasaki, Yuka W.; Siomi, Mikiko C.; Siomi, Haruhiko (2015): PIWI-Interacting RNA: Its Biogenesis and Functions. In *Annual review of biochemistry* 84, pp. 405–433. DOI: 10.1146/annurev-biochem-060614-034258.

Jiang, Hui; Wong, Wing Hung (2008): SeqMap: mapping massive amount of oligonucleotides to the genome. In *Bioinformatics (Oxford, England)* 24 (20), pp. 2395–2396. DOI: 10.1093/bioinformatics/btn429.

Johnson, L. Steven; Eddy, Sean R.; Portugaly, Elon (2010): Hidden Markov model speed heuristic and iterative HMM search procedure. In *BMC bioinformatics* 11 (1), p. 403. DOI: 10.1186/1471-2105-11-431.

Jones, Brian C.; Wood, Jason G.; Chang, Chengyi; Tam, Austin D.; Franklin, Michael J.; Siegel, Emily R.; Helfand, Stephen L. (2016): A somatic piRNA pathway in the Drosophila fat body ensures metabolic homeostasis and normal lifespan. In *Nat Commun* 7 (1), p. 691. DOI: 10.1038/ncomms13856.

Juliano, Celina E.; Reich, Adrian; Liu, Na; Götzfried, Jessica; Zhong, Mei; Uman, Selen et al. (2014): PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (1), pp. 337–342. DOI: 10.1073/pnas.1320965111.

Kawaoka, Shinpei; Izumi, Natsuko; Katsuma, Susumu; Tomari, Yukihide (2011): 3' end formation of PIWI-interacting RNAs in vitro. In *Molecular cell* 43 (6), pp. 1015–1022. DOI: 10.1016/j.molcel.2011.07.029.

Keane, Thomas M.; Creevey, Christopher J.; Pentony, Melissa M.; Naughton, Thomas J.; Mclnerney, James O. (2006): Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. In *BMC evolutionary biology* 6, p. 29. DOI: 10.1186/1471-2148-6-29.

Kerner, Pierre; Degnan, Sandie M.; Marchand, Lauriane; Degnan, Bernard M.; Vervoort, Michel (2011): Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge Amphimedon queenslandica. In *Molecular biology and evolution* 28 (8), pp. 2289–2303. DOI: 10.1093/molbev/msr046.

Kiuchi, Takashi; Koga, Hikaru; Kawamoto, Munetaka; Shoji, Keisuke; Sakai, Hiroki; Arai, Yuji et al. (2014): A single female-specific piRNA is the primary determiner of sex in the silkworm. In *Nature* 509 (7502), pp. 633–636. DOI: 10.1038/nature13315.

Lagesen, Karin; Hallin, Peter; Rødland, Einar Andreas; Staerfeldt, Hans-Henrik; Rognes, Torbjørn; Ussery, David W. (2007): RNAmmer: consistent and rapid annotation of ribosomal RNA genes. In *Nucleic acids research* 35 (9), pp. 3100–3108. DOI: 10.1093/nar/gkm160.

Lau, Yuk-Ting; Sussman, Lauren; Pales Espinosa, Emmanuelle; Katalay, Selma; Allam, Bassem (2017): Characterization of hemocytes from different body fluids of the eastern oyster Crassostrea virginica. In *Fish & shellfish immunology* 71, pp. 372–379. DOI: 10.1016/j.fsi.2017.10.025.

Le, Si Quang; Gascuel, Olivier (2008): An improved general amino acid replacement matrix. In *Molecular biology and evolution* 25 (7), pp. 1307–1320. DOI: 10.1093/molbev/msn067.

Lewis, Samuel H.; Quarles, Kaycee A.; Yang, Yujing; Tanguy, Melanie; Frézal, Lise; Smith, Stephen A. et al. (2018): Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. In *Nature ecology & evolution* 2 (1), pp. 174–181. DOI: 10.1038/s41559-017-0403-4.

Lewis, Samuel H.; Salmela, Heli; Obbard, Darren J. (2016): Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. In *Genome biology and evolution* 8 (3), pp. 507–518. DOI: 10.1093/gbe/evw018.

Li, Xin Zhiguo; Roy, Christian K.; Dong, Xianjun; Bolcun-Filas, Ewelina; Wang, Jie; Han, Bo W. et al. (2013): An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. In *Molecular cell* 50 (1), pp. 67–81. DOI: 10.1016/j.molcel.2013.02.016.

Lim, Robyn S. M.; Anand, Amit; Nishimiya-Fujisawa, Chiemi; Kobayashi, Satoru; Kai, Toshie (2014): Analysis of Hydra PIWI proteins and piRNAs uncover early evolutionary origins of the piRNA pathway. In *Developmental biology* 386 (1), pp. 237–251. DOI: 10.1016/j.ydbio.2013.12.007.

Llonga, Natalia; Ylla, Guillem; Bau, Josep; Belles, Xavier; Piulachs, Maria-Dolors (2018): Diversity of piRNA expression patterns during the ontogeny of the German cockroach. In *J. Exp. Zool. (Mol. Dev. Evol.)* 330 (5), pp. 288–295. DOI: 10.1002/jez.b.22815.

Lowe, Todd M.; Chan, Patricia P. (2016): tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. In *Nucleic acids research* 44 (W1), W54-7. DOI: 10.1093/nar/gkw413.

Ma, Xiaoshi; Ji, Aichang; Zhang, Zhifeng; Yang, Dandan; Liang, Shaoshuai; Wang, Yuhan; Qin, Zhenkui (2017): Piwi1 is essential for gametogenesis in mollusk Chlamys farreri. In *PeerJ* 5, e3412. DOI: 10.7717/peerj.3412.

Madison-Villar, M. J.; Sun, Cheng; Lau, Nelson C.; Settles, Matthew L.; Mueller, Rachel Lockridge (2016): Small RNAs from a Big Genome: The piRNA Pathway and Transposable Elements in the Salamander Species Desmognathus fuscus. In *Journal of molecular evolution* 83 (3-4), pp. 126–136. DOI: 10.1007/s00239-016-9759-3.

Manakov, Sergei A.; Pezic, Dubravka; Marinov, Georgi K.; Pastor, William A.; Sachidanandam, Ravi; Aravin, Alexei A. (2015): MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation. In *Cell reports* 12 (8), pp. 1234–1243. DOI: 10.1016/j.celrep.2015.07.036.

Marchler-Bauer, Aron; Derbyshire, Myra K.; Gonzales, Noreen R.; Lu, Shennan; Chitsaz, Farideh; Geer, Lewis Y. et al. (2015): CDD: NCBI's conserved domain database. In *Nucleic acids research* 43 (Database issue), D222-6. DOI: 10.1093/nar/gku1221.

Miesen, Pascal; Girardi, Erika; van Rij, Ronald P. (2015): Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. In *Nucleic acids research* 43 (13), pp. 6545–6556. DOI: 10.1093/nar/gkv590.

Morgulis, Aleksandr; Gertz, E. Michael; Schäffer, Alejandro A.; Agarwala, Richa (2006): WindowMasker: window-based masker for sequenced genomes. In *Bioinformatics (Oxford, England)* 22 (2), pp. 134–141. DOI: 10.1093/bioinformatics/bti774.

Murchison, Elizabeth P.; Kheradpour, Pouya; Sachidanandam, Ravi; Smith, Carly; Hodges, Emily; Xuan, Zhenyu et al. (2008): Conservation of small RNA pathways in platypus. In *Genome research* 18 (6), pp. 995–1004. DOI: 10.1101/gr.073056.107.

Nandi, Sayan; Chandramohan, Dhruva; Fioriti, Luana; Melnick, Ari M.; Hébert, Jean M.; Mason, Christopher E. et al. (2016): Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. In *Proceedings of the National Academy of Sciences of the United States of America* 113 (45), pp. 12697–12702. DOI: 10.1073/pnas.1609287113.

Palakodeti, Dasaradhi; Smielewska, Magda; Lu, Yi-Chien; Yeo, Gene W.; Graveley, Brenton R. (2008): The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. In *RNA (New York, N.Y.)* 14 (6), pp. 1174–1186. DOI: 10.1261/rna.1085008.

Perrat, Paola N.; DasGupta, Shamik; Wang, Jie; Theurkauf, William; Weng, Zhiping; Rosbash, Michael; Waddell, Scott (2013): Transposition-driven genomic heterogeneity in the Drosophila brain. In *Science (New York, N.Y.)* 340 (6128), pp. 91–95. DOI: 10.1126/science.1231965.

Pezic, Dubravka; Manakov, Sergei A.; Sachidanandam, Ravi; Aravin, Alexei A. (2014): piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. In *Genes & Development* 28 (13), pp. 1410–1428. DOI: 10.1101/gad.240895.114.

Praher, Daniela; Zimmermann, Bob; Genikhovich, Grigory; Columbus-Shenkar, Yaara; Modepalli, Vengamanaidu; Aharoni, Reuven et al. (2017): Characterization of the piRNA pathway during development of the sea anemone Nematostella vectensis. In *RNA biology* 14 (12), pp. 1727–1741. DOI: 10.1080/15476286.2017.1349048.

Price, Alkes L.; Jones, Neil C.; Pevzner, Pavel A. (2005): De novo identification of repeat families in large genomes. In *Bioinformatics (Oxford, England)* 21 Suppl 1, i351-8. DOI: 10.1093/bioinformatics/bti1018.

Rajasethupathy, Priyamvada; Antonov, Igor; Sheridan, Robert; Frey, Sebastian; Sander, Chris; Tuschl, Thomas; Kandel, Eric R. (2012): A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. In *Cell* 149 (3), pp. 693–707. DOI: 10.1016/j.cell.2012.02.057.

Reuter, Michael; Berninger, Philipp; Chuma, Shinichiro; Shah, Hardik; Hosokawa, Mihoko; Funaya, Charlotta et al. (2011): Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. In *Nature* 480 (7376), pp. 264–267. DOI: 10.1038/nature10672.

Rojas-Ríos, Patricia; Chartier, Aymeric; Pierson, Stéphanie; Simonelig, Martine (2017): Aubergine and piRNAs promote germline stem cell self-renewal by repressing the proto-oncogene Cbl. In *The EMBO journal* 36 (21), pp. 3194–3211. DOI: 10.15252/embj.201797259.

Roovers, Elke F.; Rosenkranz, David; Mahdipour, Mahdi; Han, Chung-Ting; He, Nannan; Chuva de Sousa Lopes, Susana M. et al. (2015): Piwi proteins and piRNAs in mammalian oocytes and early embryos. In *Cell reports* 10 (12), pp. 2069–2082. DOI: 10.1016/j.celrep.2015.02.062.

Rosenberg, Gary (2014): A New Critical Estimate of Named Species-Level Diversity of the Recent Mollusca*. In *American Malacological Bulletin* 32 (2), p. 308. DOI: 10.4003/006.032.0204.

Rosenkranz, David; Han, Chung-Ting; Roovers, Elke F.; Zischler, Hans; Ketting, René F. (2015a): Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence. In *Genomics data* 5, pp. 309–313. DOI: 10.1016/j.gdata.2015.06.026.

Rosenkranz, David; Rudloff, Stefanie; Bastuck, Katharina; Ketting, René F.; Zischler, Hans (2015b): Tupaia small RNAs provide insights into function and evolution of RNAi-based transposon defense in mammals. In *RNA (New York, N.Y.)* 21 (5), pp. 911–922. DOI: 10.1261/rna.048603.114.

Rosenkranz, David; Zischler, Hans (2012): proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. In *BMC bioinformatics* 13 (1), p. 355. DOI: 10.1186/1471-2105-13-5.

Ross, Robert J.; Weiner, Molly M.; Lin, Haifan (2014): PIWI proteins and PIWI-interacting RNAs in the soma. In *Nature* 505 (7483), pp. 353–359. DOI: 10.1038/nature12987.

Rouget, Christel; Papin, Catherine; Boureux, Anthony; Meunier, Anne-Cécile; Franco, Bénédicte; Robine, Nicolas et al. (2010): Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. In *Nature* 467 (7319), pp. 1128–1132. DOI: 10.1038/nature09465.

Rouse, Greg W.; Wilson, Nerida G.; Carvajal, Jose I.; Vrijenhoek, Robert C. (2016): New deep-sea species of Xenoturbella and the position of Xenacoelomorpha. In *Nature* 530 (7588), pp. 94–97. DOI: 10.1038/nature16545.

Russell, Stewart; Patel, Mehool; Gilchrist, Graham; Stalker, Leanne; Gillis, Daniel; Rosenkranz, David; LaMarre, Jonathan (2017): Bovine piRNA-like RNAs are associated with both transposable elements and mRNAs. In *Reproduction (Cambridge, England)* 153 (3), pp. 305–318. DOI: 10.1530/REP-16-0620.

Sasaki, Takashi; Shiohama, Aiko; Minoshima, Shinsei; Shimizu, Nobuyoshi (2003): Identification of eight members of the Argonaute family in the human genome. In *Genomics* 82 (3), pp. 323–330.

Schorn, Andrea J.; Gutbrod, Michael J.; LeBlanc, Chantal; Martienssen, Rob (2017): LTR-Retrotransposon Control by tRNA-Derived Small RNAs. In *Cell* 170 (1), 61-71.e11. DOI: 10.1016/j.cell.2017.06.013.

Schurko, Andrew M.; Logsdon, John M.; Eads, Brian D. (2009): Meiosis genes in Daphnia pulex and the role of parthenogenesis in genome evolution. In *BMC Evol Biol* 9 (1), p. 78. DOI: 10.1186/1471-2148-9-78.

Senti, Kirsten-André; Jurczak, Daniel; Sachidanandam, Ravi; Brennecke, Julius (2015): piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. In *Genes & Development* 29 (16), pp. 1747–1762. DOI: 10.1101/gad.267252.115.

Teixeira, Felipe Karam; Okuniewska, Martyna; Malone, Colin D.; Coux, Rémi-Xavier; Rio, Donald C.; Lehmann, Ruth (2017): piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. In *Nature* 552 (7684), pp. 268–272. DOI: 10.1038/nature25018.

Thomson, Travis; Lin, Haifan (2009): The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. In *Annual review of cell and developmental biology* 25, pp. 355–376. DOI: 10.1146/annurev.cellbio.24.110707.175327.

Tong, Ying; Zhang, Yang; Huang, Jiaomei; Xiao, Shu; Zhang, Yuehuan; Li, Jun et al. (2015): Transcriptomics Analysis of Crassostrea hongkongensis for the Discovery of Reproduction-Related Genes. In *PLoS ONE* 10 (8), e0134280. DOI: 10.1371/journal.pone.0134280.

Tosar, Juan Pablo; Rovira, Carlos; Cayota, Alfonso (2018): Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. In *Commun Biol* 1 (1), p. 94. DOI: 10.1038/s42003-017-0001-7.

Vourekas, Anastassios; Zheng, Qi; Alexiou, Panagiotis; Maragkakis, Manolis; Kirino, Yohei; Gregory, Brian D.; Mourelatos, Zissimos (2012): Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. In *Nature structural & molecular biology* 19 (8), pp. 773–781. DOI: 10.1038/nsmb.2347.

Waldron, Fergal M.; Stone, Graham N.; Obbard, Darren J. (2018): Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. In *PLoS genetics* 14 (7), e1007533.

Watanabe, Toshiaki; Lin, Haifan (2014): Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs. In *Molecular cell* 56 (1), pp. 18–27. DOI: 10.1016/j.molcel.2014.09.012.

Xu, Fei; Wang, Xiaotong; Feng, Yue; Huang, Wen; Wang, Wei; Li, Li et al. (2014): Identification of Conserved and Novel MicroRNAs in the Pacific Oyster Crassostrea gigas by Deep Sequencing. In *PloS one* 9 (8), e104371. DOI: 10.1371/journal.pone.0104371.

Zhang, Peng; Kang, Jun-Yan; Gou, Lan-Tao; Wang, Jiajia; Xue, Yuanchao; Skogerboe, Geir et al. (2015): MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. In *Cell research* 25 (2), pp. 193–207. DOI: 10.1038/cr.2015.4.

Zhang, Zhao; Xu, Jia; Koppetsch, Birgit S.; Wang, Jie; Tipping, Cindy; Ma, Shengmei et al. (2011): Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. In *Molecular cell* 44 (4), pp. 572–584. DOI: 10.1016/j.molcel.2011.10.011.

Zhao, Xuelin; Yu, Hong; Kong, Lingfeng; Liu, Shikai; Li, Qi (2016): High throughput sequencing of small RNAs transcriptomes in two Crassostrea oysters identifies microRNAs involved in osmotic stress response. In *Sci Rep* 6 (1), p. 1161. DOI: 10.1038/srep22687.

Zhou, Xue; Liao, Zhen; Jia, Qidong; Cheng, Luogen; Li, Fei (2007): Identification and characterization of Piwi subfamily in insects. In *Biochemical and biophysical research communications* 362 (1), pp. 126–131. DOI: 10.1016/j.bbrc.2007.07.179.

Zhou, Zhi; Wang, Lingling; Song, Linsheng; Liu, Rui; Zhang, Huan; Huang, Mengmeng; Chen, Hao (2014): The Identification and Characteristics of Immune-Related MicroRNAs in Haemocytes of Oyster Crassostrea gigas. In *PloS one* 9 (2), e88397. DOI: 10.1371/journal.pone.0088397.

4. PLD6-dependent small RNAs – An RNAi-based approach to identify somatic piRNAs

Julia Jehn[1], Julian S. T. Kiefer[1], David Rosenkranz[1]

[1]Institute of Organismic and Molecular Evolution, Anthropology, Johannes Gutenberg University Mainz, Anselm-Franz-von-Bentzel-Weg 7, 55099 Mainz, Germany

This chapter has not yet been published.

4.1. Abstract

Recently it was shown that a functional somatic piRNA pathway is not only a special trait of lower metazoans like the *Cnidaria* but is also widespread in protostomian species like arthropods and mollusks. While it seems that a somatic piRNA pathway got lost in the branch of vertebrates, several studies indicate that it may still be active in specific niches like stem cells of mammalian brains. To identify piRNAs, co-immunoprecipitation with PIWI proteins is the gold standard. As appropriate antibodies against PIWI proteins are not available, crawling small RNA sequencing data for sequence-homology and certain characteristics is the general approach to identify putative piRNAs in the soma. This method is only predictive and does not allow the identification of novel characteristics and functions of this pathway. Here, we developed a RNAi-based approach to verify the expression of small RNAs whose expression depend on Phospholipase D family member 6 (PLD6). PLD6 is an endonuclease that is critical for production of primary piRNAs in the animal germline. By siRNA-mediated knockdown of PLD6 in cultured cells, followed by small RNA sequencing, we identify somatic PLD6-dependent small RNAs (putative piRNAs) *in vitro*. We demonstrate that in somatic cells PLD6 acts as a functional endonuclease generating small RNAs with the typical size range of piRNAs.

4.2. Introduction

While PIWI proteins and piRNAs are ubiquitously expressed in arthropods and mollusks (Lewis et al. 2018; Jehn et al. 2018), the piRNA pathway appears to be restricted to the germline in vertebrates. Reports on somatic piRNAs in mammals base on pure sequence homology (Lee et al. 2011; Yan et al. 2011; Wu et al. 2010; Rizzo et al. 2014; Ghosheh et al. 2016; Dharap et al. 2011; Zhang et al. 2013; Ortogero et al. 2014; Phay et al. 2018; Nandi et al. 2016) and are controversially discussed in the field (Tosar et al. 2018). This skepticism is encouraged by the fact, that transposons are highly methylated and thus inactivated in somatic cells. As transposons are the main targets of the piRNA pathway this challenges the functional relevance of somatic piRNAs. In mammals, transposons were thought to be active solely in the germline, where the genome of gametes undergoes epigenetic reprogramming including global genomic de- and re-methylation. Recently transposon activity was also found in human brain tissue, particularly in the hippocampus (Upton et al. 2015). Interestingly, the dentate

gyrus within the hippocampus is one of the few brain regions known to have high rates of adult neurogenesis (Eriksson et al. 1998; Boldrini et al. 2018). Epigenetic modifications taking place during the early stages of neuronal differentiation seem to de-repress the promoter of L1 retrotransposons for a short time (Muotri et al. 2005; Coufal et al. 2009), which is reminiscent of the gonadal processes. It is therefore possible, that transposon-regulating mechanisms like the piRNA pathway are also active in the human brain. Thus, in vertebrates, a somatic piRNA pathway might have remained to be active in this particular niche. This hypothesis is supported by the finding that induced pluripotent stem cells (iPS cells) of non-human primates, which have a lower expression of Piwil2 than human iPS cells, show higher L1 transposition rates (Marchetto et al. 2013).

As due to the lack of appropriate antibodies there is no reliable method to identify human piRNAs, we sought to develop an alternative, antibody-independent approach that targets a key factor of piRNA biogenesis: Phospholipase D family member 6 (PLD6). PLD6, also called Zucchini, is a mitochondria-bound endonuclease that generates primary piRNAs in the germline. In a 3'-directed and phased process it catalyzes the endonucleolytic cleavage of large, single-stranded piRNA precursor transcripts into almost fully mature piRNAs (Ipsaro et al. 2012; Nishimasu et al. 2012; Mohn et al. 2015; Han et al. 2015). Knockout experiments in mice demonstrated that no piRNAs are formed during spermatogenesis in the absence of PLD6 (Watanabe et al. 2011). Recently, it was shown that PLD6 is also involved in the processing of secondary piRNAs in germline cells (Mohn et al. 2015). Figure 4.1 schematically shows the function of PLD6 during piRNA biogenesis in the female germline of *Drosophila*. As PLD6 is broadly expressed in somatic tissues, it might also produce potential somatic piRNAs.

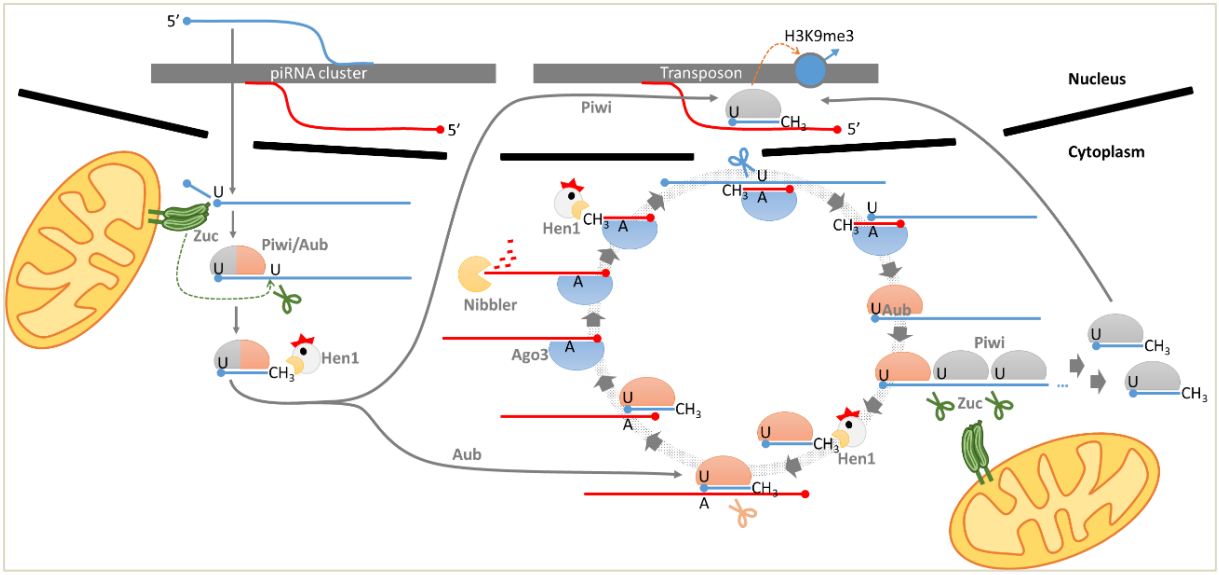


Figure 4.1: Schematic overview of the piRNA biogenesis and transposon silencing mechanism in the female germline of *Drosophila*. Highlighted are the processes where PLD6 (Zucchini in *Drosophila*) is involved in the generation of piRNAs.

Hence, in order to identify putative somatic piRNAs, our approach is to downregulate PLD6 in cultured cells by transfecting an antisense siRNA and compare the small RNA profile to a control. As a proof of concept, we tested this RNAi-based approach in HEK293T cells, where we characterized the detected PLD6-dependent small RNAs further and investigated potential target transcripts.

4.3. Results

4.3.1. PLD6 acts as an endonuclease in HEK293T cells

To test whether PLD6 generates small RNAs in somatic cells in a similar way as it does in germ cells, we performed siRNA mediated PLD6-knockdown experiments in HEK293T cells. After the knockdown efficacy was validated by qPCR and western blot (S-Figure 4.1), small RNA libraries from knockdown and control cells were generated and sequenced.

The obtained sequence reads (see S-Figure 4.2 and S-Figure 4.3 for the read length distribution) were subjected to several quality filters (S-Table 4.1) and mapped to the human genome. We then analyzed the differential small RNA read abundancy between the two libraries using the annotation tool unitas. According to these calculations, 899 non-identical sequences represented by 2,704,936 reads are significantly more abundant in the non-target control (these are presumably PLD6-dependent small RNAs), while 172 non-identical sequences represented by 249,829 reads are higher expressed in the PLD6-knockdown cells (figure 4.2, S-Table 4.1).

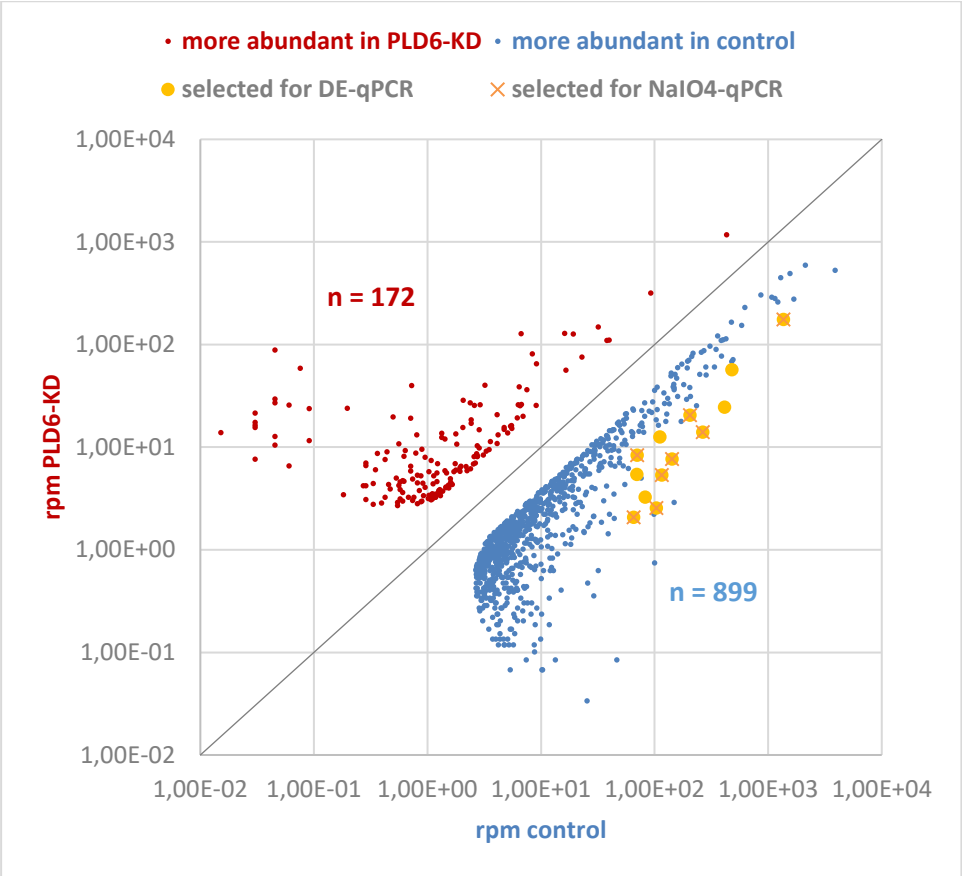


Figure 4.2: Plotted rpm values of small RNAs being differentially expressed between PLD6-KD and control as calculated by unitas (`-diffexpr_estdev 0.1`). The highlighted sequences were selected for further qPCR analysis to crosscheck the differential expression (DE-qPCR) and to check for 2'-O-methylation of the 3' ends of PLD6-dependent small RNAs (NaIO4-qPCR).

Strikingly, the majority of the small RNAs likely to be PLD6-dependent has the typical length of piRNAs (24-32 nt, figure 4.3). In contrast to that, only few of the sequences being higher expressed in the PLD6-knockdown HEK293T cells lies in this size range. This indicates that PLD6 works as an endonuclease not only in the germline, but also in somatic cells. Although the PLD6-dependent small RNAs exhibit the typical piRNA length distribution, they mainly originate from different transcripts compared to piRNAs in the germline (figure 4.4).

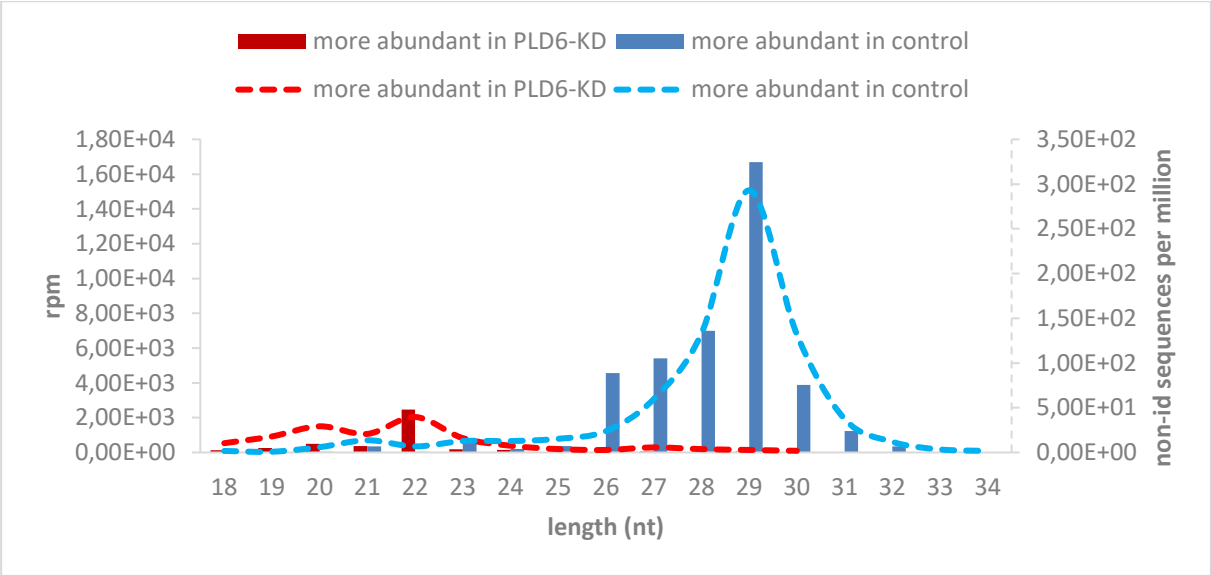


Figure 4.3: Length distribution of small RNAs being differentially expressed between PLD6-KD and control. It shows that upon PLD6-knockdown sequences with the typical length of piRNAs are diminished. The dashed line depicts the non-identical sequences per million being represented by the reads of a specific length.

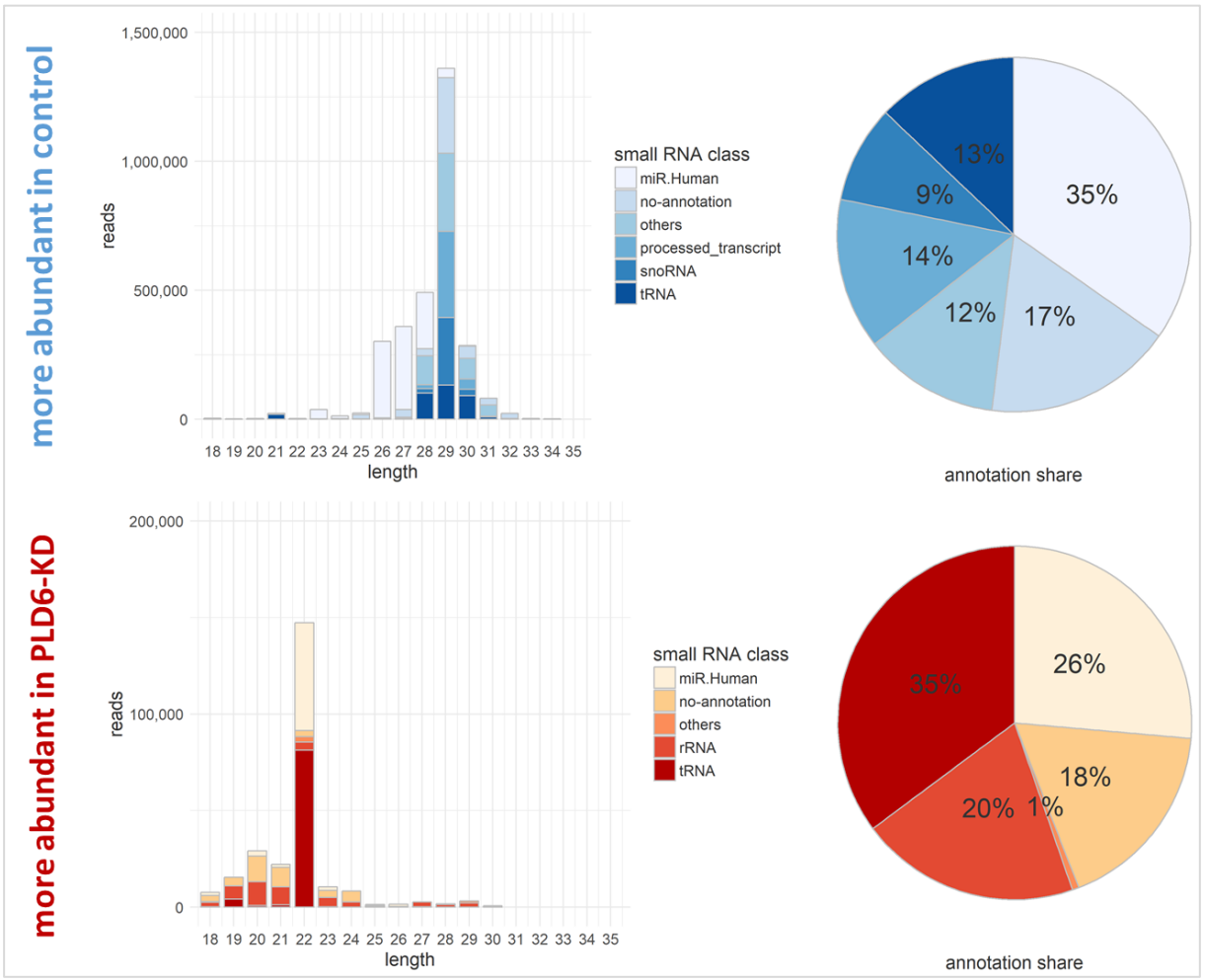


Figure 4.4: Unitas annotation of significantly higher expressed small RNAs in control (upper) and PLD6-KD (lower) HEK293T cells.

To rule out erroneous distortions arising from library preparations, we crosschecked the differential expression of 14 selected small RNAs by using qPCR. These 14 small RNAs, that cover the whole length spectrum, were chosen due to their relative high abundance in the control library (figure 4.2, yellow bullets). Blast analyses of these small RNAs show hits for e.g. miRNAs, tRNA fragments, snoRNAs and piRNAs.

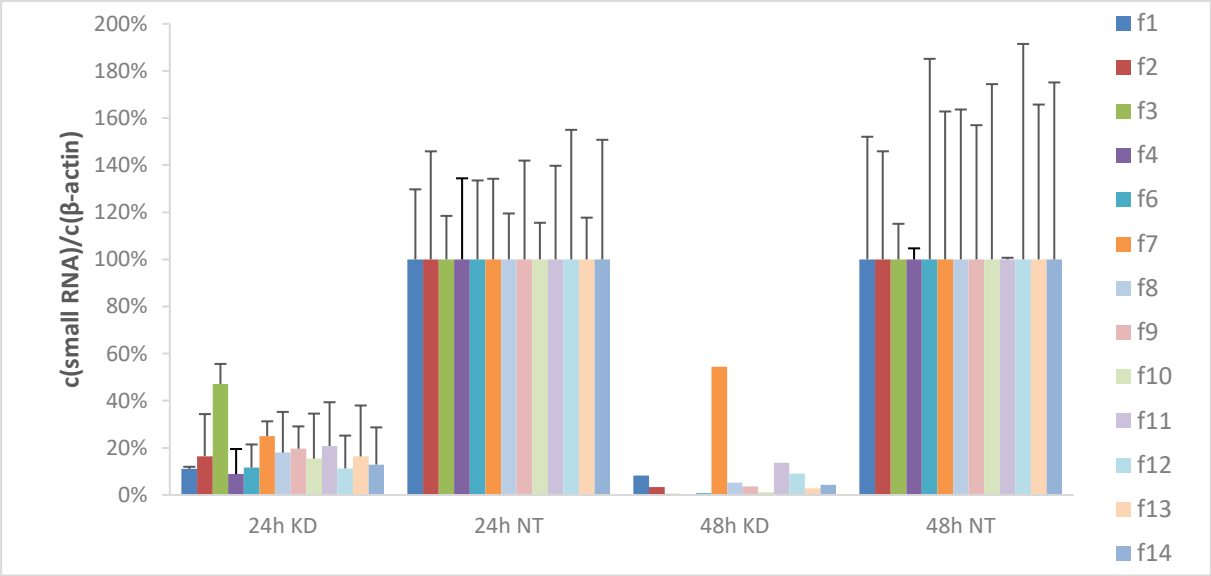


Figure 4.5: qPCR on selected small RNAs (n=2) supports the differential expression in PLD6-knockdown (KD) and non-target control (NT) HEK293T cells at 24 and 48 hours as determined by unitas analysis of the small RNA sequencing data. Primer pair f5 did not amplify any product.

The expression profile of the tested small RNAs confirms the differential expression revealed by sequencing (figure 4.5). Hence, qPCR analysis on 14 small RNAs from PLD6-knockdown and control HEK293T cells supports the notion that PLD6 can generate small RNAs in somatic cells by acting as an endonuclease.

4.3.2. PLD6-dependent small RNAs in HEK293T cells do not have typical piRNA characteristics

As small RNA sequences with the typical length of piRNAs were significantly depleted in the PLD6-knockdown cells, we asked whether these small RNAs bear the typical characteristics of piRNAs. In the germline, PLD6-generated piRNAs are usually primary piRNAs, which show a bias for a uridine at the first position of the 5' end. Like all piRNAs, they have a protective 2' O-methylation at their 3' end.

We analyzed the sequences that are more abundant in the control library for the base proportion at their first position of the 5’ end. Neither for the non-identical sequences, nor for the total reads a 1U bias could be observed (figure 4.6A). Thus, the sequences likely to be PLD6-dependent in somatic cells do not seem to have a 1U bias.

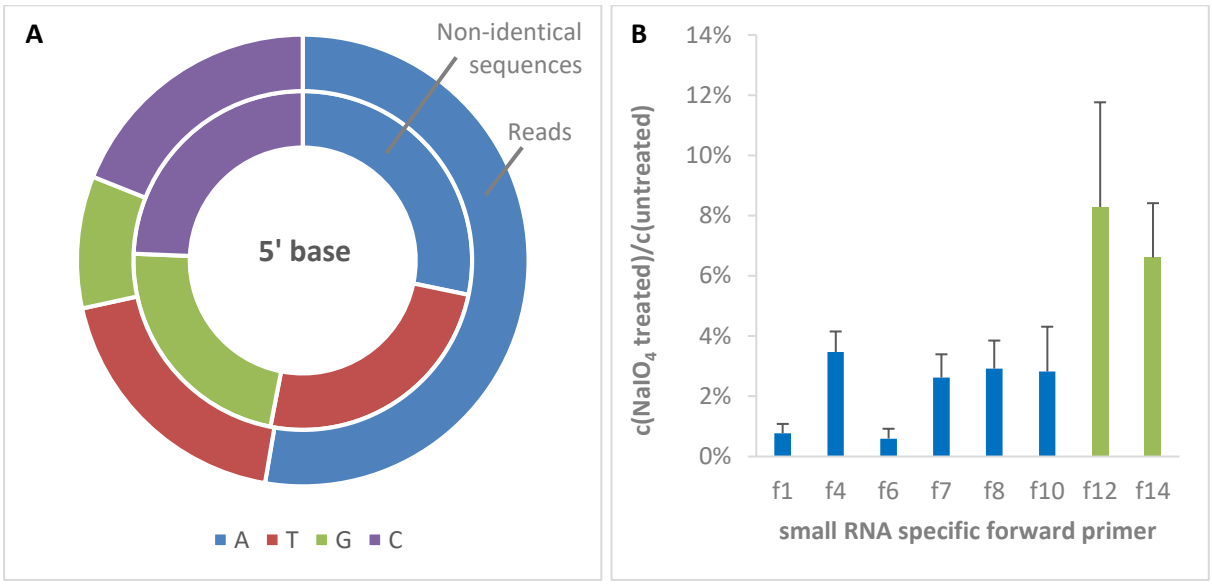


Figure 4.6: (A) Unlike piRNAs, somatic PLD6-dependent small RNAs do not show a 1U bias at the 5’ end. (B) Sodium periodate treatment followed by qPCR analysis (n=2) reveals that the tested PLD6-dependent small RNAs are not 2’-O-methylated at their 3’ end. Small RNAs listed as piRNAs (green) show a slightly higher resistance to this treatment indicating a residual modification.

To check for potential 2’-O-methylation of the 3’ ends, eight of the previously tested small RNAs were submitted to sodium periodate treatment (figure 4.2, orange crosses). As figure 4.6B shows, for each of the eight tested small RNAs, almost no amplification was detectable by qPCR compared to a non-treatment control. Hence, PLD6-dependent small RNAs in HEK293T cells do not seem to be 2’-O-methylated at their 3’ ends.

Interestingly, the two small RNAs, which are listed as piRNA sequences in the NCBI database (figure 4.6B, green bars), show a slightly higher resistance to the treatment than the other ones. This suggests that there might be a residual activity of the piRNA processing pathway in somatic cells.

4.3.3. RNA sequencing of PLD6-knockdown HEK293T cells does not reveal target transcripts of PLD6-dependent small RNAs

To investigate, whether the PLD6-dependent small RNAs have a gene regulatory function, we sequenced the transcriptome of the PLD6-knockdown HEK293T cells and analyzed the differential expression in comparison to non-target control HEK293T cells. As figure 4.7A

shows, 8265 genes were found to be significantly differentially expressed between the PLD6-knockdown and the control cells (adjusted p-value < 0.01). While slightly more genes get upregulated (4,179 genes) upon PLD6-knockdown, genes that are downregulated (4,086 genes) tend to have higher n-fold changes. As expected, PLD6 is significantly downregulated. Its expression is reduced about 70% (log2FC = -1.73).

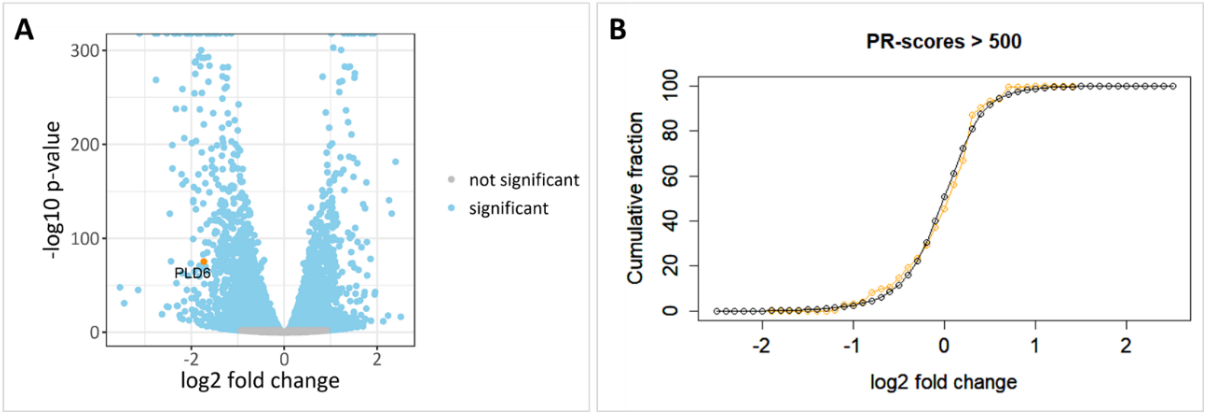


Figure 4.7: (A) Volcano plot of differential expression analysis for protein-coding genes of PLD6-knockdown and control HEK293T cells. Genes with an adjusted p-value below 0.01 were considered significantly differentially expressed (blue). (B) Cumulative plot for log2-fold-change values of genes that were identified by piRanha as potential targets of PLD6-dependent small RNAs (orange) and of all genes (black).

In order to check for genes that might be higher expressed in the PLD6-knockdown cells due to the loss of regulating PLD6-dependent small RNAs, we mapped the identified PLD6-dependent small RNAs to the human transcriptome. Depending on the number of alignments that follow the piRNA targeting rules according to Zhang et al. 2018 and the differential expression of the matching small RNAs a score for each transcript was calculated using the self-developed Perl script piRanha.pl. If the PLD6-dependent small RNAs silence genes by following the piRNA targeting rules, transcripts with high piRanha scores should be enriched to get upregulated when PLD6-dependent small RNAs are depleted by PLD6-knockdown.

When analyzing the differential expression of transcripts that might be targets according to piRanha (piRanha score > 500) we do not see any difference to the transcriptome in general (figure 4.7B). This suggests that the PLD6-dependent small RNAs in HEK293T cells do not regulate protein-coding transcripts by a piRNA-like RNAi mechanism.

In order to find an explanation for the high number of differentially expressed genes upon PLD6-knockdown in HEK293T cells, we performed GO term enrichment analysis. The only molecular function that was significantly enriched for the upregulated genes was the structural molecule activity of ribosome components (figure 4.8A; 13 genes downregulated, 76 genes upregulated). This is in line with the small RNA sequencing data, where ribosomal small RNAs were found to be upregulated upon PLD6-knockdown (figure 4.4). Thus, knocking down PLD6 seems to influence the ribosomal dynamics in HEK293T cells.

As PLD6 is a major component of the lipid-signaling pathway of mitochondrial fusion and fission processes, we also checked for the differential expression of genes involved in mitochondrial fusion and fission. As figure 4.8B shows, most of the 41 analyzed mitochondria-related genes are significantly down- or upregulated (15 down and 14 up; adjusted p-value < 0.01), with the extent of regulation being stronger for the downregulated genes.

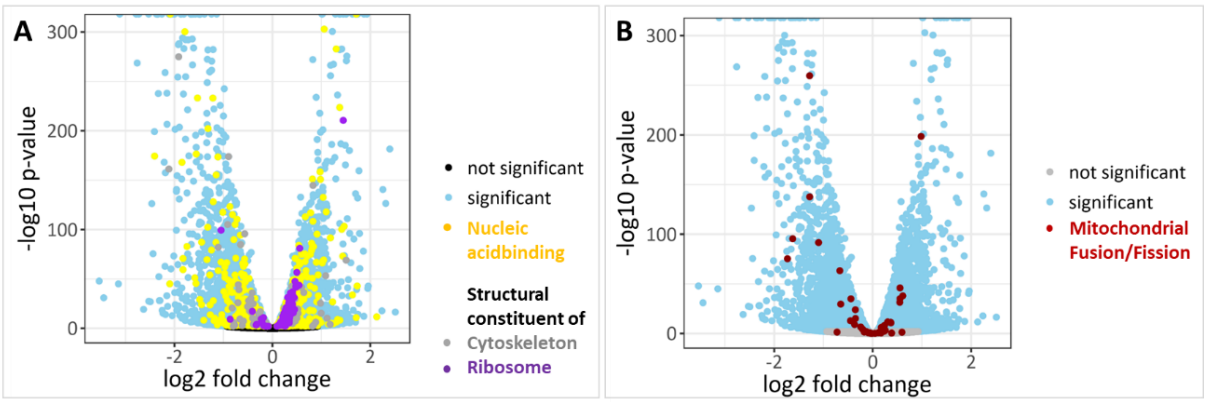


Figure 4.8: Volcano plot of differential expression analysis for protein-coding genes of PLD6-knockdown and control HEK293T cells with (A) genes assigned to selected GO terms of molecular function being highlighted and (B) genes involved in mitochondrial dynamics being marked in red.

As the major targets of piRNAs are transposable elements, we wanted to check whether the expression of repetitive sequences goes up upon PLD6-knockdown. According to DESeq2 analysis (figure 4.9A), 186 of the 1077 analyzed repetitive elements were significantly differentially expressed (adjusted p-value < 0.01). However, there is no difference in number or extend of regulation when comparing down- to upregulated repetitive elements. This suggests that PLD6-dependent small RNAs are not likely to downregulate repetitive elements in HEK293T cells. This assumption is further supported by the fact that most of the PLD6-dependent small RNAs are mapping in sense orientation to the repetitive elements (figure 4.9C). Moreover, two third of the differentially expressed repetitive elements are not transposable elements, but tRNAs, scRNAs or rRNAs (figure 4.9B). Thus, transposable elements do not get upregulated upon diminishing the population of PLD6-dependent small RNAs in HEK293T cells.

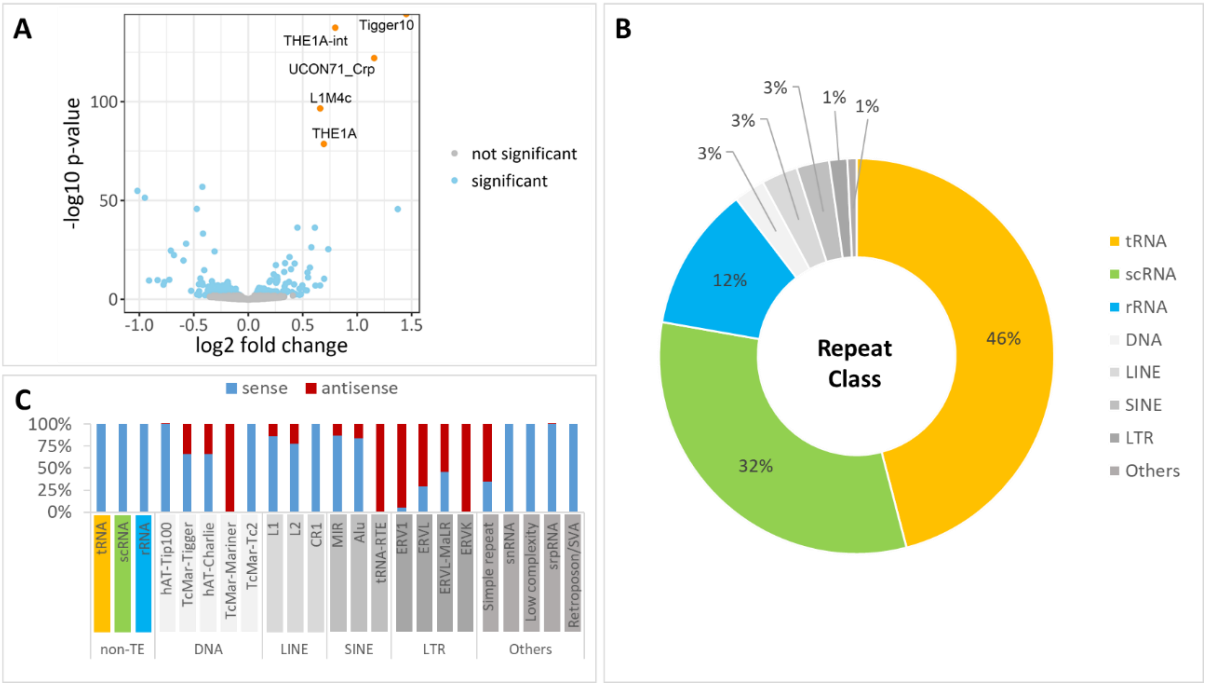


Figure 4.9: (A) Volcano plot of differential expression analysis for repetitive elements of PLD6-knockdown and control HEK293T cells. (B) Percentage share of differentially expressed repetitive elements to assigned classes. (C) Mapping orientation of PLD6-dependent small RNAs to the respective repetitive element classes.

4.4. Discussion

Since its discovery, the piRNA pathway is generally considered to be restricted to the germline and early developmental processes. Only few studies on non-vertebrate metazoans like *Aedes aegypti* and *Aplysia californica* could reliably prove an expression and function in gonadal-unrelated processes like virus-defense and memory-related synaptic plasticity (Miesen et al. 2015; Rajasethupathy et al. 2012). Recently, this speckled map of somatic piRNA existence was essentially extended by large-scale analyses of arthropods and mollusks (Lewis et al. 2018; Jehn et al. 2018). The two studies suggest that a somatic piRNA pathway targeting transposable elements existed in the ancestral protostome and is maintained by most protostomian species to day. As PIWI proteins were found to be expressed in somatic stem cells of nonbilaterian species as well (Funayama et al. 2010; Alié et al. 2011; Seipel et al. 2004; Juliano et al. 2014), it is probable that somatic function might have been a common state that later got lost in some species, especially in the branch of vertebrates. Despite this, some studies claim that a somatic function persist in vertebrates in few niches like the hippocampus (Nandi et al. 2016; Ross et al. 2014). However, data supporting this hypothesis only consists of favorable indications by bioinformatics analyses, while crushing *in vitro* evidence do not exist yet. One obstacle is that adequate antibodies against the PIWI proteins are not available. To circumvent this, we developed an RNAi-based approach to pin down small RNAs that depend on PLD6, which is the central player of piRNA biogenesis.

As a proof of principle, this method was first applied on HEK293T cells, which are easy to transfect and where PLD6 is reasonably expressed (data not shown). By siRNA-mediated

knockdown of PLD6, followed by small RNA sequencing we could identify a small RNA population that is diminished upon PLD6-knockdown in HEK293T cells (figure 4.3). As these PLD6-dependent small RNAs have the typical size range of piRNAs, PLD6 seems to act as an endonuclease also in the soma. This function could not be expected since several studies claimed that in somatic cells PLD6 and its mouse and fly homologues only act as phospholipases that degrade cardiolipin to phosphatidic acid thereby triggering mitochondrial fusion, while their endonucleolytic function is only active within the piRNA processing center of germline cells called nuage (Nureki 2014; Huang et al. 2011; Choi et al. 2006; Watanabe et al. 2011). A recent study even proposed a mechanism in which the phospholipase reactivity is induced by a conformational change due to the interaction with the two orthologues of *MIGA* (Zhang et al. 2016). Despite this, our small RNA sequencing data suggest that PLD6 executes its nucleolytic function also in somatic cells. Thus, our PLD6 knockdown approach might be applicable to identify potential somatic piRNAs. Although the PLD6-dependent small RNAs in HEK293T cells exhibit the typical piRNA length distribution, they mainly originate from different transcripts compared to piRNAs in the germline (figure 4.4). This illustrates an advantage our method has over homology-based bioinformatics approaches, as piRNAs in the soma might have alternative origins apart from the canonical germline piRNA clusters. Unlike germline-expressed piRNAs, PLD6-dependent RNAs in HEK293T cells do not have a 1U bias at their 5' end (figure 4.6A). As it was shown that PLD6 does not have a nucleotide preference *in vitro*, this might be the consequence of the absence of PIWI proteins or other components of the piRNA pathway machinery that would direct PLD6 to cleave before U (Ipsaro et al. 2012; Nishimasu et al. 2012). Furthermore, PLD6-dependent RNAs in HEK293T cells hardly show the piRNA-typical 2' O-methylation at their 3' end (low resistance to $NaIO_4$ treatment). However, those PLD6-dependent RNAs that are listed as piRNAs in the NCBI database have a relatively higher $NaIO_4$ resistance, suggesting that there is a residual piRNA maturation machinery in this artificial cell system (figure 4.6B). Even though HEK293T cells lack most of the relevant piRNA pathway components, our initial study in HEK293T cells suggests that our PLD6-knockdown and sequencing approach allows the identification of piRNAs in the soma.

Another advantage of our method is that it allows to assess the functional relevance of these putative piRNAs by studying knockdown phenotypes (e.g. via transcriptome sequencing). A limitation that comes along with the approach is that besides its function in piRNA biogenesis, PLD6 is also a major constituent of the lipid-signaling pathway of mitochondrial fusion and fission. Knocking down PLD6 therefore not only inhibits the production of potential piRNAs, but also disturbs the mitochondrial metabolism. This is also reflected in our transcriptome sequencing data of PLD6-knockdown and control cells, where most of the analyzed mitochondria-related genes are differentially expressed (figure 4.8B). Additionally, analysis of our transcriptome and small RNA sequencing data revealed that knocking down PLD6 seems to increase the abundancy of ribosomal constituents in HEK293T cells (figure 4.8A, figure 4.4). As noticed by Tanwar et al. 2016, genes involved in mitochondrial dynamics like Drp1 and Mfn2 are associated with changes in ribosomal pathways. Therefore, it is probable that the upregulation of structural constituents of the ribosome is a consequence of disturbing mitochondrial dynamics by knocking down PLD6. This needs to be considered, when analyzing

differential expression of protein-coding genes in order to identify potential targets of the putative piRNAs.

As transposable elements are the major targets of gonadal piRNAs, we checked the differential expression of repetitive elements in PLD6-knockdown and control cells. Only few repetitive elements were differentially expressed and most of them are repeats of structural RNAs rather than transposable elements (figure 4.9A+C). Thus, transposable elements are not de-repressed in HEK293T upon depleting the PLD6-dependent small RNA population. This is not surprising, since transposable elements are considered to be stably epigenetically silenced in somatic non-stem cells like HEK293T cells (Bourque et al. 2018). Furthermore, most of the PLD6-dependent small RNAs map in sense direction to the repetitive elements, indicating that these molecules are not generated in order to serve as transposon control (figure 4.9B).

Beyond transposon control, it was shown that the gonadal piRNA pathway is also involved in the posttranscriptional regulation of protein-coding genes (Rouget et al. 2010; Gou et al. 2014; Gebert et al. 2015; Zhang et al. 2015). We therefore analyzed the differential expression of protein-coding genes to check whether PLD6-dependent small RNAs have a similar function in HEK293T cells. Although many protein-coding genes were differentially expressed upon PLD6-knockdown, we did not identify an upregulation that could be explained by piRNA-like targeting of the PLD6-dependent small RNAs (figure 4.7B). Yet, we do not see this as a disproof of the explanatory power of our approach, as the lack of gene regulatory function of the PLD6-dependent small RNAs in HEK293T cells might be explained by the absence of PIWI proteins in this cell line. To evaluate the value of our method on a final basis, it is therefore essential to test the method on physiologically more relevant stem cells, where PLD6 and the PIWI proteins are substantially expressed (Sharma 2001; Wu et al. 2010; Nolde et al. 2013; Nandi et al. 2016). A good starting point might be primary cells of the human dentate gyrus, as adult neurogenesis and active retrotransposition is taking place in this particular subregion of the adult hippocampus (Eriksson et al. 1998; Boldrini et al. 2018; Muotri et al. 2005; Coufal et al. 2009).

4.5. Materials and Methods

4.5.1. PLD6-knockdown in HEK293T cells by antisense-siRNA transfection

HEK293T cells (5E4 cells/well) were seeded in 24-well plates and cultured in 1x GlutaMAX™-I DMEM supplemented with 10% FBS (Thermo Fisher). The next day the cells were transfected with 10 nM Silencer™ Select antisense-PLD6 No. 3 (4392420_s47323; Thermo Fisher) or with 10 nM Silencer™ Select Negative Control No. 1 siRNA (4390843; Thermo Fisher) using Lipofectamine™ RNAiMAX (Thermo Fisher) according to the manufacturer's protocol. 24 and 48 hours after transfection, the total RNA and proteins were isolated according to the TRI Reagent™ protocol (Thermo Fisher).

4.5.2. qPCR to validate PLD6-knockdown

In order to determine the relative mRNA concentration of PLD6 in the cultured cells, the total RNA was reverse transcribed to cDNA with SuperScript™ IV Reverse Transcriptase (Thermo Fisher) using random hexamers. Exon-spanning primers amplifying ~200 bp amplicons were designed for *PLD6* and *ACTB* (S-Table 4.2) using the NCBI tool Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). These primers were utilized to analyze the obtained cDNA as technical duplicates on a Corbett Rotor-Gene 6000 real-time PCR cycler. The copy numbers of the respective genes were quantified by standard curves of the individual primer pair amplicons and normalized by the housekeeping gene *ACTB* to calibrate for variabilities in sample preparation.

4.5.3. Western blot to validate PLD6-knockdown

The protein expression of PLD6 in the cultured cells was quantified in relation to β-actin by western blot analysis. Therefore, the protein-containing phenol phase of the TRI Reagent™ extraction was dialyzed at 4°C for 20 h, 4 h and finally 2 h in a MWCO 14000 cellulose dialysis tube (Carl Roth) against a 1% SDS solution. Afterwards Halt™ Protease Inhibitor Cocktail (Thermo Fisher) was added and the protein solution was concentrated at room temperature in 10K-Amicon columns (Merck Millipore) for 15 min at 14,000 x g. The protein concentration was determined with the Qubit™ Protein Assay Kit (Thermo Fisher) and 40 µg of total protein were applied on a SDS-polyacrylamide gel (4% stacking gel, 12.5% running gel) after they had been denaturated for 5 mins at 100°C in the presence of equal amounts of 2x Lämmli sample buffer. After electrophoretic separation, the proteins were blotted on a methanol-activated 0.45 µm Roti®-PVDF membrane (Carl Roth). The membrane was blocked in 5% milk blocking buffer and incubated with the primary antibodies for PLD6 (ab170183; Abcam) and β-actin (ab8226; Abcam). For signal detection with a Bio-Rad ChemiDoc™ Imaging System, HRP-conjugated secondary antibodies (ab6721; Abcam | ABIN101802; antikoerper-online.de) were added and a luminol-containing ECL solution was applied. Relative signal intensity was calculated with the supplied Image Lab™ software.

4.5.4. Small RNA sequencing, annotation and data analysis

After the knockdown efficacy was validated by qPCR and western blot, total RNA (m > 1 µg, c > 50 ng/µL) of PLD6-knockdown and non-target control HEK293T cells isolated 24 h after transfection was sent to BGI in Hong Kong for library preparation and 50 bp single-end sequencing of the small RNA population. On average 86 million single-end reads were obtained, where adapter sequences had been already clipped off. After conversion to the fasta-format by *TBr2_fastq2fasta.pl*, sequences were filtered for length (15-40 nt) by *TBr2_length-filter.pl*, collapsed to non-identical reads by *TBr2_collapse.pl* and depleted for low-complexity sequences by *TBr2_duster.pl*. Furthermore, reads obtained for the knockdown cells that contain the sequence of the transfected PLD6-siRNA were removed by *RemovePLD6-*

siRNA-Nr3.pl. The remaining reads were mapped to the human genome (Ensembl, GRCh38.p11) by *sRNAmapper.pl* (version 1.0.4), while setting the parameters to keep only the best alignments. Using *unitas.pl* (version 1.0.0), the mapped sequences were annotated and differential expression of small RNAs was analyzed for an estimated deviation of 10% among replicates. The used Perl scripts, except *for RemovePLD6-siRNA-Nr3.pl*, are part of the NGS toolbox (version 2.1).

4.5.5. qPCR to validate small RNA sequencing analysis and check for 2'OMe-modification at 3' end

To crosscheck the differential expression revealed by small RNA sequencing, we selected 14 small RNAs to measure their relative expression using qPCR. These 14 small RNAs cover the whole length spectrum and were selected as they have relatively high read counts in the control library. To be able to quantify the small RNAs via qPCR without disturbing precursors, the small RNA fraction was size-separated (15-40 nt) by a urea-polyacrylamide gel. The small RNAs were eluted from the gel by shaking the crushed gel slices in RNase-free water in the presence of RiboLock™ RNase Inhibitor (Thermo Fisher) for 2 h at 37°C. After removing the gel pieces with a 0.45 µm Ultrafree-MC column (Merck Millipore), the small RNAs were washed twice with RNase-free water using 3K-Amicon columns (Merck Millipore). In order to provide an annealing site for a reverse primer, the purified small RNAs were then polyadenylated at their 3' end using the A-Plus Poly(A) Polymerase Tailing Kit (CellScript). After ethanol precipitation, the polyadenylated small RNAs were reverse transcribed with the SuperScript™ IV reverse transcriptase (Thermo Fisher) using the "RT-PolyT" primer. Using the respective small RNA sequence as forward primer and "PCR against RT-PolyT" as reverse primer, the cDNA copies of the 14 small RNAs were quantified as technical duplicates on a Corbett Rotor-Gene 6000 real-time PCR cycler using the standard curve method (for primer sequences refer to S-Table 4.2). To normalize for sample quantity the copy numbers were relativized by the β-actin copy numbers determined during knockdown-validation. To check for potential 2'-O-methylation of the 3' ends, eight of these 14 small RNAs were submitted to sodium periodate treatment (like described in Rajasethupathy et al. 2012) prior to polyadenylation. By this approach, the sugar of the 3' end nucleoside of small RNAs without the protective modification is broken up and therefore not accessible for the addition of the polyA-tail necessary for qPCR amplification. Apart from the sodium periodate treatment the quantification procedure was unchanged.

4.5.6. RNA sequencing, data processing and differential expression analysis

Total RNA isolated 48 h after transfection from PLD6-knockdown and non-target control HEK293T cells (n=3) was send for library construction and paired-end sequencing to BGI (Hong Kong). On average 35 million paired-end reads were obtained. Using the online platform Galaxy (usegalaxy.org) the reads were first mapped to the human genome (Galaxy hg38) by

RNA STAR (Galaxy Tool Version 2.6.0b-1). Afterwards, gene wise counting was performed with featureCounts (Galaxy Tool Version 1.6.0.6) on basis of an Ensemble GTF-file (Homo_sapiens.GRCh38.90), which had been converted to UCSC coordinates using the File Chameleon tool of Ensembl. Based on the generated count tables, DESeq2 (Galaxy Tool Version 2.11.40.2) was used to identify significantly differentially expressed genes (adjusted p-value < 0.01). Volcano plots based on the DESeq2 result file were created with R (version 3.4.3). In order to check whether the expression of repetitive sequences would go up upon PLD6-knockdown, we repeated the differential expression analysis this time allowing the reads to map to multiple locations in the genome (Maximum number of alignments to output a read's alignment results: 100). We then fractionally counted mapping events for repetitive elements based on a GTF file containing the coordinates of repeatmasker-identified repetitive elements ("hg38_rmsk_TE.gtf.gz", version 18-Mar-2018 15:32, downloaded from labshare.cshl.edu/shares/mhammelllab/www-data/TEToolkit/TE_GTF/) with featureCounts and likewise analyzed the differential expression with DESeq2 and R.

4.5.7. Evaluation of piRNA-like gene regulation and GO term analysis

In order to check for genes that might be higher expressed in the PLD6-knockdown cells due to the loss of regulating PLD6-dependent small RNAs, we used the self-developed perl script *piRanha.pl*. By using the option *-use_scores*, piRanha calculated a score for each transcript that reflects how likely it is that the transcript is a target of the PLD6-dependent small RNAs according to the piRNA targeting rules (Zhang et al. 2018) including the number of target sites and the differential expression of the small RNAs. The score is calculated as described in figure 4.10. To test for an enrichment of differential expression, cumulative functions of log2FC values were plotted for the subset of potential target genes of PLD6-dependent small RNAs (piRanha score > 500) in comparison to all quantified genes. Additionally, GO term analysis for significantly differentially up- and downregulated genes was performed with the Panther functional classification tool (http://pantherdb.org/).

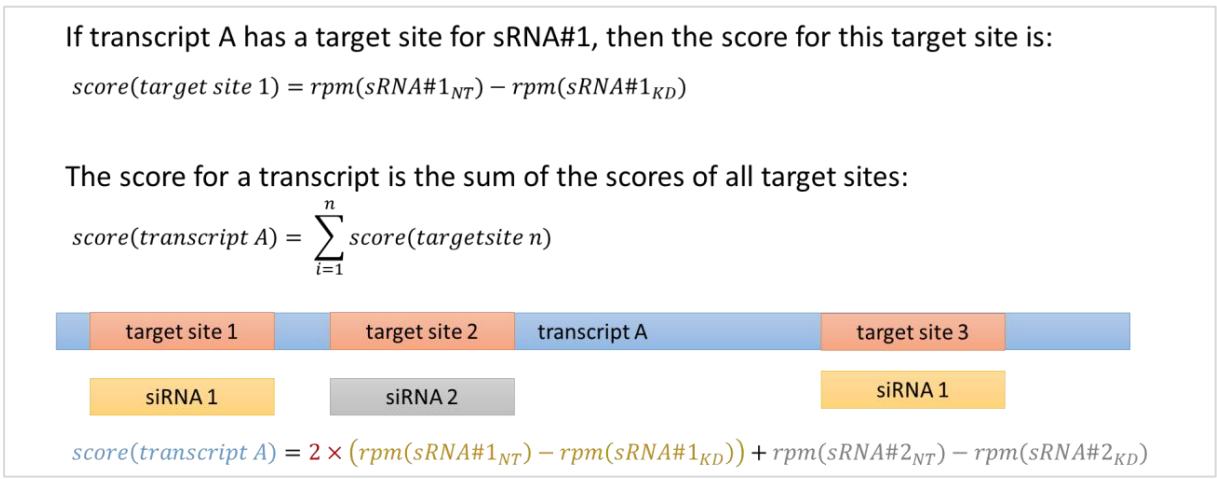


Figure 4.10: Exemplary representation of how piRanha scores were calculated.

4.6. Author Contributions

DR designed the study. **JJ** and JSTK established the siRNA transfection conditions and analyzed the small RNA sequencing data. **JJ** characterized the PLD6-dependent small RNAs further and analyzed the RNA sequencing data. **JJ** wrote this chapter.

4.7. Declarations and Acknowledgements

Sequencing datasets have been uploaded to NCBI's Sequence Read Archive (SRA) and can be accessed via the BioProject accessions PRJNA564396 (sRNAseq data) and PRJNA564901 (RNAseq data). The used Perl scripts are freely available at GitHub (github.com/jjehn/PLD6-dependent-sRNAs) or are as part of the NGS toolbox (version 2.1) available from http://www.smallrnagroup.uni-mainz.de/.

We thank the Wolfrum lab for kindly providing HEK293T cells.

4.8. References

Alié, Alexandre; Leclère, Lucas; Jager, Muriel; Dayraud, Cyrielle; Chang, Patrick; Le Guyader, Hervé et al. (2011): Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: ancient association of "germline genes" with stemness. In *Developmental biology* 350 (1), pp. 183–197. DOI: 10.1016/j.ydbio.2010.10.019.

Boldrini, Maura; Fulmore, Camille A.; Tartt, Alexandria N.; Simeon, Laika R.; Pavlova, Ina; Poposka, Verica et al. (2018): Human Hippocampal Neurogenesis Persists throughout Aging. In *Cell stem cell* 22 (4), 589-599.e5. DOI: 10.1016/j.stem.2018.03.015.

Bourque, Guillaume; Burns, Kathleen H.; Gehring, Mary; Gorbunova, Vera; Seluanov, Andrei; Hammell, Molly et al. (2018): Ten things you should know about transposable elements. In *Genome biology* 19 (1), p. 199. DOI: 10.1186/s13059-018-1577-z.

Choi, Seok-Yong; Huang, Ping; Jenkins, Gary M.; Chan, David C.; Schiller, Juergen; Frohman, Michael A. (2006): A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. In *Nature cell biology* 8 (11), pp. 1255–1262. DOI: 10.1038/ncb1487.

Coufal, Nicole G.; Garcia-Perez, José L.; Peng, Grace E.; Yeo, Gene W.; Mu, Yangling; Lovci, Michael T. et al. (2009): L1 retrotransposition in human neural progenitor cells. In *Nature* 460 (7259), pp. 1127–1131. DOI: 10.1038/nature08248.

Dharap, Ashutosh; Nakka, Venkata Prasuja; Vemuganti, Raghu (2011): Altered expression of PIWI RNA in the rat brain after transient focal ischemia. In *Stroke* 42 (4), pp. 1105–1109. DOI: 10.1161/STROKEAHA.110.598391.

Eriksson, P. S.; Perfilieva, E.; Björk-Eriksson, T.; Alborn, A. M.; Nordborg, C.; Peterson, D. A.; Gage, F. H. (1998): Neurogenesis in the adult human hippocampus. In *Nature medicine* 4 (11), pp. 1313–1317. DOI: 10.1038/3305.

Funayama, Noriko; Nakatsukasa, Mikiko; Mohri, Kurato; Masuda, Yoshiki; Agata, Kiyokazu (2010): Piwi expression in archeocytes and choanocytes in demosponges: insights into the stem cell system in demosponges. In *Evolution & development* 12 (3), pp. 275–287. DOI: 10.1111/j.1525-142X.2010.00413.x.

Gebert, Daniel; Ketting, René F.; Zischler, Hans; Rosenkranz, David (2015): piRNAs from Pig Testis Provide Evidence for a Conserved Role of the Piwi Pathway in Post-Transcriptional Gene Regulation in Mammals. In *PloS one* 10 (5), e0124860. DOI: 10.1371/journal.pone.0124860.

Ghosheh, Yanal; Seridi, Loqmane; Ryu, Taewoo; Takahashi, Hazuki; Orlando, Valerio; Carninci, Piero; Ravasi, Timothy (2016): Characterization of piRNAs across postnatal development in mouse brain. In *Scientific reports* 6, p. 25039. DOI: 10.1038/srep25039.

Gou, Lan-Tao; Dai, Peng; Yang, Jian-Hua; Xue, Yuanchao; Hu, Yun-Ping; Zhou, Yu et al. (2014): Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. In *Cell research* 24, 680 EP -. DOI: 10.1038/cr.2014.41.

Han, Bo W.; Wang, Wei; Li, Chengjian; Weng, Zhiping; Zamore, Phillip D. (2015): piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. In *Science (New York, N.Y.)* 348 (6236), pp. 817–821. DOI: 10.1126/science.aaa1264.

Huang, Huiyan; Gao, Qun; Peng, Xiaoxue; Choi, Seok-Yong; Sarma, Krishna; Ren, Hongmei et al. (2011): piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. In *Developmental cell* 20 (3), pp. 376–387. DOI: 10.1016/j.devcel.2011.01.004.

Ipsaro, Jonathan J.; Haase, Astrid D.; Knott, Simon R.; Joshua-Tor, Leemor; Hannon, Gregory J. (2012): The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. In *Nature* 491 (7423), pp. 279–283. DOI: 10.1038/nature11502.

Jehn, Julia; Gebert, Daniel; Pipilescu, Frank; Stern, Sarah; Kiefer, Julian Simon Thilo; Hewel, Charlotte; Rosenkranz, David (2018): PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. In *Communications biology* 1, p. 137. DOI: 10.1038/s42003-018-0141-4.

Juliano, Celina E.; Reich, Adrian; Liu, Na; Götzfried, Jessica; Zhong, Mei; Uman, Selen et al. (2014): PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (1), pp. 337–342. DOI: 10.1073/pnas.1320965111.

Lee, Eun Joo; Banerjee, Sourav; Zhou, Hongjun; Jammalamadaka, Aruna; Arcila, Mary; Manjunath, B. S.; Kosik, Kenneth S. (2011): Identification of piRNAs in the central nervous system. In *RNA (New York, N.Y.)* 17 (6), pp. 1090–1099. DOI: 10.1261/rna.2565011.

Lewis, Samuel H.; Quarles, Kaycee A.; Yang, Yujing; Tanguy, Melanie; Frézal, Lise; Smith, Stephen A. et al. (2018): Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. In *Nature ecology & evolution* 2 (1), pp. 174–181. DOI: 10.1038/s41559-017-0403-4.

Marchetto, Maria C. N.; Narvaiza, Iñigo; Denli, Ahmet M.; Benner, Christopher; Lazzarini, Thomas A.; Nathanson, Jason L. et al. (2013): Differential L1 regulation in pluripotent stem cells of humans and apes. In *Nature* 503 (7477), pp. 525–529. DOI: 10.1038/nature12686.

Miesen, Pascal; Girardi, Erika; van Rij, Ronald P. (2015): Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. In *Nucleic acids research* 43 (13), pp. 6545–6556. DOI: 10.1093/nar/gkv590.

Mohn, Fabio; Handler, Dominik; Brennecke, Julius (2015): Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. In *Science (New York, N.Y.)* 348 (6236), pp. 812–817. DOI: 10.1126/science.aaa1039.

Muotri, Alysson R.; Chu, Vi T.; Marchetto, Maria C. N.; Deng, Wei; Moran, John V.; Gage, Fred H. (2005): Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. In *Nature* 435 (7044), pp. 903–910. DOI: 10.1038/nature03663.

Nandi, Sayan; Chandramohan, Dhruva; Fioriti, Luana; Melnick, Ari M.; Hébert, Jean M.; Mason, Christopher E. et al. (2016): Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. In *Proceedings of the National Academy of Sciences of the United States of America* 113 (45), pp. 12697–12702. DOI: 10.1073/pnas.1609287113.

Nishimasu, Hiroshi; Ishizu, Hirotsugu; Saito, Kuniaki; Fukuhara, Satoshi; Kamatani, Miharu K.; Bonnefond, Luc et al. (2012): Structure and function of Zucchini endoribonuclease in piRNA biogenesis. In *Nature* 491 (7423), pp. 284–287. DOI: 10.1038/nature11509.

Nureki, Osamu (2014): Is zucchini a phosphodiesterase or a ribonuclease? In *Biomedical journal* 37 (6), pp. 369–374. DOI: 10.4103/2319-4170.132909.

Ortogero, Nicole; Schuster, Andrew S.; Oliver, Daniel K.; Riordan, Connor R.; Hong, Annie S.; Hennig, Grant W. et al. (2014): A novel class of somatic small RNAs similar to germ cell pachytene PIWI-interacting small RNAs. In *The Journal of biological chemistry* 289 (47), pp. 32824–32834. DOI: 10.1074/jbc.M114.613232.

Phay, Monichan; Kim, Hak Hee; Yoo, Soonmoon (2018): Analysis of piRNA-Like Small Non-coding RNAs Present in Axons of Adult Sensory Neurons. In *Molecular neurobiology* 55 (1), pp. 483–494. DOI: 10.1007/s12035-016-0340-2.

Rajasethupathy, Priyamvada; Antonov, Igor; Sheridan, Robert; Frey, Sebastian; Sander, Chris; Tuschl, Thomas; Kandel, Eric R. (2012): A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. In *Cell* 149 (3), pp. 693–707. DOI: 10.1016/j.cell.2012.02.057.

Rizzo, Francesca; Hashim, Adnan; Marchese, Giovanna; Ravo, Maria; Tarallo, Roberta; Nassa, Giovanni et al. (2014): Timed regulation of P-element-induced wimpy testis-interacting RNA expression during rat liver regeneration. In *Hepatology (Baltimore, Md.)* 60 (3), pp. 798–806. DOI: 10.1002/hep.27267.

Ross, Robert J.; Weiner, Molly M.; Lin, Haifan (2014): PIWI proteins and PIWI-interacting RNAs in the soma. In *Nature* 505 (7483), pp. 353–359. DOI: 10.1038/nature12987.

Rouget, Christel; Papin, Catherine; Boureux, Anthony; Meunier, Anne-Cécile; Franco, Bénédicte; Robine, Nicolas et al. (2010): Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. In *Nature* 467 (7319), pp. 1128–1132. DOI: 10.1038/nature09465.

Seipel, Katja; Yanze, Nathalie; Schmid, Volker (2004): The germ line and somatic stem cell gene Cniwi in the jellyfish Podocoryne carnea. In *Int. J. Dev. Biol.* 48 (1), pp. 1–7. DOI: 10.1387/ijdb.15005568.

Tanwar, Deepak Kumar; Parker, Danitra J.; Gupta, Priyanka; Spurlock, Brian; Alvarez, Ronald D.; Basu, Malay Kumar; Mitra, Kasturi (2016): Crosstalk between the mitochondrial fission protein, Drp1, and the cell cycle is identified across various cancer types and can impact survival of epithelial ovarian cancer patients. In *Oncotarget* 7 (37), pp. 60021–60037. DOI: 10.18632/oncotarget.11047.

Tosar, Juan Pablo; Rovira, Carlos; Cayota, Alfonso (2018): Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. In *Communications biology* 1, p. 2. DOI: 10.1038/s42003-017-0001-7.

Upton, Kyle R.; Gerhardt, Daniel J.; Jesuadian, J. Samuel; Richardson, Sandra R.; Sánchez-Luque, Francisco J.; Bodea, Gabriela O. et al. (2015): Ubiquitous L1 mosaicism in hippocampal neurons. In *Cell* 161 (2), pp. 228–239. DOI: 10.1016/j.cell.2015.03.026.

Watanabe, Toshiaki; Chuma, Shinichiro; Yamamoto, Yasuhiro; Kuramochi-Miyagawa, Satomi; Totoki, Yasushi; Toyoda, Atsushi et al. (2011): MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. In *Developmental cell* 20 (3), pp. 364–375. DOI: 10.1016/j.devcel.2011.01.005.

Wu, Qiuling; Ma, Qi; Shehadeh, Lina A.; Wilson, Amber; Xia, Linghui; Yu, Hong; Webster, Keith A. (2010): Expression of the Argonaute protein PiwiL2 and piRNAs in adult mouse mesenchymal stem cells. In *Biochemical and biophysical research communications* 396 (4), pp. 915–920. DOI: 10.1016/j.bbrc.2010.05.022.

Yan, Zheng; Hu, Hai Yang; Jiang, Xi; Maierhofer, Vera; Neb, Elena; He, Liu et al. (2011): Widespread expression of piRNA-like molecules in somatic tissues. In *Nucleic acids research* 39 (15), pp. 6596–6607. DOI: 10.1093/nar/gkr298.

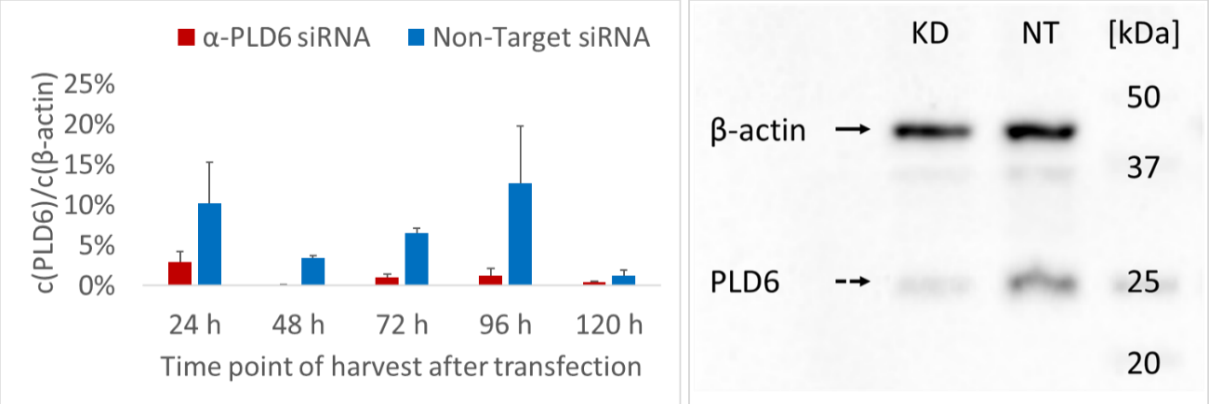
Zhang, Donglei; Tu, Shikui; Stubna, Michael; Wu, Wei-Sheng; Huang, Wei-Che; Weng, Zhiping; Lee, Heng-Chi (2018): The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. In *Science (New York, N.Y.)* 359 (6375), pp. 587–592. DOI: 10.1126/science.aao2840.

Zhang, Peng; Kang, Jun-Yan; Gou, Lan-Tao; Wang, Jiajia; Xue, Yuanchao; Skogerboe, Geir et al. (2015): MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. In *Cell research* 25 (2), pp. 193–207. DOI: 10.1038/cr.2015.4.

Zhang, Ying; Li, Jianchao; Chen, Rong; Dai, Aiqin; Luan, Deqin; Ma, Teng et al. (2013): Cloning, characterization and widespread expression analysis of testicular piRNA-like chicken RNAs. In *Molecular biology reports* 40 (4), pp. 2799–2807. DOI: 10.1007/s11033-012-2295-3.

Zhang, Yongping; Liu, Xiaoman; Bai, Jian; Tian, Xuejun; Zhao, Xiaocui; Liu, Wei et al. (2016): Mitoguardin Regulates Mitochondrial Fusion through MitoPLD and Is Required for Neuronal Homeostasis. In *Molecular cell* 61 (1), pp. 111–124. DOI: 10.1016/j.molcel.2015.11.017.

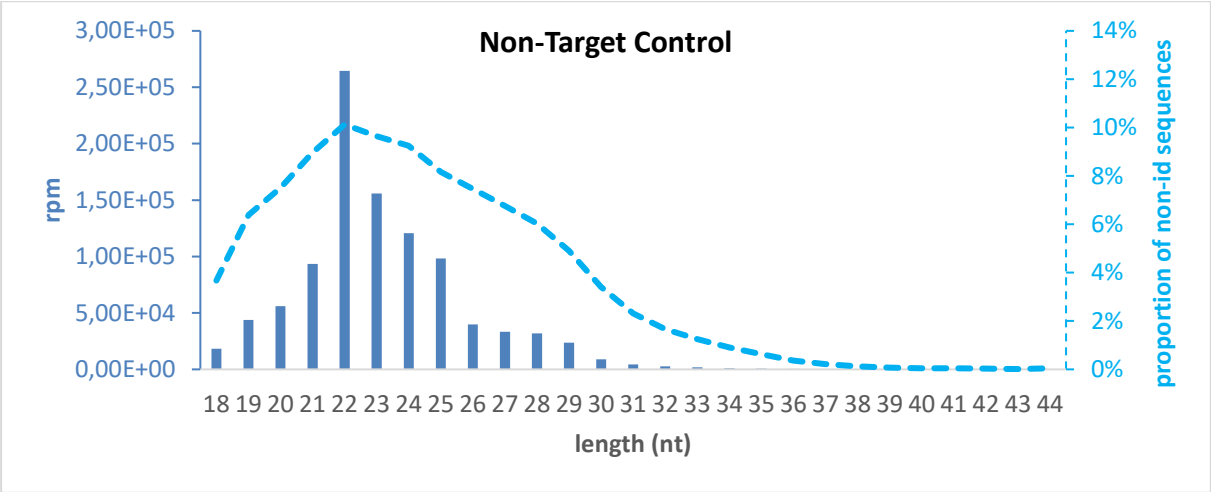
4.9. Supplement



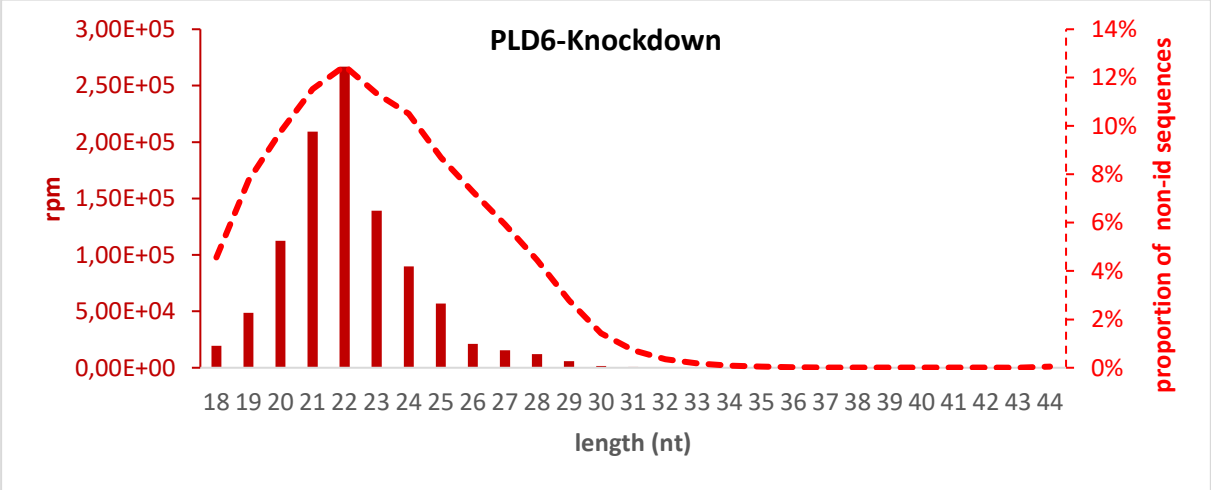
S-Figure 4.1: Validation of PLD6-knockdown on the transcript level via qPCR (left) and on the protein level via western blot (right).

S-Table 4.1: Numerical summary of the output generated by the respective analysis steps.

Applied processes	**Non-Target Control**	**PLD6-Knockdown**
Clean Reads BGI	82,491,565 reads (3,415,186 non-id)	89,592,331 reads (2,937,180 non-id)
Fastq2fasta	Not affected	Not affected
Length-filter 15-40 nt	82,485,954 reads (3,410,868 non-id)	89,590,131 reads (2,935,278 non-id)
Collapse	Not affected	Not affected
Duster	3,405,257 non-id	2,929,545 non-id
RemovePLD6-siRNA-Nr3	Not affected	2,908,766 non-id
sRNA-mapper (`-a best`)	66,711,467 reads (1,173,992 non-id)	59,241,468 reads (1,059,552 non-id)
Unitas-filter	66,235,564 reads	58,784,723 reads
Unitas diffexpr (`-diffexpr_estdev 0.1`)	2,704,936 reads (899 non-id)	249,829 reads (172 non-id)



S-Figure 4.2: Length profile of small RNA reads sequenced from HEK293T cells transfected with a non-target-control-siRNA (control). The dashed line depicts the proportion of non-identical sequences being represented by the reads of a specific length.



S-Figure 4.3: Length profile of small RNA reads sequenced from HEK293T cells transfected with an anti-PLD6-siRNA (PLD6-KD). The dashed line depicts the proportion of non-identical sequences being represented by the reads of a specific length.

S-Table 4.2: Sequences of oligonucleotide ordered from biomers.net.

Name	Stock	Sequence 5' -> 3'	Gene	Design by
PLD6 fwd	1711	ATCTCTGCCTGTTCGCCTT	*PLD6*	Julian Kiefer
PLD6 rev	1706	CCACGATGGCAAACTTGTGA	*PLD6*	Julian Kiefer
ACTB fwd	1715	CGAGCACAGAGCCTCGCCTTT	*ACTB*	Julian Kiefer
ACTB rev	1716	CATGCCCACCATCACGCCCTGG	*ACTB*	Julian Kiefer

Name	Stock	Sequence 5' -> 3'	Gene	Design by
f1	1730	AGCTACATCTGGCTACTGGGTCTCAA	*MIR222*	Julia Jehn
f2	1731	ATCTGTGATGATCTTATCCCGAACCTGAA	*SNORD50A*	Julia Jehn
f3	1732	TCCCTGGTGGTCTAGTGGTTAGGATTCGG	*LncRNA /tRNA-Glu*	Julia Jehn
f4	1727	GGCTGGTCCGAAGGTAGTGAGTTATCTCAAT	*RNY1*	Julia Jehn
f5	1728	TCTGGTGATGAAATGGAACGTTTCTGATG	*IL37/DIAPH 1*	Julia Jehn
f6	1729	ACCGCCTGGGAATACCGGGTGCTGTAGGCTT	*RNA5SP358*	Julia Jehn
f7	1724	CTTCGTGATCGATGTGGTGACGTCGTGCTC	*45S pre-ribosomal 5*	Julia Jehn
f8	1725	AGAGGGTCTTTTTCACCCCGCTGTTGCTCTT	*Gly-tRNA*	Julia Jehn
f9	1726	AATGTGTGACTGAAAGGTATTTTCTGAGC	*SNORD31*	Julia Jehn
f10	1721	CGAGAGGGGCTGTGCTCGCAAGGTTTCTT	*SLC6A6*	Julia Jehn
f11	1722	ATTAATGATGAGATATAACCTTGACTGAAG	*SNORD119*	Julia Jehn
f12	1723	GTCAATGATGAATGGTAAAAGGTCTGAGT	*piR-36338 /SCARNA6*	Julia Jehn
f13	1718	GTGGGGTTCGTTTTCGGGCATGAAAATTT	*tRNA-Arg*	Julia Jehn
f14	1719	TTCGATGAAGAGATGATGACGAGTCTGACT	*piRNA piR-60668*	Julia Jehn
PCR against RT-PolyT	RNA #4	CGAATTCTAGAGCTCGAGGCAGG		David Rosenkranz
RT-PolyT	RNA #3	CGAATTCTAGAGCTCGAGGCAGGCGACATGT$_{25}$VN		David Rosenkranz

5. Selection for extreme secondary structure in coding sequences of viruses - An evolutionary arms race borne defense mechanism?

Daniel Gebert[1], Julia Jehn[1], David Rosenkranz[1]

[1]Institute of Organismic and Molecular Evolution, Anthropology, Johannes Gutenberg University Mainz, Anselm-Franz-von-Bentzel-Weg 7, 55099 Mainz, Germany

This chapter was published as a Research Article in Open Biology under the title "Widespread selection for extremely high and low levels of secondary structure in coding sequences across all domains of life" (Gebert, Jehn and Rosenkranz, Open Biology 2019 9:190020). Numbering of figures and tables as well as the citation style have been adjusted.

5.1. Abstract

Codon composition, GC-content and local RNA secondary structures can have a profound effect on gene expression and mutations affecting these parameters, even though they do not alter the protein sequence, are not neutral in terms of selection. Although evidence exists that in some cases selection favors more stable RNA secondary structures, we currently lack a concrete idea of how many genes are affected within a species, and if this is a universal phenomenon in nature. We searched for signs of structural selection in a global manner, analyzing a set of one million coding sequences from 73 species representing all domains of life, as well as viruses, by means of our newly developed software PACKEIS. We show that codon composition and amino acid identity are main determinants of RNA secondary structure. In addition, we show that the arrangement of synonymous codons within coding sequences is non-random, yielding extremely high, but also extremely low RNA structuredness significantly more often than expected by chance. Together, we demonstrate that selection for high and low levels of secondary structure is a widespread phenomenon. Our results provide another line of evidence that synonymous mutations are less neutral than commonly thought, which is of importance for many evolutionary models.

5.2. Introduction

The genetic code of DNA uses units of three nucleotides (codons) to code for one amino acid. Since the number of possible codons exceeds the number of proteogenic amino acids, most amino acids are encoded not by a single but several different codons. Therefore, mutations on the DNA level do not necessarily result in an altered amino acid sequence of the corresponding protein. These silent (synonymous) substitutions have long been assumed to be neutral in terms of natural selection (Kimura 1977). However, silent substitutions will necessarily result in altered codon composition of a gene and further have the potential to alter a gene's GC-content, both being features that can indeed be subject to selection (Sharp

et al. 1995). Moreover, silent substitutions can change the secondary structure of an mRNA, thereby affecting the process of translation (Babendure et al. 2006; Kudla et al. 2009; Mao et al. 2014; Huang et al. 2019) and non-random patterns of secondary structures within protein coding genes in different species have been explained by natural selection (Katz and Burge 2003; Chamary and Hurst 2005; Hoede et al. 2006; Fricke et al. 2018).

However, the currently available data does not allow to assess whether selection that acts on secondary structures within coding sequences represents a peculiarity of a few genomic loci in a limited number of species, or rather a widespread phenomenon, affecting many genes in species throughout the domains of life. It is further unknown, if selection acts only in one direction, favoring strong secondary structures as suggested by previous studies, or alternatively yields extremes at both ends of the spectrum.

To address these issues, we analyzed protein coding sequences of 73 species representing all domains of life and further included more than 240 thousand non-identical viral coding sequences. Using our newly developed software PACKEIS, we compared the predicted secondary structure of the evolutionary realized variants with that of corresponding artificial coding sequences that could have been realized in order to encode the same peptide sequence. We show that codon usage and amino acid identity both massively influence secondary structures. Beyond that, we identified protein coding sequences that exhibit extremely high or low structuring, compared to their artificial counterparts, independent of altered GC-content or codon usage, which is due to a non-random arrangement of synonymous codons within the coding sequence. Importantly, these extreme solutions occur significantly more often than we would expect when assuming absence of selection that favors structural extremes (structural selection). For the species under examination, we conservatively evaluate the fraction of protein coding sequences being subject to structural selection at on average 2-3%. We propose that altered structures of coding sequences affect a transcript's stability and/or its translation efficiency, which in turn are traits that evolution can act on, this way yielding the unexpectedly high number of structural extremes. A remarkably high number of coding sequences under structural selection was found in RNA viruses and we speculate that small RNA based host immune systems have exerted a selective regime on viral genomes favoring highly backfolded transcripts to avoid targeting by anti-viral small RNAs (Ding and Voinnet 2007; Ding and Lu 2011).

5.3. Results

5.3.1. Approach and software development

Prior to our actual survey, we had to develop a software tool that allowed us to assess whether or not a realized open reading frame (original ORF, oORF) represents an extreme solution in terms of backfolding (base pairing with itself through self-complementarity), considering the alternative ORFs (aORFs) that could have been realized in order to encode the given peptide sequence based on usage of synonymous codons. To this end, we have developed the highly

parallelizable software PACKEIS which compares the degree of backfolding (DBF) of the oORFs with that of a defined number of aORFs, this way yielding a DBF-score ranging from 0 to 1, referring to the degree of backfolding in the light of alternative codon usage with 0 representing extremely low structuredness and 1 representing extremely high structuredness. Details on the PACKEIS algorithm can be found in the Methods section.

5.3.2. ORFs exhibit extreme structures more often than expected by chance

Initially we speculated that particularly ORFs of viral genes may exhibit high levels of backfolding in order to escape small RNA based anti-viral responses of host immune systems. A piRNA-based immunity against viruses has been described in mosquito species such as *Aedes aegyptii*, and virus derived siRNAs that confer antiviral immunity can be found in plants, mosquitoes as well as in mice and human somatic cells (Ding and Voinnet 2007; Ding and Lu 2011; Qiu et al. 2017; Varjak et al. 2018; Huang and Li 2018). We thus started our analyses with ORFs of polyproteins from 13 human-pathogenic mosquito-borne viruses (Powell 2018). Contrasting our expectation, we found that only the Edge Hill virus ORF exhibits DBF-scores that imply a significant high degree of backfolding (DBF-score$_{model0}$=0.99, DBF-score$_{model2}$=1.00) while the ORFs of the other tested viral polyproteins have unremarkable DBF-scores in the range of >0.05 to <0.95 (figure 5.1A).

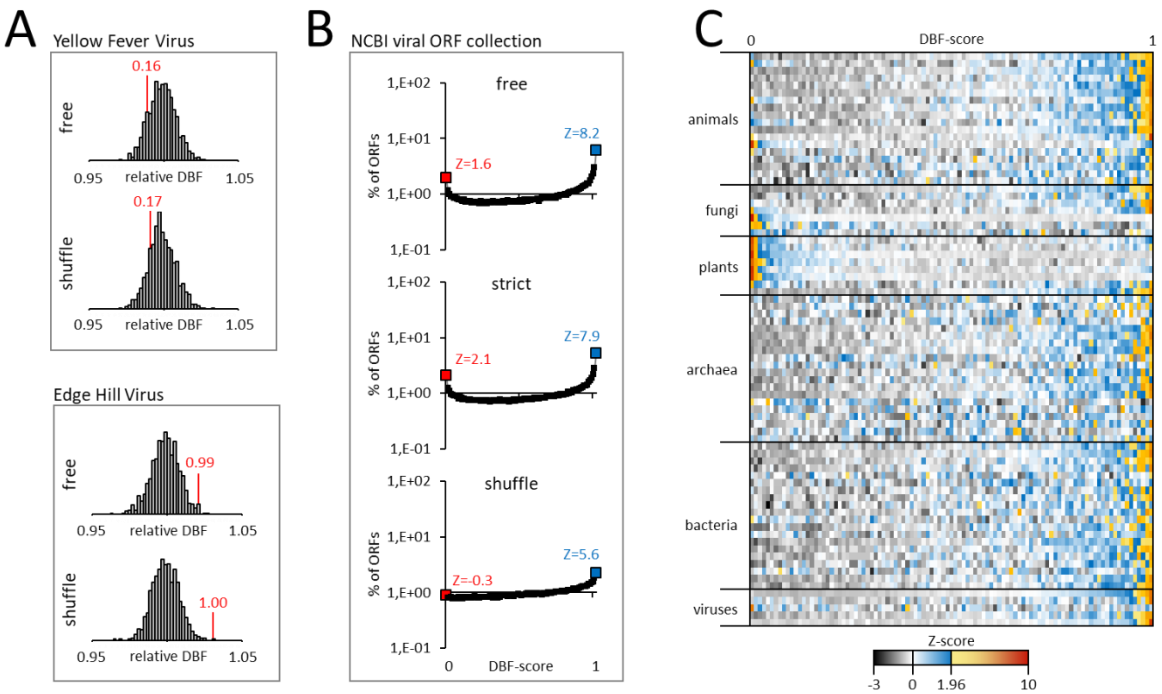


Figure 5.1: (A) DBF of aORFs from mosquito-borne RNA viruses for the purpose of illustration. X-axis shows the distribution of DBFs relative to the average value of all aORFs. Genomes of the Yellow Fever virus and the Edge Hill virus genomes both encode a single polyprotein. DBFs and the corresponding DBF-scores are indicated in red. Only the ORF of the Edge Hill virus shows an exceptional DBF which is considerably higher compared to the corresponding aORFs (DBF-score$_{model0}$=0.99, DBF-score$_{model2}$=1.00). (B) The analysis of DBF-scores for all available

viral ORF sequences reveals a consistent and significant enrichment for extremely high DBF-scores. (C) Lines in the heatmap represent species, rows represent DBF-scores from 0 to 1 in steps of 0.01 using model2 (shuffle). The color indicates row Z-scores with Z-scores above 1.96 ($p<0.05$ for the two-tailed hypothesis) indicated in shades from yellow to red. Lines for viruses represent dsDNA-, dsRNA, ssDNA-, ssRNA(-)- and ssRNA(+)-viruses.

In order to assess whether at all ORFs with extremely high or low DBF-scores occur significantly more often than expected by chance, we extended our analyses to a complete collection of non-identical viral ORFs (n=244,314) that are deposited at NCBI's GenBank sequence database. Indeed, we observed a significant enrichment for ORFs with high DBF-scores (figure 5.1B). While we would expect that under neutral conditions each DBF-score from 0 to 1 in steps of 0.01 accounts for roughly the same amount (1/101) of analyzed ORFs, the fraction of ORFs with a DBF-score of 1 accounts for 6.2%, 5.3% and 2.3% of all viral ORFs according to model0, model1 and model2, respectively (figure 5.1B). Despite the significant enrichment for ORFs with high DBF-scores, the fact that the large majority of ORFs showed no signs of selection for strong secondary structures casted doubt on our initial speculation that small RNA based immune responses have exerted strong selective regimes favoring highly backfolded ORFs.

Alternative explanations for favoring extreme levels secondary structures base on altered RNA stability and altered translation efficiency or a tradeoff between both, which would be a fundamental principal whose footprints should be present in all organisms. We thus sampled 73 representatives from all domains of life and calculated DBF-scores of their ORFs in a genome wide manner (electronic supplementary material, table S1). We found that applying model2 yielded the most modest results and we will thus refer to the results that base on model2 in the following, in order to give conservative estimates on the number of ORFs that we assume to be under structural selection, and to exclude effects related to GC content and biased codon usage. For the entirety of analyzed aORFs the fraction of paired bases relative to the oORF follows a Gaussian distribution ranging from 91% to 109% compared to the oORF. For 59 species we observed a significant enrichment of ORFs with DBF-score$_{model2}$>0.95 (figure 5.1C). Of these, nine species additionally showed a significant enrichment of ORFs with DBF-score$_{model2}$<0.05. Nine species exhibited only a significant enrichment of ORFs with DBF-score$_{model2}$<0.05 but not ORFs with DBF-score$_{model2}$>0.95. For the remaining five species, neither significant enrichment for extremely high nor low DBF-scores could be observed (figure 5.1C; electronic supplementary material, table S1). When comparing the DBF-scores of homologous genes across different species we did not find significant correlations, suggesting fluctuating structural selective forces on homologous genes along different phylogenetic branches (electronic supplementary material, table S2). We did further find no correlation between DBFscores$_{model2}$ and gene expression, neither on the transcript nor the protein level (electronic supplementary material, table S3).

Remarkably, an enrichment for ORFs with DBF-score$_{model2}$<0.05, suggesting selection that favors low levels of secondary structure, is present in 16 out of 33 sampled eukaryotic species (and 7 out of 8 plant species), while the same applies only to 1 out of 20 sampled archaeal,

and 1 out of 20 sampled bacterial species. To quantify the fraction of ORFs that is presumably subject to structural selection within a given species, we summed up the fraction of ORFs with DBF-scores$_{model2}$<0.05 and DBF-scores$_{model2}$>0.95, taking only quantiles with Z-score≥1.96 ($p<0.05$ for the two tailed hypothesis) into account. We then subtracted the share of ORFs expected to be allotted to the corresponding quantiles assuming a uniform distribution of DBF-scores under absence of selection (see Methods section). On average, we obtained similar fractions of ORFs under structural selection per species in the three domains of life ranging from 1.98% in archaea to 2.06% in eukaryotes and 2.54% in bacteria (figure 5.3A; electronic supplementary material, table S4).

Interestingly, the number of ORFs under structural selection from species representing all three domains of life is considerably lower compared to what we initially observed for virus ORFs, reviving the idea that small RNA based immune systems may have contributed to the realized structuring patterns. If so, we would expect that particularly viruses that encode their genome in form of single stranded RNA, which represents a putative target for antisense small RNAs of a host, would be exposed to a selective pressure that favors highly structured ORFs. We thus grouped viruses according to virus types into five classes (double stranded (ds) DNA viruses, single stranded (ss) DNA viruses, dsRNA viruses, ssRNA plus-strand viruses and ssRNA minus-strand viruses) and checked the structuring patterns of ORFs for each class separately (figure 5.2A). Remarkably, the fraction of genes under structural selection is significantly higher in viruses and on average more than twice as high as in any of the three domains of life (figure 5.2B). Moreover, we found that ssDNA and dsDNA viruses show the lowest amount of ORFs under structural selection within viruses, while ssRNA plus-strand and ssRNA minus-strand viruses exhibit the highest amount of ORFs under structural selection. Notably, 11.5% of ORFs from ssRNA minus-strand viruses are presumably subject to structural selection, a value that surpasses that of any other of the 73 species under examination (figure 5.2A).

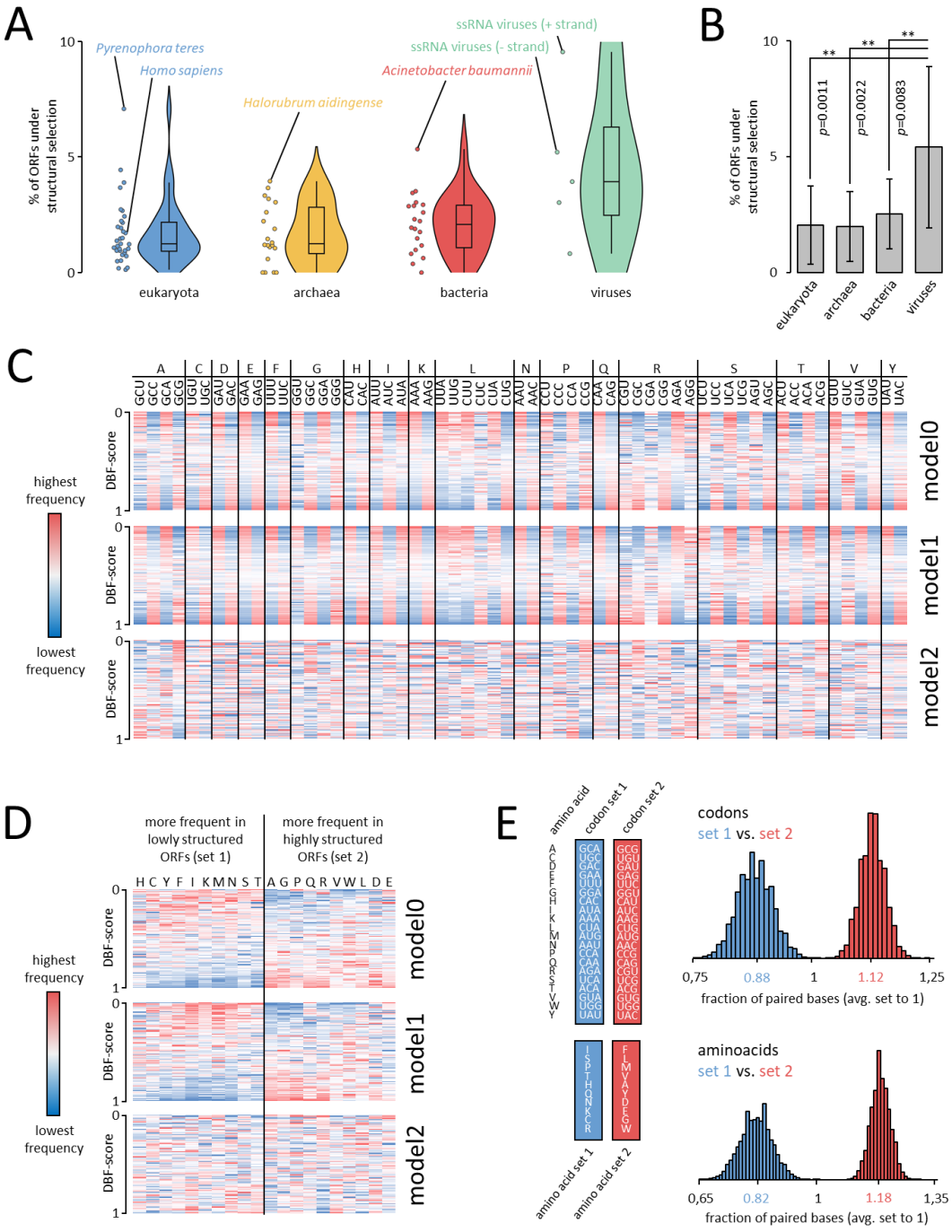


Figure 5.2: (A) Estimation on the fraction of ORFs under structural selection. (B) Virus ORFs are significantly more often under structural selection. P values represent two-tailed p values from unpaired t-tests. Error bars refer to standard deviation. (C) Codon frequencies in ORFs sorted by DBF-score. Data exemplarily taken from Mus musculus. (D) Amino acid frequencies in ORFs sorted by DBF-score. Data exemplarily taken from Mus musculus. (E) Structuring of ORFs that code for identical peptide sequences but use different sets of codons (top) and structuring of ORFs that code for peptides being composed of different sets of amino acids (bottom).

5.3.3. Codon composition and amino acid identity contribute to extreme secondary structures in ORFs

When we compared DBF-scores obtained with different models we were surprised by the observation that a considerable fraction of ORFs shows extremely high or low score values applying model0 and model1 while showing only intermediate values when applying model2. Hence, we considered it unlikely that these ORFs were in fact subject to structural selection. Instead we presumed that this pattern is caused by features other than ORF structure alone, features which must be differentially implemented in the different models.

In contrast to model2, model0 and model1 allow to generate aORFs that differ in GC content and codon usage from the oORF. This applies in particular to oORFs that have an extremely biased codon usage compared to a random codon usage (in case of model0) or the global codon usage of the corresponding species (in case of model1). We thus assumed that aberrant codon usage can lead to extreme structuring of ORFs and checked codon frequencies of oORFs separately for each DBF-score quantile and species, using the three different models. Remarkably, when applying model0 and model1 we observed that codons of all amino acids (except for M and W which are only encoded by one codon) can be divided into those that are found more frequently in highly structured oORFs, or those codons that are found more frequently in lowly structured oORFs, while a corresponding bias is absent when applying model2 (figure 5.2C). Interestingly, we made an analogous finding when focusing on different amino acids instead of codons, suggesting that amino acid identity also influences the degree of structuring (figure 5.2D). The above described division of codons for model0 and model1 invariably follows the combined number of G and C bases in the respective codons for each amino acid, where a higher codon GC content correlates with greater frequency in highly structured oORFs (electronic supplementary material, figure S2a), an observation recently also made by Fricke et al (2018). Similarly, amino acids that are more frequently found in highly structured oORFs tend to exhibit higher mean GC shares in their respective codons ($p<0.001$, Mann-Whitney-U test) (electronic supplementary material, figure S2b,c).

To verify the influence of divergent codon usage on secondary structure, we built a set of 10,000 random peptides with a length of 500 amino acids each. For each peptide, we constructed two corresponding aORFs, using only those codons that are most frequent in lowly structured oORFs (set 1) for the first aORF, and only those codons that are most frequent in highly structured oORFs (set 2) for the second aORF. Then we compared the degree of backfolding for both aORF groups as measured in the amount of paired bases. Confirming our observation on codon bias across highly and lowly structured oORFs, we found that using different sets of synonymous codons can have a massive effect on the degree of backfolding with aORFs being composed of set 1 codons having an average amount of paired bases corresponding to the 0.88-fold of the total average (figure 5.2E). Accordingly, aORFs being composed of set 2 codons have an average amount of paired bases corresponding to the 1.12-fold of the total average (figure 5.2E). Since species with large effective population size (N_e) are those where codon preferences correlate strongly with tRNA abundance (dos Reis and

Wernisch 2008), we checked for a connection between the number of genes under structural selection and N_e but did not observe any correlation.

To check whether amino acid identity also contributes to oORF secondary structure, we built another two sets of 10,000 random peptides with a length of 500 amino acids each. Peptides in the first set were composed only of those 10 amino acids which are more frequently encoded in lowly structured oORFs (set 1 amino acids), while peptides in the second set were composed only of those 10 amino acids which are more frequently encoded in highly structured oORFs (set 2 amino acids). Equal probabilities for each codon of a given amino acid were used. As is the case for divergent codon usage, we found that also amino acid identity is an important determinant of oORF secondary structure, with aORFs of peptides being composed of set 1 amino acids having an average amount of paired bases corresponding to the 0.82-fold of the total average (figure 5.2E). Accordingly, ORFs of peptides being composed of set 1 amino acids have an average amount of paired bases corresponding to the 1.28-fold of the total average (figure 5.2E).

5.4. Discussion

Owing to the degenerate nature of the genetic code, synonymous substitutions have initially been regarded as neutral in terms of natural selection. Since that, a plethora of studies has demonstrated that synonymous codon usage can ultimately alter gene expression, clearly a trait that selection can act on.

As early as 1988, Denis Shields and coworkers found that usage of synonymous codons among 91 *Drosophila melanogaster* genes varied greatly (Shields et al. 1988). Further they observed that enhanced GC-content due to preference for C-ending synonymous codons correlates with gene expression, a finding that has been earlier established also for different unicellular organisms (Gouy and Gautier 1982; Shields and Sharp 1987; Bennetzen and Hall 1982). Mechanistically, this can be explained by more stable and efficient transcription of GC-rich genes and the fact that unfavorable GC-contents can trigger heterochromatization (Kudla et al. 2006; Barahimipour et al. 2015; Newman et al. 2016). Apart from GC-content, it is well known that codon usage also correlates with tRNA abundance, which is likely the outcome of a co-evolution of codon usage and tRNA expression to optimize translation of highly expressed genes (Moriyama and Powell 1997; Duret and Mouchiroud 1999; Duret 2000; Kanaya et al. 2001).

In contrast to our knowledge on how GC-content and synonymous codon usage affects gene expression, far less is known about how secondary structure itself shapes gene expression patterns, which is possibly due to the fact that it is not trivial to disentangle these factors since one will influence the other and any difference in amino acid or synonymous codon usage will likely result in different degrees of RNA secondary structure (Mathews et al. 1999). In regards to this issue, we show here that ORFs with unusual levels of backfolding are indeed biased for specific sets of codons and amino acids, and our analysis of artificial coding sequences using

these different codon and amino acid sets confirmed the close interweaving of codon usage bias, amino acid identity and extreme secondary structures. With these results in mind, we want to point out that any inferences on which traits evolution in fact acts on should be made with particular caution.

The possible role of mRNA secondary structure in the regulation of gene expression has been evaluated previously, though in a limited number of studies and species. Carlini et al. (2001) compared two related drosophilid genes Adh and Adhr with respect to codon bias, expression and ability to form secondary structures. They noticed that the weakly expressed and weakly biased gene Adhr has a much stronger potential for backfolding compared to its heavily expressed and biased counterpart. Soon after, a more comprehensive study that compared folding energies of original and corresponding artificial coding sequences generated by codon shuffling reported widespread selection for local RNA secondary structure, particularly in bacterial species but also in some archaea and eukaryotic organisms (Katz and Burge 2003). Similar findings were subsequently presented for mammals (Chamary and Hurst 2005). Functional evidence for the importance of mRNA structures was provided by Kudla and colleagues, who showed that the stability of mRNA folding near the ribosomal binding site is a major determinant for the expression of a GFP reporter protein encoded by a set of mRNAs that randomly differ at synonymous sites (Kudla et al. 2009). A general correlation of folding energies and profiles of ribosomal density in *Escherichia coli* and *Sacharomyces cerevisae* emphasized the importance of mRNA secondary structure and translation efficiency (Tuller et al. 2010). Finally, a subtle large-scale analysis of coding sequences conducted by Fricke and colleagues revealed that not only the amount of secondary structure but also their nature is non-random, with base pairing events between the first bases of two opposing codons being significantly underrepresented, suggesting presence of selective forces (Fricke et al. 2018).

Noteworthily, Hoede et al. (2006) have proposed that selection also acts on the level of DNA structure during transcription and favors local intra-strand secondary structures to reduce the extent of transcriptional mutagenesis, a phenomenon that can particularly observed at highly expressed genes. An important contribution to this complex issue was recently made by Lai et al. (2018) who demonstrated that mRNAs as well as long non-coding RNAs intrinsically form secondary structures that result in short 5’- to 3’-end distances which possibly affects the process of translation initiation.

Considering the available data, the unveiled widespread selection for high or low levels of secondary structure in coding sequences throughout the analyzed species is not surprising, and we propose that evolutionary adjustment of the degree of secondary structure in ORFs contributes to fine-tuning of gene expression. If at all, one could argue that our estimates on the fraction of genes that is subject to this kind of selection appears surprisingly small. However, we want to emphasize that our estimates represent the lower limit, deduced from the enrichment of genes with extremely high or low DBF-scores, additionally excluding those genes were we cannot rule out that selection acts on the level of codon usage and/or GC-content instead of secondary structure alone. For these genes we assume that restrictions on the amino acid sequence level prevent mRNAs to be optimally folded and that synonymous

codon usage is exhaustively used to shift the mRNA towards the optimum. This would be a plausible explanation for the enrichment of genes at both ends of the DBF-score spectrum. For an undeterminable fraction of genes, the optimal folding might be realized without requiring a suspicious arrangement of synonymous codons, though possibly not being less subject to structural selection that maintains the current state.

Interestingly, we found the highest amount of genes under structural selection in RNA viruses. Many viral RNA structures, so called *cis*-acting elements, are important for viral replication but are typically restricted to non-translated regions of the viral genome and the function of structures within coding sequences is poorly understood (Liu et al. 2009). We speculate that highly structured viral coding sequences could be at least in part promoted by anti-viral host RNAi pathways. Many species have developed siRNA- and piRNA-based defense strategies to combat viral infections (Ding and Voinnet 2007; Ding and Lu 2011; Szittya and Burgyán 2013; Bronkhorst and van Rij 2014). Since it has been shown that target secondary structure is a major determinant of RNAi efficiency (Shao et al. 2007; Rosenkranz et al. 2015; Fast et al. 2017), we assume that the evolutionary arms race between viruses and hosts in many cases gave rise to viral transcripts that are characterized by a high degree of backfolding in order to provide as little attack surface as possible for single-stranded guiding RNAs.

In summary, our results demonstrate the close connection between codon usage bias, amino acid identity and RNA secondary structure. Moreover, using a yet unrivaled broad data basis and an algorithm that excludes the effect of altered codon usage and GC-content, we show that selection for extreme secondary structures within coding sequences is a widespread phenomenon throughout life and, with respect to viral ORFs, even beyond. Finally, with PACKEIS we provide a tool that allows other researchers to easily conduct corresponding genome wide analyses in any species of their choice not considered in the course of this study.

5.5. Materials and Methods

5.5.1. Data selection

In total 73 representative species from the three domains of life (eukaryotes n=33, archaea n=20, bacteria n=20) were selected, paying attention to a balanced representation of sub-clades within a given domain. ORF sequences from eukaryotes were downloaded from Ensembl database (release 94, Zerbino et al. 2017). The longest transcripts for each gene were extracted using the custom Perl script select_longest_transcripts.pl. For archaea and bacteria species, we downloaded cDNA data and peptide data from Ensembl (release 94). We translated cDNA in all possible forward frames and checked for presence of the resulting peptide sequences in the corresponding peptide dataset. Candidate ORFs with a match in the peptide dataset were collected for each species. Prediction of ORF sequences from cDNA and peptide datasets was conducted using the custom Perl script ORF_from_cDNA.pl.

Viral genome sequences were downloaded from NCBI GenBank. ORF sequences were extracted from GenBank files and converted into FASTA format using the custom Perl script

GB_2_FASTA.pl. Viral sequences were further sorted into separate files according to viral classes (ssRNA positive strand, ssRNA negative strand, dsRNA, ssDNA, dsDNA) and hosts (algae, archaea, bacteria, environment, human, invertebrates, plants, protozoa, vertebrates) using the custom Perl script sort_viral_genomes.pl. All custom Perl scripts used in the course of this study are available at https://sourceforge.net/p/packeis.

Gene expression data was downloaded from PaxDb Protein Abundance Database and EBI Expression Atlas (Wang et al. 2015; Petryszak et al. 2016).

5.5.2. The PACKEIS algorithm

In a first step, PACKEIS calculates average probabilities for base pairing in predicted local RNA structures using algorithms of the ViennaRNA package (Lorenz et al. 2011). Therefore it runs RNAfold or RNAplfold (depending on the input ORF length) on a number of input sequences (FASTA file) and parses the resulting output files. Next, it calculates the degree of backfolding (DBF) as measured in the fraction of paired bases within the original ORF (oORF) defined by the sum of the average base pairing probabilities for each position divided by the number of total bases. In a second step, PACKEIS generates a set of alternative ORFs (aORFs, default n=100), each of which still codes for the same amino acid sequence, and calculates the DBF for each aORF as described above. By comparing the DBF of the oORF with those of the aORFs, PACKEIS outputs a measurand (DBF-score) that allows to assess the probability that the DBF of the oORF is a product of chance, where a DBF-score of 0 means that none of the aORFs exhibits a lower DBF, while a DBF-score of 1 means that none of the aORFs exhibits a higher DBF. The DBF-score is calculated according to the following formula

$$\mathrm{DBF-score} = \frac{i \,/\, (i + j + k \,/\, 2)}{j \,/\, (i + j + k \,/\, 2)}$$

Where *i* refers to the number of aORFs with higher DBF and *j* refers to the number of aORFs with lower DBF. *k* refers to the number of aORFs with identical DBF compared to the oORF so that the DBF score amounts to 0.5 in case that all aORFs behave exactly as the oORF.

For the construction of aORFs, PACKEIS implements three different models. Model0 (free) uses equal frequencies for all synonymous codons of a specific amino acid. Model1 (strict) uses specified codon frequencies that reflect the codon usage of the species in question (or alternatively codon usage frequencies as defined by the user). Model2 (shuffle) constructs aORFs by shuffling those codons that are already present in a given oORF. In contrast to model0 and model1, codon frequencies and GC-content are perfectly preserved in each of the aORFs compared to the oORF when applying model2 (figure 5.3). Thus, applying model2 allows to exclude the impact of aberrant codon usage or GC-content of a given oORF and to assess whether the present codons are arranged in a non-random fashion regarding the effect on secondary structure (figure 5.3).

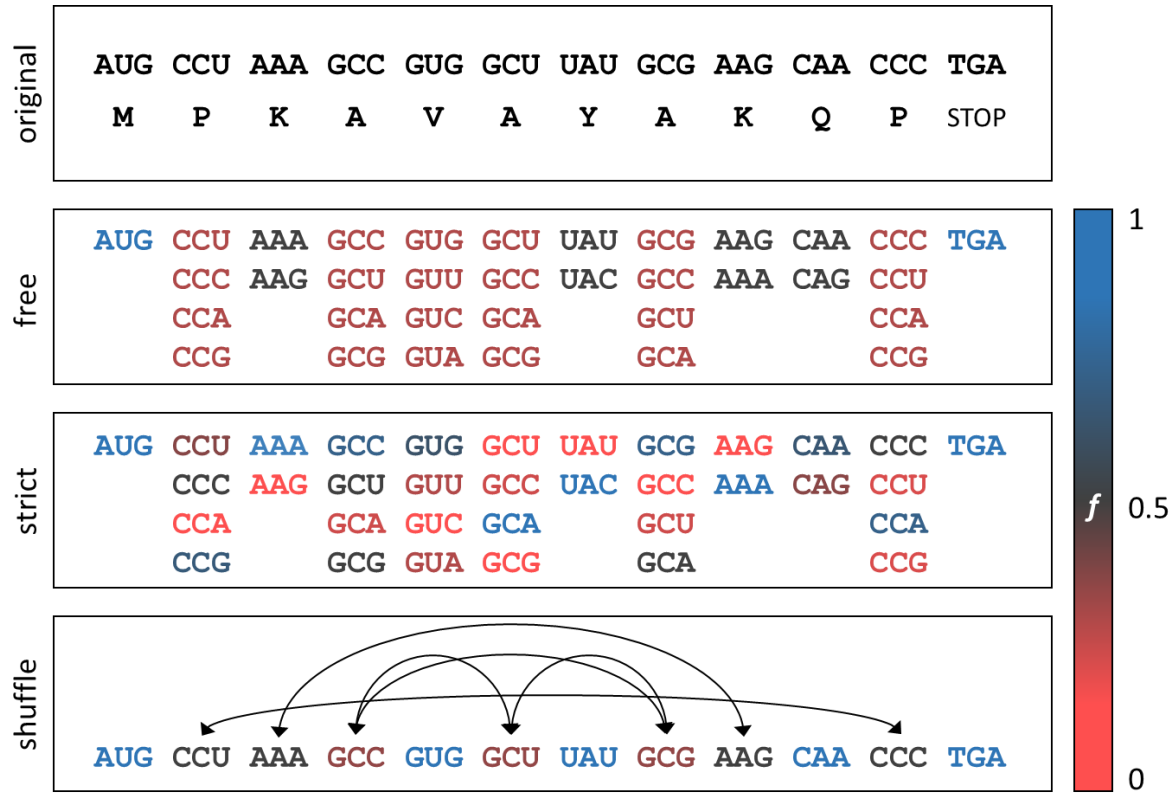


Figure 5.3: PACKEIS uses different models to generate artificial ORFs based on the original ORF. Colors refer to the probability for a specific codon to be placed at a given position. When applying model0 (free) PACKEIS uses equal probabilities for all codons of a specific amino acid. When applying model1 (strict) the probabilities are derived from the global codon usage of the species in question. When applying model2 (shuffle), codons of the original ORF are randomly shuffled. Note that in the above example Valine can only be encoded by GUG when applying model2 since no alternative Valine codons are present in the given ORF.

PACKEIS produces a number of output files including a table that lists the DBF-scores for each input sequence, a text file that shows the distribution of DBF-scores from 0 to 1 in steps of 0.01, a table that refers to codon composition, amino acid identity and GC content for ORFs with a specific DBF-score, and finally one text file per input sequence that lists the DBF-scores for each of the corresponding aORFs.

To check whether the implemented models yield coherent results, we pairwise compared DBF-scores for 27,628 *Arabidopsis thaliana* ORFs obtained when applying the three different models. Indeed we observed a high degree of correlation across the results obtained applying the different models as deduced from Pearson's correlation coefficients ranging from r=0.86 to r=0.97, supporting the general validity of the results (electronic supplementary material, figure S1).

The PACKEIS software including a detailed documentation and test datasets is freely available at https://sourceforge.net/p/packeis and http://www.smallRNAgroup.uni-mainz.de/software.html.

5.5.3. DBF score calculation with PACKEIS

DBF-scores were calculated using the PACKEIS software which was developed in the course of this study. PACKEIS will use RNAplfold to calculate base pairing probabilities based on local rather than global RNA folding which is more reliable for larger sequences. Therefore, we set the minimum length [nt] of an input sequence for PACKEIS to run RNAplfold instead of RNAfold to 100 with the option -l 100. PACKEIS was run three times using different models for the construction of aORFs applying the option -m 0, -m 1 and -m 2, respectively.

5.5.4. Quantifying the number of ORFs under structural selection

We sectioned DBF-scores into 101 quantiles ranging from 0 to 1 in steps of 0.01 ($Q_{0.00}$ ... $Q_{1.00}$). ORFs that fell in the range of 0 to 0.04 (p<0.05) were considered as candidates for being subject to selection that favors low structuring. ORFs that fell in the range of 0.96 to 1 (p>0.95) were considered as candidates for being subject to selection that favors high structuring. The null hypothesis (absence of selection) was rejected, if any of the lower or upper five quantiles showed a significant enrichment for ORFs. The summed fractions of genes in the lower (s_{low}) or upper (s_{high}) five quantiles with Z scores ≥1.96, deducting the share of ORFs that would be expected for each quantile assuming an even distribution in absence of selection (1/101), was considered as the fraction of genes that is subject to structural selection (S=s_{low}+s_{high}) according to the following formula:

$$s_{low/high} = \sum_{k=i}^{j} f\left(z_{Q\frac{k}{100}}\right) \cdot \left(a_{Q\frac{k}{100}} - \frac{1}{101}\right)$$

where f (Z<1.96)=0 and f (Z≥1.96)=1. For s_{low} i=0 and j=4, for s_{high} i=96 and j=100. a refers to the number of ORFs within the specific quantile. According to the definition of f, each of the lowest and highest five quantiles is only taken into account if ORFs from the corresponding quantile are significantly over-represented (p<0.05, Z>=1.96). The term $z_{Q\frac{k}{100}}$ refers to the Z-score of a specific quantile, e.g. $Q_{1.00}$ for k=100. The function value $f\left(z_{Q\frac{k}{100}}\right)$ becomes 0 for $z_{Q\frac{k}{100}}$<1.96. s_{low} represents the number of genes were selection acts to reduce structuredness while s_{high} represents the number of genes were selection acts to enhance structuredness. Both fractions together represent the total number of genes S that is subject to structural selection.

5.5.5. Generating random peptides and ORFs

To simulate the effect of using different sets of codons on secondary structure we first sorted codons according to a bias in codon frequency across oORFs with high and low $DBFscores_{model1}$. A bias was attested in case that we observed a steady increase or decrease of the average codon frequency from the upper five quantiles via the middle 91 quantiles to the lower five quantiles. Since the observed bias was not always consistent across all species under examination, we decided based on the status in the majority of species (electronic supplementary material, table S5). Set 1 codons were those being more frequent in oORFs with $DBFscores_{model1}$ ranging from 0 to 0.04 in the majority of species, set 2 codons were those being more frequent in oORFs with $DBFscores_{model1}$ ranging from 0.96 to 1 in the majority of species. Amino acids were grouped into set 1 and set 2 amino acids accordingly.

To analyze the effect of using different sets of codons on secondary structure, we built a set of 10,000 random peptides with a length of 500 amino acids each. The first amino acid of each peptide was methionine. For each of the 10,000 random peptides we constructed one corresponding aORF using set 1 codons and one corresponding aORF using set 2 codons. Then we predicted and compared the fraction of paired bases for aORFs being composed of set 1 codons and aORFs being compared of set 2 codons using the custom Perl script test_codon_sets.pl.

To analyze the effect of using different sets of amino acids on secondary structure, we built two sets of 10,000 random peptides with a length of 500 amino acids each, with the first peptide set being composed of set 1 amino acids and the second peptide set being composed of set 2 amino acids. One ORFs for each of the 20,000 peptides was constructed with equal probabilities for possible codons of a given amino acid. We predicted and compared the fraction of paired bases for ORFs from set 1 peptides and ORFs from set 2 peptides using the custom Perl script test_aa_sets.pl. Note that in contrast to the analysis of set 1 and set 2 codons, we did not compare sets of two aORFs that encode an identical peptide, but rather independent ORFs that code for different peptides (set 1 and set 2 peptides). All custom Perl scripts used in the course of this study are available at https://sourceforge.net/p/packeis.

5.6. Author Contributions

DG, **JJ** and DR performed the relevant bioinformatics analyses and coded custom Perl scripts. DG, **JJ** and DR interpreted the results. DG and DR wrote the paper. The PACKEIS software was developed by DR. DR is responsible for study design and concept.

5.7. Declarations and Acknowledgements

Figures and tables supporting this paper have been uploaded as electronic supplementary material (https://doi.org/10.1098/rsob.190020).

The authors declare to have no conflict of interest.

This work was supported by the intramural “Impulsfonds” of the Johannes Gutenberg University, Mainz.

We thank Prof. Hans Zischler for helpful discussion and administrative support.

5.8. References

Babendure, Jeremy R.; Babendure, Jennie L.; Ding, Jian-Hua; Tsien, Roger Y. (2006): Control of mammalian translation by mRNA structure near caps. In *RNA* 12 (5), pp. 851–861. DOI: 10.1261/rna.2309906.

Barahimipour, Rouhollah; Strenkert, Daniela; Neupert, Juliane; Schroda, Michael; Merchant, Sabeeha S.; Bock, Ralph (2015): Dissecting the contributions of GC content and codon usage to gene expression in the model alga Chlamydomonas reinhardtii. In *The Plant Journal* 84 (4), pp. 704–717.

Bennetzen, Jeffrey L.; Hall, Benjamin D. (1982): Codon selection in yeast. In *The Journal of biological chemistry* 257 (6), pp. 3026–3031.

Bronkhorst, Alfred W.; van Rij, Ronald P. (2014): The long and short of antiviral defense: small RNA-based immunity in insects. In *Current opinion in virology* 7, pp. 19–28.

Carlini, David B.; Chen, Ying; Stephan, Wolfgang (2001): The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. In *Genetics* 159 (2), pp. 623–633.

Chamary, J. V.; Hurst, Laurence D. (2005): Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. In *Genome biology* 6 (9), R75.

Ding, Shou-Wei; Lu, Rui (2011): Virus-derived siRNAs and piRNAs in immunity and pathogenesis. In *Current opinion in virology* 1 (6), pp. 533–544.

Ding, Shou-Wei; Voinnet, Olivier (2007): Antiviral immunity directed by small RNAs. In *Cell* 130 (3), pp. 413–426.

dos Reis, Mario; Wernisch, Lorenz (2008): Estimating translational selection in eukaryotic genomes. In *Molecular biology and evolution* 26 (2), pp. 451–461.

Duret, Laurent (2000): tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. In *Trends in genetics : TIG* 16 (7), pp. 287–289.

Duret, Laurent; Mouchiroud, Dominique (1999): Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. In *Proceedings of the National Academy of Sciences* 96 (8), pp. 4482–4487.

Fast, Isabel; Hewel, Charlotte; Wester, Laura; Schumacher, Julia; Gebert, Daniel; Zischler, Hans et al. (2017): Temperature-responsive miRNAs in Drosophila orchestrate adaptation to different ambient temperatures. In *RNA* 23 (9), pp. 1352–1364.

Fricke, Markus; Gerst, Ruman; Ibrahim, Bashar; Niepmann, Michael; Marz, Manja (2018): Global importance of RNA secondary structures in protein-coding sequences. In *Bioinformatics (Oxford, England)* 35 (4), pp. 579–583.

Gouy, Manolo; Gautier, Christian (1982): Codon usage in bacteria: correlation with gene expressivity. In *Nucleic Acids Res* 10 (22), pp. 7055–7074.

Hoede, Claire; Denamur, Erick; Tenaillon, Olivier (2006): Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. In *PLoS genetics* 2 (11), e176.

Huang, Lin; Ashraf, Saira; Lilley, David M. J. (2019): The role of RNA structure in translational regulation by L7Ae protein in archaea. In *RNA* 25 (1), pp. 60–69.

Huang, Yu; Li, Yi (2018): Secondary siRNAs rescue virus-infected plants. In *Nature plants* 4 (3), p. 136.

Kanaya, Shigehiko; Yamada, Yuko; Kinouchi, Makoto; Kudo, Yoshihiro; Ikemura, Toshimichi (2001): Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. In *J Mol Evol* 53 (4-5), pp. 290–298.

Katz, Luba; Burge, Christopher B. (2003): Widespread selection for local RNA secondary structure in coding regions of bacterial genes. In *Genome research* 13 (9), pp. 2042–2051.

Kimura, Motoo (1977): Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. In *Nature* 267 (5608), p. 275.

Kudla, Grzegorz; Lipinski, Leszek; Caffin, Fanny; Helwak, Aleksandra; Zylicz, Maciej (2006): High guanine and cytosine content increases mRNA levels in mammalian cells. In *PLoS biology* 4 (6), e180.

Kudla, Grzegorz; Murray, Andrew W.; Tollervey, David; Plotkin, Joshua B. (2009): Coding-sequence determinants of gene expression in Escherichia coli. In *Science* 324 (5924), pp. 255–258.

Lai, Wan-Jung C.; Kayedkhordeh, Mohammad; Cornell, Erica V.; Farah, Elie; Bellaousov, Stanislav; Rietmeijer, Robert et al. (2018): mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances. In *Nature communications* 9 (1), p. 4328.

Liu, Ying; Wimmer, Eckard; Paul, Aniko V. (2009): Cis-acting RNA elements in human and animal plus-strand RNA viruses. In *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1789 (9-10), pp. 495–517.

Lorenz, Ronny; Bernhart, Stephan H.; Siederdissen, Christian Höner zu; Tafer, Hakim; Flamm, Christoph; Stadler, Peter F.; Hofacker, Ivo L. (2011): ViennaRNA Package 2.0. In *Algorithms for molecular biology* 6 (1), p. 26.

Mao, Yuanhui; Liu, HuiLing; Liu, Yanlin; Tao, Shiheng (2014): Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. In *Nucleic Acids Res* 42 (8), pp. 4813–4822.

Mathews, David H.; Sabina, Jeffrey; Zuker, Michael; Turner, Douglas H. (1999): Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. In *Journal of molecular biology* 288 (5), pp. 911–940.

Moriyama, Etsuko N.; Powell, Jeffrey R. (1997): Codon usage bias and tRNA abundance in Drosophila. In *J Mol Evol* 45 (5), pp. 514–523.

Newman, Zachary R.; Young, Janet M.; Ingolia, Nicholas T.; Barton, Gregory M. (2016): Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. In *Proceedings of the National Academy of Sciences* 113 (10), E1362-E1371.

Petryszak, Robert; Keays, Maria; Tang, Y. Amy; Fonseca, Nuno A.; Barrera, Elisabet; Burdett, Tony et al. (2016): Expression Atlas update--an integrated database of gene and protein expression in humans, animals and plants. In *Nucleic acids research* 44 (D1), D746-52. DOI: 10.1093/nar/gkv1045.

Powell, Jeffrey R. (2018): Mosquito-borne human viral diseases: why Aedes aegypti? In *The American journal of tropical medicine and hygiene* 98 (6), pp. 1563–1565.

Qiu, Yang; Xu, Yanpeng; Zhang, Yao; Zhou, Hui; Deng, Yong-Qiang; Li, Xiao-Feng et al. (2017): Human virus-derived small RNAs can confer antiviral immunity in mammals. In *Immunity* 46 (6), 992-1004. e5.

Rosenkranz, David; Rudloff, Stefanie; Bastuck, Katharina; Ketting, René F.; Zischler, Hans (2015): Tupaia small RNAs provide insights into function and evolution of RNAi-based transposon defense in mammals. In *RNA* 21 (5), pp. 911–922.

Shao, Yu; Chan, Chi Yu; Maliyekkel, Anil; Lawrence, Charles E.; Roninson, Igor B.; Ding, Ye (2007): Effect of target secondary structure on RNAi efficiency. In *RNA* 13 (10), pp. 1631–1640.

Sharp, Paul Martin; Averof, Michalis; Lloyd, Andrew T.; Matassi, Giorgio; Peden, John F. (1995): DNA sequence evolution: the sounds of silence. In *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 349 (1329), pp. 241–247.

Shields, Denis C.; Sharp, Paul M. (1987): Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases. In *Nucleic Acids Res* 15 (19), pp. 8023–8040.

Shields, Denis C.; Sharp, Paul M.; Higgins, Desmond G.; Wright, Frank (1988): " Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. In *Molecular biology and evolution* 5 (6), pp. 704–716.

Szittya, György; Burgyán, József (2013): RNA interference-mediated intrinsic antiviral immunity in plants. In : Intrinsic Immunity: Springer, pp. 153–181.

Tuller, Tamir; Waldman, Yedael Y.; Kupiec, Martin; Ruppin, Eytan (2010): Translation efficiency is determined by both codon bias and folding energy. In *Proceedings of the National Academy of Sciences* 107 (8), pp. 3645–3650.

Varjak, Margus; Leggewie, Mayke; Schnettler, Esther (2018): The antiviral piRNA response in mosquitoes? In *Journal of General Virology* 99 (12), pp. 1551–1562.

Wang, Mingcong; Herrmann, Christina J.; Simonovic, Milan; Szklarczyk, Damian; Mering, Christian von (2015): Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. In *Proteomics* 15 (18), pp. 3163–3168.

Zerbino, Daniel R.; Achuthan, Premanand; Akanni, Wasiu; Amode, M. Ridwan; Barrell, Daniel; Bhai, Jyothish et al. (2017): Ensembl 2018. In *Nucleic Acids Res* 46 (D1), D754-D761.

Conclusion

The overarching aim of this thesis is to tackle various unexplored aspects of small RNA pathways. Following this goal, the study on tRNA fragments provides a first overview of tRF expression across tissues and species. It revealed that 5' tRNA halves are highly abundant in the hippocampus of primates, but not in the hippocampus of other analyzed mammals. Thus, high levels of 5' tRHs in the hippocampus might be a primate-specific trait, which would be the first report of a lineage-specific differential tRF expression in the animal kingdom to the best of my knowledge. As transcripts that are presumably stabilized by 5' tRH targeting were enriched for a function in neurogenesis, while transcripts that are presumably degraded upon 5' tRH targeting were depleted of this function, it can be hypothesized that 5' tRHs play a role in fine-tuning primate neurogenesis. Furthermore, the dissection of 5' tRH targeting rules via a novel k-mer approach provides surprising new findings suggesting a sequence-specific, but non-miRNA-like mechanism of target recognition that erodes the AGO-centric view on tRF-mediated gene regulation of previous studies (Haussecker et al. 2010; Kuscu et al. 2018). It will be interesting to see with which proteins certain tRF subsets interact in distinct tissues and cell states. As the developed k-mer approach may not only be used to study the targeting rules of other tRFs, but also other (yet unknown) small RNAs, new insights in small RNA target recognition may be gained with this approach in the future.

Lewis et al. have recently shown that somatically expressed piRNAs are a common state in arthropods (Lewis et al. 2018). Here, the study of molluskan piRNAs and PIWI proteins demonstrates that PIWI paralogs are ubiquitously expressed in the representative mollusk species *L. stagnalis* and *C. gigas*, suggesting that a somatic piRNA pathway was already established in an early protostomian ancestor. To reinforce this theory, it will be interesting to investigate the situation of other protostomes in this context. The characterization of the putative molluskan piRNAs showed that transposons as well as messenger RNAs are targeted both in the germline and in the soma and enhances the view that a presumably germline-restricted piRNA pathway targeting only transposons represents a vertebrate-specific trait rather than the general rule within metazoans. The revealed lineage-specific adaptations of molluskan piRNA cluster composition and the dynamic expression of piRNA clusters during oyster development emphasize once more the evolutionary plasticity of the PIWI/piRNA system.

Lately, also the dogma of a germline-restricted piRNA pathway in vertebrates was challenged, as some studies suggest that a somatic piRNA pathway function might have persisted in certain niches (Nandi et al. 2016; Ross et al. 2014). However, these findings are controversially discussed as (due to the lack of appropriate anti-PIWI-antibodies) they are mainly based on sequence-homology approaches and piRNA databases are known to be contaminated by other small RNA classes (Tosar et al. 2018). With the approach to knockdown PLD6 in cultured cells and identify the small RNAs that depend on PLD6 via small RNA sequencing, a method is provided to identify putative piRNAs without the need of anti-PIWI antibodies. The underlying idea is that small RNAs that depend on PLD6 can be considered as putative piRNAs, as PLD6 is

the central factor of primary piRNA biogenesis. As a proof of concept, this RNAi-based approach was tested in HEK293T cells, which showed that the expression of small RNAs in the size range of piRNAs is dependent on PLD6, suggesting that PLD6 acts as an endonuclease generating small RNAs in somatic cells like HEK293T cells similar to germline cells. piRNA-like target silencing for PLD6-dependent small RNAs was not observed in HEK293T cells, which is presumably due to the lack of PIWI protein expression in this cell line. It will be interesting to assess the full potential of this approach on physiologically more relevant stem cells, where PIWI expression has been reported (Sharma 2001; Wu et al. 2010; Nolde et al. 2013; Nandi et al. 2016).

In the end, the study on the dimension of natural selection of secondary structure of coding sequences not only revealed that the structural selection of coding regions is a widespread phenomenon throughout all domains of life, but also found highly structured coding regions to be enriched in single-stranded RNA viruses. It is likely that the genomes of single-stranded RNA viruses evolved in favor of highly structured transcripts in order to evade the RNAi-mediated defense of the host, which would be yet another example of the evolutionary arms race between RNAi pathways and their targets. Collectively, the results presented in this thesis contribute to a more variegated understanding of small RNA pathways.

References

Haussecker, Dirk; Huang, Yong; Lau, Ashley; Parameswaran, Poornima; Fire, Andrew Z.; Kay, Mark A. (2010): Human tRNA-derived small RNAs in the global regulation of RNA silencing. In *RNA (New York, N.Y.)* 16 (4), pp. 673–695. DOI: 10.1261/rna.2000810.

Kuscu, Canan; Kumar, Pankaj; Kiran, Manjari; Su, Zhangli; Malik, Asrar; Dutta, Anindya (2018): tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. In *RNA (New York, N.Y.)* 24 (8), pp. 1093–1105. DOI: 10.1261/rna.066126.118.

Lewis, Samuel H.; Quarles, Kaycee A.; Yang, Yujing; Tanguy, Melanie; Frézal, Lise; Smith, Stephen A. et al. (2018): Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. In *Nature ecology & evolution* 2 (1), pp. 174–181. DOI: 10.1038/s41559-017-0403-4.

Nandi, Sayan; Chandramohan, Dhruva; Fioriti, Luana; Melnick, Ari M.; Hébert, Jean M.; Mason, Christopher E. et al. (2016): Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. In *Proceedings of the National Academy of Sciences of the United States of America* 113 (45), pp. 12697–12702. DOI: 10.1073/pnas.1609287113.

Nolde, Mona J.; Cheng, Ee-chun; Guo, Shangqin; Lin, Haifan (2013): Piwi Genes are dispensable for normal hematopoiesis in mice. In *PloS one*. DOI: 10.1371/journal.pone.0071950.g001.

Ross, Robert J.; Weiner, Molly M.; Lin, Haifan (2014): PIWI proteins and PIWI-interacting RNAs in the soma. In *Nature* 505 (7483), pp. 353–359. DOI: 10.1038/nature12987.

Sharma, A. K. (2001): Human CD34+ stem cells express the hiwi gene, a human homologue of the Drosophila gene piwi. In *Blood* 97 (2), pp. 426–434. DOI: 10.1182/blood.V97.2.426.

Tosar, Juan Pablo; Rovira, Carlos; Cayota, Alfonso (2018): Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. In *Communications biology* 1, p. 2. DOI: 10.1038/s42003-017-0001-7.

Wu, Qiuling; Ma, Qi; Shehadeh, Lina A.; Wilson, Amber; Xia, Linghui; Yu, Hong; Webster, Keith A. (2010): Expression of the Argonaute protein PiwiL2 and piRNAs in adult mouse mesenchymal stem cells. In *Biochemical and biophysical research communications* 396 (4), pp. 915–920. DOI: 10.1016/j.bbrc.2010.05.022.

Autorenbeiträge

Kapitel 1: tRNA fragments in health and disease

Das Manuskript wurde von Julia Jehn unter Mitwirkung von David Rosenkranz geschrieben. Julia Jehn erstellte alle Abbildungen.

Kapitel 2: Expression profile and targeting rules of 5' tRNA halves

Die Studie wurde von David Rosenkranz unter Einbezug wesentlicher Beiträge von Julia Jehn konzipiert. Charlotte Hewel selektierte und prozessierte öffentlich zugängliche sRNAseq-Datensätze von menschlichen Geweben, während Verena Erb die Annotation kleiner RNAs und RPM-Berechnung dieser Datensätze durchführte. Laura Wester generierte und analysierte die sRNA-Library des menschlichen Hippocampus mit Hilfe von Isabel Fast. Roxana N. Kooijmans sezierte das Makaken-Gehirn und homogenisierte die Hippocampus-Gewebeprobe. Isabel Fast isolierte die Gesamt-RNA aus der Makaken-Probe und koordinierte die Beschaffung und Sequenzierung aller Primaten-Hirnproben. Julia Jehn prozessierte und analysierte die sRNAseq-Datensätze der Zelllinien, der Primaten-Hirnproben und des Hippocampus von Schwein, Ratte und Maus. Svenja Wulsch und Benjamin Ottum transfizierten die Zellen mit synthetischen tRF-Analoga, prognostizierten potenzielle Zieltranskripte und qPCR-quantifizierten die Expression der potenziellen Zieltranskripte und des tRF-Levels mit Unterstützung von Julia Jehn. Julia Jehn transfizierte die Zellen mit tRF-antisense-RNAs und qPCR-quantifizierte das tRF-Level. Die für die k-mer-Analyse entwickelten Perl-Skripte wurden von David Rosenkranz kodiert. Die Parameteroptimierung des Perl-Skripts map_kmers.pl wurde von Jana Treml durchgeführt. Jana Treml und David Rosenkranz analysierten öffentlich zugängliche RNA-Seq-Daten, während Julia Jehn die im Rahmen des Projekts generierten RNA-Seq-Daten analysierte. Julia Jehn schrieb die R-Skripte, führte die k-mer-Analyse durch, analysierte die thermodynamisch favorisierten tRF-Transkript-Interaktionen und untersuchte die potenzielle Funktion von tRHs in der Neurogenese. Julia Jehn schrieb das Manuskript, zu dem David Rosenkranz wichtige Beiträge leistete.

Kapitel 3: PIWI genes and piRNAs are ubiquitously expressed in somatic tissues of mollusks

David Rosenkranz hat die Studie entworfen. Daniel Gebert identifizierte die PIWI-Paraloge in sequenzierten, aber unannotierten Mollusken-Genomen. Basierend auf den identifizierten Sequenzen entwarf Julia Jehn die Isoform-spezifischen qPCR-Primer und extrahierte zusammen mit Julian S. T. Kiefer die Gesamt-RNA aus den von David Rosenkranz präparierten Geweben. Julia Jehn und Julian S. T. Kiefer verarbeiteten die Gesamt-RNA und führten die qPCR-Quantifizierung durch, die von Julia Jehn analysiert und visualisiert wurde. Daniel Gebert rekonstruierte und visualisierte die PIWI-Gen-Stammbäume. Frank Pipilescu kultivierte und

sezierte die Spitzschlammschnecken. Frank Pipilescu und führte zusammen mit Sarah Stern die RNA-Extraktion für die anschließende sRNA-Sequenzierung durch. Charlotte Hewel, Frank Pipilescu und Sarah Stern führten de novo miRNA-Annotation basierend auf den erhaltenen sRNAseq-Daten durch. Daniel Gebert führte die bioinformatische Analyse von piRNA-Clustern durch und erstellte die entsprechenden Abbildungen. David Rosenkranz analysierte die sRNAseq-Daten und entwickelte die Perl-Skripte zur Datenanalyse. David Rosenkranz schrieb das Manuskript mit Beiträgen von Daniel Gebert und Julia Jehn. Julian S. T. Kiefer, Frank Pipilescu und Charlotte Hewel lieferten Anregungen für Korrekturen und Verbesserungen des Manuskripts.

Kapitel 4: PLD6-dependent small RNAs – An RNAi-based approach to identify somatic piRNAs

Die Studie wurde von David Rosenkranz entworfen. Julia Jehn und Julian S. T. Kiefer führten die siRNA-Transfektion durch und validierten den PLD6-Knockdown mittels qPCR-Analyse und Western-Blot. Julia Jehn und Julian S. T. Kiefer extrahierten die Gesamt-RNA der transfizierten Zellen und analysierten die daraus gewonnenen sRNAseq-Daten. Julia Jehn schrieb das Perl-Skript RemovePLD6-siRNA-Nr3.pl und führte weitere biochemische und bioinformatische Analysen von PLD6-abhängigen kleinen RNAs sowie die Analyse der RNAseq-Daten durch. Das Perl-Skript piRanha.pl wurde von David Rosenkranz kodiert. Dieses Kapitel wurde von Julia Jehn geschrieben.

Kapitel 5: Selection for extreme secondary structure in coding sequences of viruses – An evolutionary arms race borne defense mechanism?

Die Studie wurde von David Rosenkranz entworfen, der auch die PACKEIS-Software entwickelte. Die Perl-Skripte wurden von David Rosenkranz unter Beitrag von Daniel Gebert kodiert. Die Analyse der Beziehung zwischen dem DBF-Score und der Genexpression wurde von Julia Jehn durchgeführt. Daniel Gebert, Julia Jehn und David Rosenkranz interpretierten die Ergebnisse. David Rosenkranz schrieb das Manuskript, mit Beiträgen von Daniel Gebert.

Danksagung

An erster Stelle möchte ich ein riesengroßes Dankeschön an D. richten, dafür dass er mir die Möglichkeit gegeben hat als Teil seiner Gruppe an verschiedensten Projekten mitzuarbeiten, Verantwortung zu übernehmen und mir dabei stets viele Freiheiten gegeben, sowie das nötige Vertrauen entgegengebracht hat. Zudem möchte ich mich bei Herrn Z. für die Zweitbetreuung meiner Arbeit und die herzliche Atmosphäre im Institut bedanken. Des Weiteren gilt mein Dank Herr K. und Frau S. für ihre investierte Zeit und die hilfreichen Kommentare im Rahmen der TAC-Meetings. E. und K. möchte ich für ihre tatkräftige Unterstützung bei allen IPP-bezogenen Angelegenheiten danken. Danken möchte ich auch allen „IOMesen“ im ersten Stock der Anthro für den kollegialen Umgang und die vielen schönen BBQ Sessions vor dem Institut. Hierbei möchte ich mich vor allem bei J. für die super Zusammenarbeit und bei C. für ihre phänomenale Unterstützung bei unserem Einstieg in die Bioinformatik bedanken. Außerdem möchte ich K. für all die hilfreichen Diskussionen im Büro und die kurzweiligen Mittagspausen Danke sagen. Vergessen werden darf natürlich auch nicht La Familia: vielen Dank, dass ihr so ein geiler Haufen toller Freunde seid und stets für allerlei Aktivitäten und Events zu haben seid! Ganz besonderen Dank gilt natürlich meiner Familie, die mich immer unterstützt und in allen Lebenslagen für mich da ist.

Eidesstattliche Versicherung

Hiermit versichere ich, Julia Jehn, dass ich die vorgelegte Dissertation selbst angefertigt und ausschließlich die angegebenen Quellen und Hilfsmitte verwendet habe. Ich habe oder hatte die jetzt als Dissertation vorgelegte Arbeit noch an keiner anderen deutschen oder ausländischen Hochschule oder vergleichbaren Einrichtung zur Erlangung eines akademischen Grades eingereicht. Auch habe ich noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolglos oder erfolgreich beendet. Für die Anfertigung der vorgelegten Arbeit habe ich keine entgeltliche Hilfe Dritter, insbesondere eine Promotionsberatung oder -vermittlung in Anspruch genommen.

Mainz, 25.09.2019 _____

(Julia Jehn)

Curriculum Vitae

Name: Julia Jehn

Anschrift: Große Bleiche 10

55116 Mainz

Geburtsdatum: 15.04.1991

Geburtsort: Groß-Umstadt

Staatsangehörigkeit: deutsch

Schule und Studium

10/2016 – heute	Doktorandin, Biologie (Dr. rer. nat.) Johannes Gutenberg Universität, Mainz Institut für Organismische und Molekulare Evolutionsbiologie Stipendium: International PhD Programme (IPP), koordiniert vom Institut für Molekulare Biologie Mainz
08/2013 – 04/2016	M.Sc. Biomolecular Engineering Technische Universität Darmstadt Schwerpunkt: Rote Biotechnologie
08/2013 – 05/2014	ERASMUS-Programm (Teil des Masters) Universidad de Alcalá de Henares (Spanien) Schwerpunkt: Health Biology
10/2010 – 07/2013	B.Sc. Biomolecular Engineering Technische Universität Darmstadt
09/2001 – 06/2010	Abitur Friedrich Dessauer Gymnasium, Aschaffenburg

Konferenzbeitrag Posterpräsentation

Jehn J, Treml J, Wulsch S, Ottum B, Rosenkranz D. Target Pattern Analysis of 5' tRNAs halves reveals sequence-specific, Ago-independent gene regulation. *14th Microsymposium on Small RNA Biology, Wien* (Mai 2019)

Publikationen

Jehn J, Rosenkranz D. tRNA-Derived Small RNAs: The Good, the Bad and the Ugly. *Med One* (Veröffentlicht: 14. August 2019)

Gebert D, **Jehn J**, Rosenkranz D. Widespread selection for high and low secondary structure in coding sequences across all domains of life. *Open Biology* (Veröffentlicht: 29. Mai 2019)

Jehn J*, Gebert D*, Pipilescu F*, Stern S, Kiefer JST, Hewel C, Rosenkranz D. PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Communications Biology* (Veröffentlicht: 7. September 2018). (*gleichwertiger Beitrag)