

ANALYSIS OF TARGETING MECHANISMS OF tRNA DERIVED FRAGMENTS (tRFs)

by

SPYROS KARAIKOS

A dissertation submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computational & Integrative Biology

Written under the direction of

Dr. Andrey Grigoriev

And approved by

Dr. Andrey Grigoriev

Dr. Eric Klein

Dr. Jongmin Nam

Dr. Zoya Ignatova

Camden, New Jersey

January 2019

ABSTRACT OF THE DISSERTATION

Analysis of targeting mechanisms of tRNA derived fragments (tRFs)

by SPYROS KARAIKOS

Dissertation Director:

Dr. Andrey Grigoriev

We characterized *Drosophila melanogaster* tRFs, which appear to have a number of structural and functional features similar to those of miRNAs. These tRFs show a number of similarities with miRNAs, including seed sequences. Based on complementarity with conserved *Drosophila* regions we identified such seed sequences and their possible targets with matches in the 3' UTR regions. Strikingly, the potential target genes of the most abundant tRFs show significant Gene Ontology enrichment in development and neuronal function. The latter suggests that involvement of tRFs in the RNA interfering pathway may play a role in brain activity or brain changes with age.

Next, we observed different behavior of two types of tRNA fragments (3' and 5' tRFs) detected in significant numbers in rat brains. These fragments showed dynamic changes with age and 3' tRFs were found to be increasing from young to mid-aged to old rats while 5' tRFs displayed less consistent patterns. Further, 3' tRFs showed a narrow range of sizes

compared to 5' tRFs suggesting different biogenesis mechanisms. Putative targets of these fragments were found to be enriched in neuronal and developmental functions.

Last, we describe interactions of human tRFs with their putative target RNAs associated with human Ago1 using Crosslinking, Ligation, And Sequencing of Hybrids (CLASH). We found that Argonaute-loaded tRFs target a wide range of transcripts corresponding to various gene types, in addition to protein-coding transcripts. In the latter, 3' UTR regions are the likely primary target of tRFs, although there is a significant number of interactions of tRFs with coding sequences and a small number of interactions between tRFs and 5' UTR regions. We also report a novel phenomenon – a large number of putative interactions between tRFs and intronic sequences. We analyzed sequences of chimeras formed *in vivo* between tRFs and their targets to identify clusters of RNA-RNA interaction signatures. We also identified enriched motifs that may be responsible for these interactions and we provide ample evidence supporting the notion that tRF “seed” sequences appear to be primarily located on the 5' end of a tRF.

ACKNOWLEDGEMENTS

Thank everyone who has aided me towards this moment. Especially lab members and
family

General Introduction

tRNA molecules

Mature tRNAs are usually less than 90 nts long and their secondary structure resembles a cloverleaf. Properly folded tRNAs contain four distinct arms: D arm, anticodon loop, T arm and a variable loop. Transfer RNAs are crucial components of the cell's translational machinery. They facilitate translation of mRNA codons into amino-acids through base pairing between the mRNA codon and the anti-codon tri-nucleotide located in the middle of the tRNA molecule. Despite the fact that there are only 64 codons encoding for 20 amino acids, the number of tRNA genes ranges in hundreds for multiple species. For example, the human genome contains more than 600 tRNA genes, ranking them among the most abundant RNA molecules in the human transcriptome [1, 2]. Such abundance of tRNA genes suggests that these molecules may have additional functions and properties. Hence, it is not surprising that recently there has been an explosion of reports describing abundant levels and potential novel functions of their fragments (tRFs) in cells of different species.

tRNA-derived small RNAs

Traditionally, transfer RNAs (tRNAs) have been seen as critical components of the translational machinery acting as adaptor molecules in the transfer of amino acids. Recently there have been multiple attempts to understand them as regulatory molecules. The use of next generation sequencing (NGS) has enabled the detection of populations of these small RNAs in a more easily detectable and identifiable manner, lending to deeper analysis of them.

tRFs are derived from tRNAs, a very conserved type of RNA which suggests a primitive RNA silencing pathway. There are organisms like yeast which lack miRNAs, however tRNAs are reported across all kingdoms of life. They were first discovered in bacterial cells under specific stress conditions, serving as a protective response. Later studies reported tRF molecules to be present in protozoa, zebrafish, mouse and human[3-8]. Since then, multiple studies have contributed significantly towards understanding tRFs as regulatory molecules. There are two main species of small RNAs that derive from tRNAs and they are categorized based on length and biogenesis, tRNA halves and tRNA-derived fragments (tRFs). tRNA halves we discovered first and are usually generated from the mature tRNA, which is cleaved in the anticodon loop portion of the tRNA. Their length may range anywhere from 28 to 40 nts [9, 10]. These halves are generated and cleaved usually after enduring stress, including starvation, temperature stress, hypoxia and oxidative stress

[11]. Cleavage is carried out by the RNase A enzyme angiogenin or RNase T1 family member Rny1 in mammalian and yeast genomes, respectively. These cleaved tRNA halves can function by repressing translation, promoting stress granule assembly or directly interfering with the siRNA pathways [12]. Unlike tRNA halves, tRNA-derived fragments (tRFs) are shorter (~16-24nt) and can be classified into four distinct types based on the tRNA region from which they are generated: 5' tRF (tRF-5), i-tRF (tRF-i/tRF-2), 3' tRF (tRF-3) and 3'U tRF (tRF-1/tsRNA). The latter two types originate from the 3' end of the tRNA. tRF-3 are generated through cleavage on the D-loop after the post-transcriptional CCA addition takes place [13]. tRF-1 are generated from the pre-tRNA sequence downstream of the 3' end of the mature tRNA molecule [14]. tRF-5 are generated through cleavage on the T-loop and tRF-i are generated from the internal tRNA sequence [15, 16].

Biogenesis of tRFs

There have been multiple attempts to determine the biogenesis and function of these different types of tRFs and currently these questions are still open to investigation. Several studies suggest that the biogenesis of tRFs is similar to that of miRNAs and siRNAs. In one case, a mature tRF-5 was shown to decrease upon Dicer knockdown in HeLa and HEK293 cells. Furthermore, the same study demonstrated in vitro generation of the tRF with recombinant Dicer [3]. Another study described several tRF-3 tRFs in HEK-293 cells, which were Dicer-dependent and were 5' phosphorylated [17]. These modifications are

characteristic for Dicer products and contrast with 3' tRNA halves, which carry a 5' hydroxyl [4]. Additionally, a Dicer-dependent tRF-1 was identified in mouse ES cells and it was hypothesized to be generated from an alternative hairpin secondary structure of pre-tRNA Ile [17]. However, two alternative studies, which rely on large-scale computational analysis of sequencing libraries, report that the majority of tRF-5 and tRF-3 molecules in mammalian cells and tissues are Dicer-independent [7, 18, 19]. Concordant results for Dicer-independent generation of tRF-3 molecules have also been reported for HEK-293 cells [20]. Additionally, angiogenin which cleaves tRNAs to generate tRNA halves, is also proposed to play a role in the biogenesis by cleaving tRFs [19]. As a result, multiple pathways for tRF processing have been proposed, which may be species/context-specific [21].

Targeting modes of tRFs

tRNA-derived fragments have been hypothesized to function like or impact miRNAs, by either regulating mRNAs (similarly to miRNAs) or regulating miRNA loading and disrupting miRNA processing [3, 19, 22, 23]. tRFs have also been shown to bind to Argonaute-RISC complexes, further indicating that they participate in RISC-mediated gene silencing. For example, a tRF-1 and a tRF-3 molecule have been shown to be associated with Ago proteins in humans, suggesting a miRNA-like function [4, 8, 13, 24]. Previous studies have demonstrated regulatory function of these tRFs by postulating that they bind and repress mRNAs in a fashion similar to miRNAs and at times even compete with miRNAs. Initially,

it had been unclear if they act like plant miRNAs that are almost fully complementary to their targets, or like animal miRNAs that have a specific pairing via a “seed” region found on the 5’ end of the molecule. There have been conflicting models of such seed regions, lending to their complexity. One study has suggested a traditional miRNA-like silencing based on complementarity of the 5' seed sequence of a tRF to a short sub-sequence within a 3' UTR of a transcript; another has shown that the last 8-10 nucleotides (nts) on the 3’ end of a tRF-5 are responsible for mRNA repression. The most recent study focused on tRF-3 molecules reported that three distinct tRF-3 molecules have an identified “seed” element on the very same location that miRNA “seed” resides at position 2-8 nts. Regarding their functionality, there have been evidence that their targets and expression have been connected to metabolism, stress, and differentiation, suggesting their significance as regulatory molecules for proper cellular growth and maintenance [24-27].

Aging, neurodegeneration and small RNAs

In the context of aging, miRNAs and tRFs are differentially expressed [28]. The impact of these small RNAs on the aging process is still being unraveled, but there have been studies confirming the differential expression of certain miRNAs, suggesting a biological impact [29-31]. In particular, we found that tRFs were expressed and also loaded in a similar manner to miRNAs with regards to aging in *Drosophila Melanogaster*. We found that the amount of Ago2-loaded tRFs significantly increased between the two stages of three and thirty days [32]. Additionally, we reported that the cytoplasmic levels of tRF-3

molecules increase with time in rat brains [33]. In both cases, predicted targets for tRFs were enriched for genes controlling neurological or brain processes granting us confidence in their biological relevance. One important example of an age-related small RNA is miR-34, that is upregulated in both *C. elegans* and *Drosophila* with age [29-31, 34]. However, although mir-34 expression changes in *C. elegans*' mutants either did not affect or extend life span [29], while in *Drosophila* either caused a shorter lifespan, or an extension of life span depending on the regulation [35]. This hints at a species-specific effect of mir-34 and potentially many other miRNAs and tRFs, illustrating the complexity of small RNA interactomes.

CHAPTER 1: AGE-DRIVEN MODULATION OF tRNA-DERIVED FRAGMENTS in DROSOPHILA AND THEIR POTENTIAL TARGETS

This work has been published as follows:

Age-driven modulation of tRNA-derived fragments in Drosophila and their potential targets

Spyros Karaiskos ^{1#}, Ammar S. Naqvi ^{1#}, Karl E. Swanson¹ and Andrey Grigoriev ^{1*}

¹ Department of Biology, Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey 08102, USA

* Correspondence: Andrey Grigoriev, Department of Biology, Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey 08102, USA
andrey.grigoriev@rutgers.edu

These authors provided equal contribution to this work.

Reviewers:

This article was reviewed by Eugene Koonin, Neil Smalheiser and Alexander Kel.

Keywords:

RISC, Argonaute, Aging, Small RNA, ncRNA, tRNA, tRF

Contribution

My contribution to this study involves developing the pipeline to analyze the NGS data. In particular, I generated Figures 3-6 and I implemented the 7-mer seed identification pipeline. Finally, I also participated in writing the manuscript with emphasis on the results and materials and methods sections

Abstract

Background:

Development of sequencing technologies and supporting computation enable discovery of small RNA molecules that previously escaped detection or were ignored due to low count numbers. While the focus in the analysis of small RNA libraries has been primarily on microRNAs (miRNAs), recent studies have reported findings of fragments of transfer RNAs (tRFs) across a range of organisms.

Results:

Here we describe *Drosophila melanogaster* tRFs, which appear to have a number of structural and functional features similar to those of miRNAs but are less abundant. As is the case with miRNAs, (i) tRFs seem to have distinct isoforms preferentially originating from 5' or 3' end of a precursor molecule (in this case, tRNA), (ii) ends of tRFs appear to contain short "seed" sequences matching conserved regions across 12 *Drosophila* genomes, preferentially in 3' UTRs but also in introns and exons; (iii) tRFs display specific

isoform loading into Ago1 and Ago2 and thus likely function in RISC complexes; (iii) levels of loading in Ago1 and Ago2 differ considerably; and (iv) both tRF expression and loading appear to be age-dependent, indicating potential regulatory changes from young to adult organisms.

Conclusions:

We found that *Drosophila* tRF reads mapped to both nuclear and mitochondrial tRNA genes for all 20 amino acids, while previous studies have usually reported fragments from only a few tRNAs. These tRFs show a number of similarities with miRNAs, including seed sequences. Based on complementarity with conserved *Drosophila* regions we identified such seed sequences and their possible targets with matches in the 3' UTR regions. Strikingly, the potential target genes of the most abundant tRFs show significant Gene Ontology enrichment in development and neuronal function. The latter suggests that involvement of tRFs in the RNA interfering pathway may play a role in brain activity or brain changes with age.

Background

Transfer RNAs (tRNAs) have been traditionally seen as key players in protein translation, but recently there have been multiple attempts to understand them as regulatory molecules [7, 27, 36]. There are two main species of tRNA-derived small RNAs that are categorized based on length and biogenesis, including tRNA-derived small RNAs (tsRNAs, ~28-40 nt) and tRNA-derived fragments (tRFs, ~16-24nt) [9, 10]. In this study, we focus specifically on tRFs, represented by three different fragment types based on cleavage pattern. One type is produced from the tRNA 5' part (ending before the anticodon loop), while the other two types originate from the 3' region, and contain either multiple uracils or a CCA modification at the end [5, 7, 37]. There have been various attempts to determine the biogenesis pathways and potential cleavage events that make these tRFs distinct from one another [5, 21, 22, 36-39].

Previous studies have demonstrated regulatory function of these tRFs by postulating that they bind and repress mRNAs in a fashion similar to microRNAs (miRNAs) or even compete with miRNAs [7, 10, 21, 25, 37, 40, 41]. It is unclear if they act like plant miRNAs that are fully complementary to their targets, or like animal miRNAs that have a specific pairing “seed” region. Conflicting models of such seed regions have been proposed. One of them has suggested a traditional miRNA-like silencing based on complementarity of the 5' seed sequence of a tRF to a short sub-sequence within a 3' UTR of a transcript [22]; another

has shown that the last 8-10 nucleotides (nts) on the 3' end of the tRF in the 5' portion of the full tRNA are responsible for mRNA repression [42].

In the present study, we elucidated tRF/mRNA pairing further by developing a computational approach and a pipeline analogous to miRNA seed-pairing studies [43-45]. Searching for conserved regions among 12 *Drosophila* species, we predicted tRF seeds and hybridization patterns similar to that of miRNAs. In a striking parallel to the experimental observations, we also found cases of both 3'- and 5'-located potential seeds for different tRF species. Some of the functions of tsRNAs/tRFs have been connected to stress, metabolism, and differentiation suggesting the species may be critical regulatory molecules for proper cellular growth and maintenance [21, 22, 24, 25, 27, 37, 38, 42, 46]. Expanding this functional catalog in our study, we observed significant enrichment in neuronal function and development among potential targets of the prominent tRF isoforms.

We further analyzed the association with age. Recent studies have highlighted that miRNAs are associated with the aging process, showing differential isoform expression and differential RISC loading of specific miRNAs with age, related to modifications on the 3' end, including untemplated additions, 2'-O-methylation or imprecise Drosha/Dicer cleavages [35, 47]. Here, we present a follow-up computational analysis of the same deep-sequencing libraries, this time focusing on tRFs originating from multiple tRNAs. In

addition to the *in silico* prediction of seed regions, we examined changes in individual tRF isoforms with age. This unexpectedly revealed diverse patterns, resembling those of miRNA and suggesting that tRFs may impact age-associated events, while simultaneously being modulated with age. Taken together, these findings confirm that despite the lower counts in deep-sequencing experiments, tRFs represent not degradation products but potentially important players in Argonaute pathways, increasing our understanding of these regulatory molecules.

Results

Using four different *D. melanogaster* small RNA libraries, including co-immunoprecipitations of Ago1 and Ago2 in flies aged 3 days and 30 days [47], we observed striking patterns of age-dependent expression, structure and preferential loading of tRFs into RISC complexes. Following the similarity of tRF features with miRNAs, we predicted potential targets for further experimental validation that would be the ultimate test of the biological functionality of tRFs.

Read Distributions of tRNA Fragments are Similar to miRNAs

The read distributions mapping to known miRNAs usually show an asymmetry favoring the mature arm of a given miRNA stem-loop sequence. This is usually demonstrated by observing a high relative frequency of the reads aligning to one of the arms (5' or 3'). At times, we also observe reads that originate from the middle or loop section, which is inferred by a very low frequency of reads mapping to the middle section of the RNA molecule. This type of visualization is particularly useful because it may shed light into potential 5' or 3' modifications, which may include alternative cleavage sites, deletions, non-templated additions, and RNA editing events [48-50].

We investigated whether tRF-tRNA alignments displayed similar patterns to miRNAs in the read distributions. First, we found that tRF reads, which were more abundant in the Ago2 libraries, mapped to >100 nuclear and mitochondrial *Drosophila* tRNA genes covering the whole spectrum of 20 amino acids. This is in contrast to previous studies, which have usually reported fragments from only a few tRNAs [22, 26, 42, 46]. We also observed multiple isoforms of the same tRF being expressed. Interestingly, these mappings showed very specific patterns: the reads typically aligned to either the 5' or 3' region of the tRNA molecule, and often had identical start positions or presumed cleavage sites (see below). One of the tRF ends in these cases matched the respective end of the host tRNA, while the other showed some variability comparable to that observed in miRNA [35, 47]. In other words, the distribution of reads that mapped appeared as non-random and precise as those of miRNAs, strongly suggesting that their source was not indiscriminate degradation but rather a targeted biological process.

All detected *Drosophila* tRFs and their relative read distributions in visual format can be found on our website [50]; here we illustrate the findings with the two examples of tRFs of different level of abundance, AlaAGC and MetCAT tRFs (Fig. 1). As was typical for most tRFs, the read distributions invite comparisons to a canonical miRNA structure, suggesting that specific cleavage mechanisms may be at work. We observed clearly defined boundaries for 5' and 3' regions. The uneven read distribution allows one to speculate that, in case of AlaAGC (Fig. 1A), the 5' arm is the analog of a miRNA mature and/or functional strand, while the 3' arm is similar to a passenger strand (that would eventually be degraded). The low frequency of reads mapping to the middle region is akin to miRNA loop regions. MetCAT is an example of the opposite case of prevalent read counts in the 3'-end (Fig 1B). Generally, the majority of tRF reads showed a miRNA-like asymmetric distribution by aligning to either the 5' or 3' region of the tRNAs.

Age-associated Global Shift of Ago1 vs Ago2-loaded tRFs

A number of further similarities to miRNAs were suggested by the association of the tRFs with the Argonaute proteins (Ago1 and Ago2) of the two RISC complexes. Previously, we have analyzed Ago1 and Ago2 loading of microRNAs and found age-specific patterns [47].

As with miRNAs, we observed that the total levels of Ago-loaded tRFs changed with age. In Ago1 the normalized read counts for 3 days and 30 days stayed relatively constant at ~5,000. In contrast, in Ago2 there was a 4-fold increase (from 5,000 to ~20,000 normalized total read counts) between 3 to 30 days. Amongst tRFs with counts >100 (arbitrary threshold for illustrative purposes), 8 were downregulated and 4 upregulated in Ago1, while all 40 Ago2-associated tRFs were upregulated with age, indicating possible functional importance in an age-related manner.

Further investigating this result, we determined whether the differences in loading into Ago2 reflected an increased association of specific isoforms over others. This particular phenomenon is seen in miRNAs [47], so it was of interest to assess if there was a similarity in tRF behavior. We first identified two tRFs, GluCTC and AspGTC, that displayed multiple isoforms in both the Ago1-IP and Ago2-IP libraries and that also showed differential loading with age, with the most abundant isoform changing two-fold or more (Fig. 2). For GluCTC we observed the same isoform, the 25mer, being loaded onto both RISC complexes, but in Ago1 it showed a decrease with age, while in Ago2 it showed an increase with age, hinting at a mechanism that either actively partitions these fragments at the loading step in the biogenesis pathway or contributes to their retention with age when loaded to Ago2 (Fig. 2A-B). In the case of AspGTC, the isoform (29mer) that is most abundant was not detected at all in Ago1, while it was readily loaded into Ago2, which also showed increased loading with age (Fig. 2C-D).

Further, we considered loading ratios of 30 days to 3 days for each tRF. Our findings indicated that loading onto Ago2 increased at 30 days, while Ago1 loading decreased or stayed the same as at 3 days (Fig 3). Not all the tRFs are shown: e.g., Gly-related ones did not have any reads in the Ago1 libraries, and no reads were found in Ago1 for the major Ago2 isoform of AspGTC tRF depicted in Fig. 2. In several cases, distinct fragments from different tRNA genes with the same anticodon were detected, e.g., for GluCTC. When tRF sequences allowed us to distinguish such tRNA genes, we named them tRNAgene-1, tRNAgene-2, and so on (note that a union of all GluCTC isoforms in Fig. 2 corresponds to GluCTC-2 in Fig. 3). When tRF sequences could be assigned to more than one of such tRNA genes, we assumed all of these genes contributed equally to the observed tRF counts.

We then examined tRFs that were both Ago1- and Ago2-loaded in order to ascertain any age preference. We specifically looked at tRFs at the two different time points and compared their ratios in Ago2- and Ago1-associated libraries (Fig. 3). At 3 days, we observe that the ratios are either below 1 or very close to 1, with the exception of GluTTC. Thus at 3 days, there is either a preference for Ago1 or no preference at all. However, at 30 days the reverse is the case: for most tRFs we detected at least a two-fold increase in Ago2 loading. Hence, tRFs are more likely to be loaded onto Ago2 and not Ago1 in older flies, confirming an age preference amongst loaded tRFs.

We next focused specifically on the tRF species containing CCA at the 3'-end and examined their accumulation with age in both Ago1 and Ago2. Such species showed a two-fold increase in Ago1 libraries (6% to 12%) from 3 days to 30 days, and even higher in Ago2 libraries (6% to 16%), supporting the notion that fragments of mature tRNAs contribute to the global increase of loading with age.

Together, these data support the idea that the loading patterns of tRFs between Ago1 and Ago2 change dramatically with age, such that Ago2-loading of select isoforms increases, while Ago1-loading of tRF isoforms belonging to the same tRNA decreases. These results are similar to findings of age-dependent loading of miRNAs [47] and they also indicate that there may be distinct pathways for Ago loading by recognizing, partitioning or retaining specific isoforms, which may change as a function of age.

Seed Sequences in Conserved Regions

The mechanism of tRF action upon loading into Ago1 and Ago2 still remains unclear, but there are clues to suggest a miRNA-like pathway of execution. For example, the fragments have been detected in the cytoplasmic fraction of cells [42], several studies have shown trans-silencing capabilities of tRFs, and the silencing of a mock mRNA fully complementary to a tRF [41, 51] has been demonstrated. Some authors [22, 46] suggest a traditional

miRNA-like silencing based on complementarity of the 5' seed sequence of a tRF to a short sub-sequence within a 3' UTR of a transcript [43-45]. Another study, however, suggests a 3' seed sequence, while ruling out a 5' or a mid-tRF seed binding [42].

To further explore the notion of mRNA-targeting, we developed a computational pipeline to detect a potential location of a seed sequence (analogous to that of animal miRNAs) in the tRFs. In miRNAs, 3'-compensatory sites [52] and central pairing sites [53] have been reported in addition to the most prevalent 5' seeds [42-45]. For seed finding we followed the same approach used to identify the seed sequence in microRNAs [42-45], with short sequence windows sliding along the tRF sequence, without any location constraints. Then we found exact matches of the reverse complements of these sequence windows to the 5' UTR, 3' UTR, intron and exon (CDS) sequences in the *D. melanogaster* genome and in the conserved portions of these regions of the 12 *Drosophila* genomes [54]. For a window of length k we then compared the observed match counts with those expected by chance (estimated from k -mer genomic frequency) and with the mean frequency of all other possible k -mer sequences produced by reshuffling the nucleotides in the window. In agreement with our conjecture that tRFs may harbor miRNA-like short seed sequences, 7-nt windows showed good discrimination between the conservation levels of 5' and 3' ends of the most tRFs we analyzed (Fig. 4).

The results for the most abundant Ago2-loaded Gly-derived tRFs detected in our studies strongly supported the seed location on the 3' end of tRFs (Fig. 4A-B,D-E). We observed that the tRF GlyGCC 7mer located at position 12 (to 18) has the highest frequency of reverse complement occurrences in the conserved regions of *Drosophila* genomes (regions associated with >14,000 genes in total), making it a candidate seed sequence (Fig 4A,D). A very similar tRF, GlyTCC (attcccggccgaCgcacca), contained a one nucleotide difference to GlyGCC (attcccggccgaTgcacca) and had a candidate seed located at position 11 (to 17), shifted one nucleotide towards the 5' end (Fig 4B,E).

We found no overlap between the lists of *D. melanogaster* transcripts with matches to the seeds of GlyGCC compared to GlyTCC. Thus, although a single nucleotide difference in/near the seed region may influence tRF targeting and hybridization, it is remarkable that a very different set of conserved sequence matches/potential targets still corresponds to the same 3' location of the seed sequence. Though many tRFs showed a peak similar to those in Fig 4A-B, we noted that a few tRFs showed such peaks in the 5' region, suggesting a 5' seed targeting. For example, in the tRF mt:SerGCT the 7mer window matches peaked at the 5' end of the sequence (Fig. 4C), as opposed to a 3' end maximum found in the Gly-related tRFs. Thus we also observed potential seeds on both 5' and 3' ends of tRFs, in parallel to what was detected experimentally.

The enrichment in the counts of matches for potential seed sequences is very prominent in the conserved genomic regions (Fig. 4). The frequency of their 3'UTR matches far exceeds the expected frequency (between five and several hundred fold). At the same time, the seed matches are not among the most frequent heptamers in the *D. melanogaster* genome (e.g., the genomic frequency of the potential seed match in mt:SerGCT is less than half of the top heptamer). All these facts point to a possible functional role for the seed sequences, similar to those of miRNA seeds.

Intron sequences also produced much higher numbers of matches than expected (or than reshuffled tRF 7mers), when conserved *Drosophila* introns were analyzed (Fig. 4D-F). In GlyTCC, the intronic matches even slightly exceeded the 3'UTR matches, while the seed peaks in both cases were on the same 3' end of the tRF (Fig. 4H). This

Potential Targeted Regions in mRNA

To find potential targeted regions, we compared the 5' UTR, 3' UTR, intron and exon parts of genes in the *D. melanogaster* genome. Per unit of length the 3' UTR regions matched the potential tRF seeds most frequently (Fig. 5) suggesting a prevalence of a 3' UTR targeting mode. This further supports the idea that tRFs may behave similar to miRNAs.

We observed significant enrichment of the 3'UTR for mt:SerGCT seed matches in the *D. melanogaster* genome ($p < 0.001$), both among random heptamers and among reshuffled nucleotides comprising the seed. The Gly tRF seed matches, with less extreme AT-richness, did not show such enrichment. However, we note that shuffling of the seed sequence is not an ideal random model and statistical testing of the tRF seed regions is complicated by the fact that a tRNA sequence is under multiple selective constraints for its structure and function related to translation (and furthermore different from the constraints of a miRNA).

We also scanned for nearly perfect complementary matching between full-length tRFs and 3' UTRs, which would inform us if some of these tRFs acted like plant miRNA. This analysis, however, yielded no significant results, suggesting that the tRF binding mode may be more consistent with animal miRNAs.

Assuming the latter (animal-like) binding mode, we observed a variety of seed sequence matches in the conserved fly genome regions. As with miRNAs, there were 7mer-m8, 7mer-1a and 8mer-1a match types. These types have been studied and confirmed previously for miRNAs [43] and are as follows. 7mer-m8 is a match of 7 nts (Fig. 6A and Fig. 6B). 7mer-1a and 8mer-1a refer to matches of first 6 (Fig. 6C, GlyGCC) or 7 (Fig. 6C, mt:SerGCT) nucleotides of the seed, respectively, followed by an extra A (added to the

elongated match region in the Fig. 6C). All three illustrated targets (Fig. 6A-C) possess 3' UTR regions highly conserved amongst all 12 *Drosophila* genomes analyzed.

Notably, some seeds showed overlap with the seed of either another tRF or a miRNA (Fig. 6A and Fig. 6C). For example, both GlyGCC and mir-277 seeds overlap by 5 nts and this sometimes led to their complementarity against the same target (Fig. 6A). Such overlaps could theoretically lead to competition of tRFs and miRNAs for the same targets, potentially adding another layer of complexity to the regulatory processes.

As demonstrated by our results, there is clear evidence that tRFs interact and are loaded onto Argonaute proteins and may target the 3' UTR regions of mRNAs, suggesting a potential post-transcriptional regulatory mechanism similar to that of miRNAs. The fact that the candidate seeds aligned predominantly to the 3' UTRs indicates that one of the mechanisms for suppression may be translational inhibition. Alternatively, some tRFs may employ mRNA cleavage for regulation, since we observed CDS regions that also aligned to our candidate seeds [43-45].

Gene Ontology Analysis of Potential Targets

Given the difference in seed localization, we predicted targets for the divergent cases of the Gly and mt:SerGCT tRFs. Following the link between the Ago-loading change of miRNA and brain degeneration with age [47], we assessed whether targets of these tRF were also associated with a particular biological process. Using the identified seed sequences, we sought targets for the tRFs in the *D. melanogaster* genome based on perfect matches to 3'UTRs. We then conducted a gene ontology (GO) enrichment analysis using the AmiGO 2 software [55] to understand the nature of the predicted targets.

Stringent criteria for enrichment revealed several interesting trends. Notably, neuronal and developmental processes were the most dominant among the significantly enriched terms ($p\text{-value} < 0.001$) belonging to the GO category "biological process". In particular, for GlyGCC we observed 52% of enriched GO terms related to development and 15% related to neuronal function, while for mt:SerGCT these numbers were 39% and 12%, respectively (Additional File 1). In the GO analysis, the most populated process terms (if one counts potential targets, described by these terms) are often generic ones, like "biological process" or "biological regulation". For both of these tRFs, the most populated GO terms after the generic ones were GO:0032502 (developmental process) or GO:0048856 (anatomical structure development). Pertinent to the tRF involvement in the neuronal regulation, synapse- or axon-related GO terms accounted for 20% (in

mt:SerGCT) to about half (in GlyGCC) of the significantly enriched terms ($p\text{-value} < 0.001$) in the category "cellular localization" (Additional File 2). The targets, exemplified in Fig. 6, belong to these GO categories, e.g., Dlar, a targeted gene of mt:SerGCT (Fig. 6B) is a conserved member of the tyrosine phosphatase family with a fundamental role in axon targeting/development and organization of actin filaments [56, 57]; see Discussion). For the category "molecular function", terms related to DNA and RNA binding (with variations including regulatory region or nucleotide binding) were frequently enriched for mt:SerGCT and GlyGCC (Additional File 3).

Alternative polyadenylation and longer 3'UTRs have been observed in the transcripts in fly brains [58] and we checked if that affected our results. For all targets found above (using longest annotated 3'UTRs) we selected the shortest annotated 3'UTRs (if those were available) and again searched for matches to the seeds. As expected, there was a reduction in the numbers of both seed matches and corresponding targets. However, the reduction for the brain-associated genes (54.8% of matches and 60.1% of targets remaining) was very similar to the rest of tRF targets (53.4% of matches and 61.3% targets remaining) and this factor could not explain the GO term enrichment described above. As for the length itself, the coefficients of variation in both target sets were very high (1.24 and 1, respectively) thus the length difference was not significant between these subsets of genes.

Discussion

In this report we characterized tRFs found in Ago1 and Ago2 IP libraries from *Drosophila* to reveal expression and loading patterns in the context of age. We also identified potential targets and a likely mode for targeting.

We identified tRFs in both Ago1 and Ago2 co-immunoprecipitated libraries, indicating miRNA-like functionality of loading of these tRFs into RISC complexes. Alignment to the mature tRNA sequence revealed a high read-depth on one side of the tRNA molecule and size distributions of 16-30 base pairs in length, which suggests a similar structural motif as miRNAs. Although the library was size-selected for these distributions, we observed very precise boundaries of tRFs (similar to those in miRNA [35, 47]), strongly suggestive of a biological process rather than random degradation responsible for their generation. However, given the isoform diversity, limited degradation effects on the tRF ends cannot be ruled out, and their scale is comparable to “nibbling” in miRNA [35, 47].

By examining age-associated patterns of tRF expression, we saw distinct isoforms changes in age-dependent manner in *Drosophila*. For example, for GluCTC we observed the same isoforms present in both Ago1 and Ago2 libraries, but an increase in individual isoforms in Ago2, and a decrease in Ago1, especially for most abundant or major isoform.

Additionally, the major isoform of AspGTC in Ago2 was not present at all in Ago1 (see Fig. 2). These types of change are correlated with a shift in loading of these fragments into Ago2 vs Ago1 with age. Thus, the partitioning of multiple tRFs between Ago1 and Ago2 may be a coordinated process modulated with age in *Drosophila* (see Fig. 3).

One possible explanation proposed for the observations of differential miRNA loading with age (which can be extended to tRFs) is that the cells are adjusting their regulatory processes for upcoming age-associated stresses [47]. Since Ago2-mediated translational silencing causes retention of the polyA tail [59], Ago2-association might make it possible to respond to age-associated internal or external stimuli more rapidly and effectively by allowing for re-activation of target mRNAs. Ago2 mutants have been shown to develop neurodegenerative phenotypes in the study of miRNA involvement in the aging process [47]. This may serve as further support of our predictions of the tRF regulatory function since in these mutants the disrupted stabilizing modification, lack of tRF RISC loading, and subsequent deregulation of the neuronal targets could further contribute to such phenotypes. One possible target of GlyGCC and mt:SerGCT (see Fig. 6C), the gene *Atg8a*, is intimately linked to aging pathways, e.g., the insulin/IGF-signaling pathway that mediates the lifespan in *Drosophila* through Smad binding [60].

Other modes of tRF-driven regulation have been proposed, from inhibiting translation initiation factors to direct interaction with ribosome, etc [7, 10, 21, 25, 37, 40, 41]. Given

the base pairing in the tRNA stems, one cannot exclude potential interaction with full-length host tRNAs or their fragments. While this paper was under review, a possible role of tRFs as tumor suppressors binding to oncogenic RNA-binding protein YBX1, displacing pro-oncogenic transcripts has been described [61]. However, the patterns of conservations we observed indicate a clear possibility of miRNA-like targeting.

Although the exact mechanism is still being unraveled, our results suggest a short seed region in tRFs that is key for recognizing potential mRNA targets. While for animal miRNAs the 5' seed location is most common, 3'-compensatory sites [52] and central pairing sites [53] have been reported. In our examples, the Gly-associated tRF in *Drosophila* has a putative 3' seed region, while the mt:SerGCT tRF has a 5' seed. Thus, in parallel to experimental data showing two possible seed locations [22, 42, 46], our results demonstrate that regions of conservation can be present at either the 5' or the 3' end in different tRFs. We also provide evidence that the 3' UTR may be where targeting occurs, allowing us to speculate that the mode of action may include translational repression or mRNA cleavage.

Alternatively, some tRFs may employ mRNA cleavage for regulation, since we observed CDS regions that also aligned to our candidate seeds [43-45]. Enrichment of seed matches in the conserved intron regions may also indicate a role of tRFs in alternative splicing and transcriptional regulation, given the evidence of Ago2 involvement in these process in the

nucleus [62]. The enrichment of targets involved in development may be of particular interest in this regard as Ago2 transcriptional target genes are also bound by Polycomb group transcriptional repressor proteins and change during development [62].

Drosophila Ago1 and Ago2 employ different mechanisms to silence target mRNAs and in particular Ago2 mutants show neurodegeneration and a shortened lifespan [47]. The fact that most tRFs are loaded and/or show a dramatic change in loading with age in Ago2 suggests that these small RNAs may also be involved in such pathways. In this regard, it is notable that despite the difference in seed localization (and no common targets), putative targets of tRFs from both mt:SerGCT and GlyCTC are significantly enriched in developmental and neuronal functions (Supp. Table 1-3). Further, we found that these target lists overlap (with up to 29 targets) with the well-studied miRNAs mir-34, mir-277, mir-190, and mir-10. All of these miRNAs impact brain function, affecting neurodegeneration, bi-polar disorder, and schizophrenia [35, 63, 64], in agreement with our predictions of tRF influence on the brain and age-related events. An overlap of the tRF seed with that of mir-277 is of importance, as it may relate one of the most abundant tRFs (GlyGCC) to brain deterioration, since mir-277 has been reported to modulate neurodegeneration [65].

Amongst the common targets of GlyGCC and mir-277, we observed Dlg (FBgn0001624), coding for the *Drosophila* discs large tumor suppressor protein (see Fig 6A). The

mechanism of regulation of this gene would be of interest since it has been previously associated with neuron development [18, 66] and it also shows homology with a human tumor suppressor protein [67]. Another common target, Toll-7 (FBgn0034476), may also be of significance, since it acts as a neurotrophin receptor and neurotrophism is only starting to be elucidated in insects [68].

Some of the tRF targets in the significantly enriched GO categories are closely related to the RNA regulatory pathways, e.g., Fmr1 (FBgn0028734, a homolog of the fragile X mental retardation 1 gene in human). This is an RNA-binding protein that interacts with the RISC complex itself and is necessary for proper development [69-71]. Of note, this gene is located in the *Drosophila* genome in the immediate vicinity (a few hundred basepairs) of mir-34 and mir-277, hinting at a potentially deeper regulatory connection.

Conclusions

This is the first time such a detailed analysis has been performed on tRFs. We developed a robust pipeline to identify candidate "seed" regions that clearly showed a stronger binding pattern based on specific positions, restricting it to the 5' or 3' end and a binding preference for 3' UTRs. The results reveal tRFs features that in many respects resemble structural and functional properties of miRNAs and strongly suggest that these small RNAs are not simply tRNA degradation products, but are specific, biologically-generated species. The targets predicted with candidate seeds showed enrichment in processes

related to neuronal function and development, hinting at the biological significance of these tRF molecules. Thus, the trends observed with tRFs likely represent bona fide targeted processing of tRNAs, and the tRF association with different RISC complexes in the context of age may reflect an important regulatory function.

Methods

Mapping and quantifying tRFs

We used *Drosophila* Ago-IP libraries GSM1278635, GSM1278636, GSM1278637 and GSM1278638 available from the Gene Expression Omnibus (GEO) database, with experimental details described earlier [47]. Adaptor sequences were removed from the 3' end of the reads in the Illumina fastQ files using the fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The adaptor sequences are as follows:

5' adapter = 5'- GUUCAGAGUUCUACAGUCCGACGAUC- 3'

3' adapter = 5'- TGGAATTCTCGGGTGCCAAGG- 3'

Reads were then collapsed and annotated with the number of times each was sequenced, so only unique reads were analyzed. The reads were then mapped using Bowtie to the D.

melagonaster (dm5) genome and tRNAs obtained from FlyBase. Bowtie parameters were restricted to only output perfectly aligned matches to the tRNA sequence. The reads were aligned and mapped to the entire tRNA sequence with the CCA addition. After mapping reads to their respective tRNAs, each library was independently normalized by the total number of reads mapped to the *D. melanogaster* genome (v. R6.03).

Differential/Preferential loading with age

We identified differential loading of tRFs with age in Ago1 and Ago2 using a ratio metric. We first identified the most abundant isoform in our 30 day libraries and used the read count numbers of that specific isoform for our ratio calculations. We calculated the ratio of 30 days to 3 days for Ago1 and Ago2 of highly abundant (1000 or more reads) tRFs. We then plotted the ratios to see loading changes that may occur with age.

To observe what was preferentially loaded (Ago1 vs Ago2) with age, we obtained a different ratio. The ratio of this measure was the ratio of reads of a particular tRF of Ago2 to Ago1 at 3 days and at 30 days.

Analysis of Seeds, Targeting and GO Terms

In order to identify a potential seed sequence in our dataset, we generated k-mer subsequences of the tRF by applying a sliding window by shifting one nt towards the 3'

end after each subsequent k-mer generation. We then found exact matches for each of these subsequences to the conserved 5' UTR, 3' UTR, exon and intron regions of 12 *Drosophila* genomes provided by UCSC [72] and to those regions in the *D. melanogaster* genome. We then compared for each k-mer in a tRF the observed number of its matches in the conserved 3' UTR regions with the expected number (based on the frequency of matches across the *D. melanogaster* genome) and with the average number of matches of all possible k-mers with the same nucleotide composition in conserved 3'UTRs to identify candidate seeds. Genes with exact matches of 7mer candidate seeds to the longest annotated 3'UTR were considered potential targets. While our approach is similar to TargetScan [43-45], we did not use its "context score" as it was unlikely to be applicable for our cases of both 3' and 5' seeds. To find the preferentially targeted regions we normalized the total match counts by the total length of each respective set of regions. AmiGO [55] was used to find enriched GO-terms in our target list for each tRF.

Abbreviations

tRF: Transfer RNA fragment; tsRNA: tRNA-derived small RNA; ncRNA: Non-coding RNA; UTR: Untranslated region; RISC: RNA-induced silencing complex; GO: Gene ontology.

Competing interests

The authors declare that they have no competing interests.

Contributions

AN performed sequence alignments, tRF and isoform identification and characterization, drafted the initial version of the manuscript. SK identified seeds and targets, analyzed enrichment, performed gene ontology analysis, generated most figures and helped to draft and revise the paper. KS with participation of AN, SK and AG created the public repository of the data and helped to draft the paper. AG conceived the study, participated in its design, execution and coordination, and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We are grateful to Nancy Bonini and members of her lab for the sequencing data generation, discussions and insightful comments on the earlier versions of this paper. We thank Kevin Abbey for excellent technical support. This work was funded in part by the National Science Foundation (grant DBI-1126052 to A.G.).

Figures and Legends

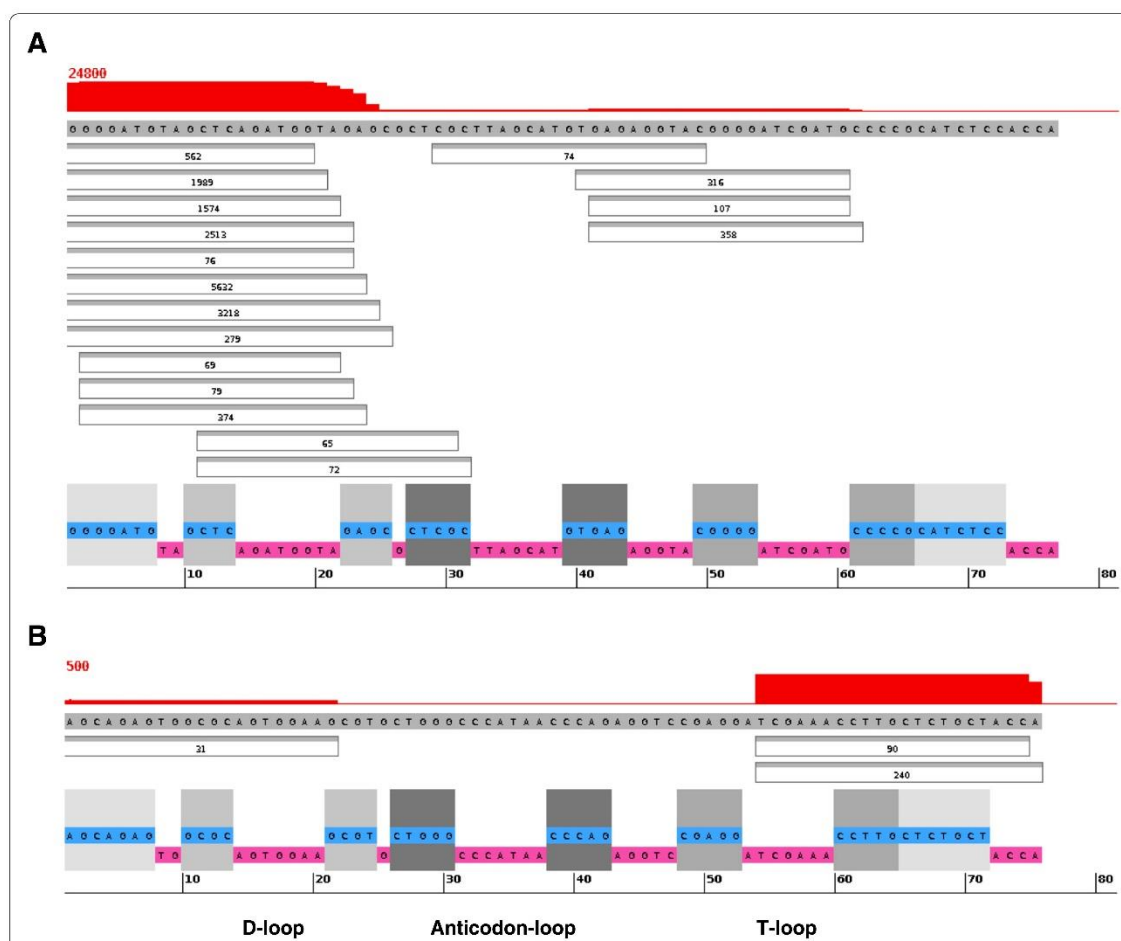


Figure 1. Examples of Read Distribution Patterns of tRFs.

Screenshot of our RNA display [73], showing reads that align to tRNA-Ala in the Ago2, 30 days library (A) and MetCAT in Ago1, 30 days (B). Sequence at the bottom with the magenta background indicates single-stranded (loop) regions in the tRNA molecule, while the cyan background and matching grey boxes indicate stems. The red on top indicates read depth coverage of specific regions of the tRNA. Reads (boxes in the middle) with

counts of at least 1% of the most abundant read are displayed; lower count reads are omitted for compact visualization.

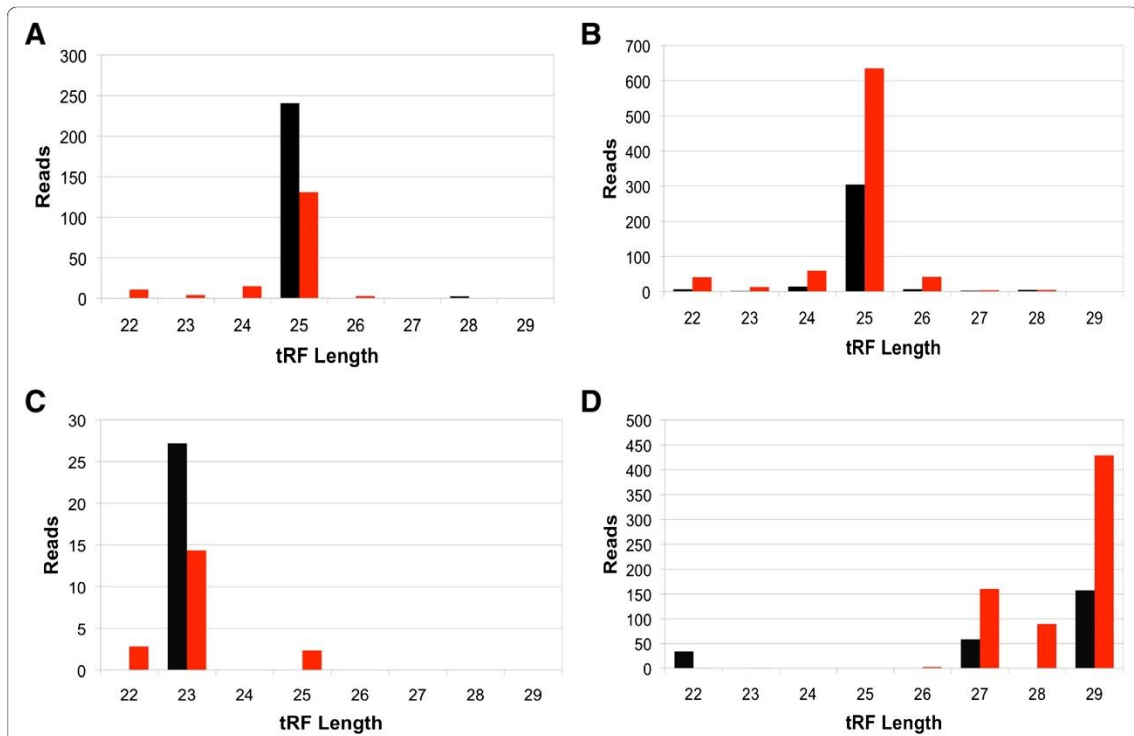


Figure 2. Distinct tRF Isoform Changes with Age.

Isoform distributions for GluCTC in (a) Ago1-IP and (b) Ago2-IP, (c) AspGTC in Ago1-IP and (d) Ago2-IP libraries. Both tRFs show a decrease in Ago1, but an increase of specific isoforms in Ago2 with age. Black bars represent normalized counts at 3 days, while red bars represent normalized counts at 30 days. The same GluCTC is also presented in Fig. 3 (GluCTC-2).

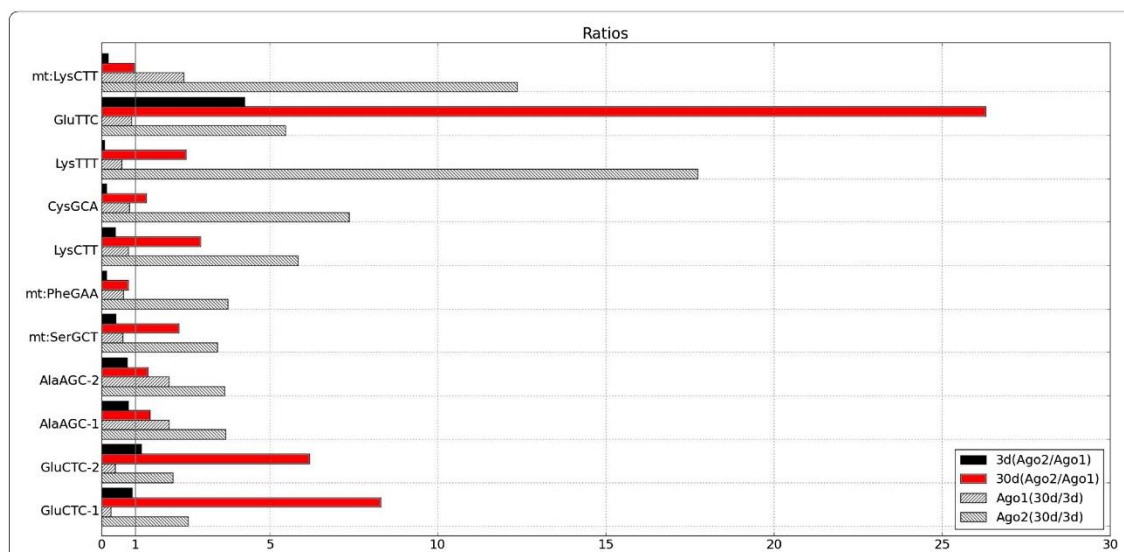


Figure 3. Differential and Preferential Loading.

Plot of abundant tRFs that are present in all four libraries. Plots show relative ratios of reads: Ago2 to Ago1 in 3 days (black) and 30 days (red); 30 days to 3 days in Ago1 (slash) and Ago2 (back slash).

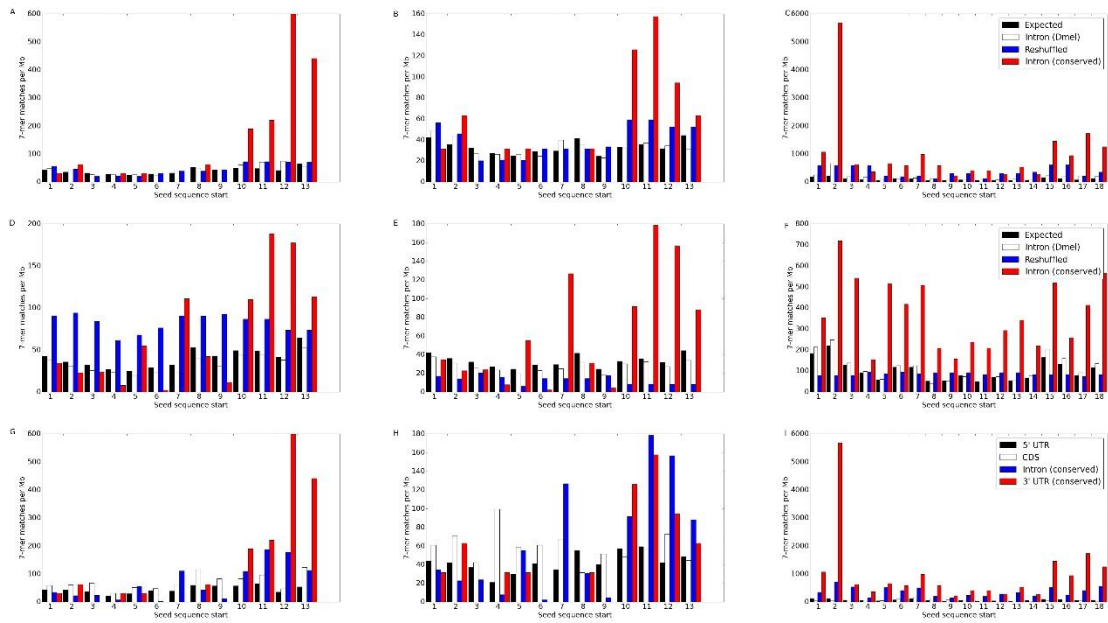


Figure 4. Candidate Seed Regions for tRFs.

The numbers of sequence matches in the 3' UTR regions (A-C) and introns (D-F) are plotted vs window start positions of 7mer windows of (A, D) GlyGCC, (B, E) GlyTCC and (C, F) mt:SerGCT tRFs in *Drosophila*. Color key for the top row is given in (C), for the bottom row in (F). Expected number of matches is shown in black and average number of matches for all other 7mers with the same nucleotide composition as the given window is shown in blue. The observed number of matches in the *D. melanogaster* genome is shown in white and in the conserved regions of 12 *Drosophila* genomes is shown in red.

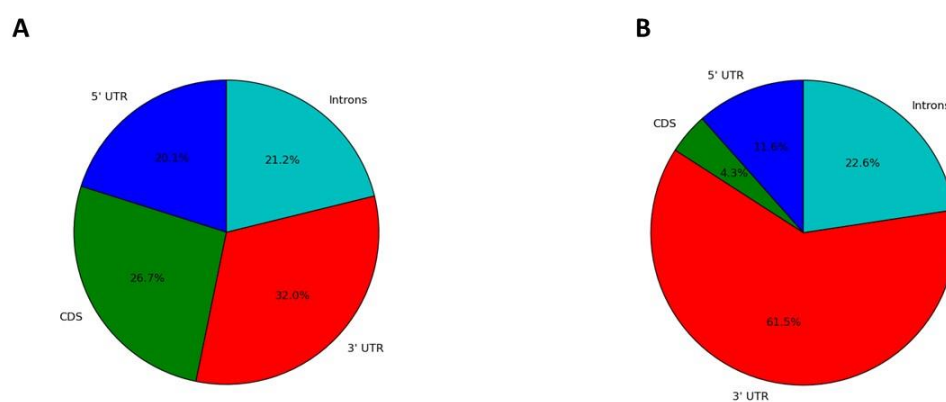


Figure 5. Percentage of Seed Alignments by Region.

The percentages by region (5' UTR, 3' UTR, introns and CDS in the *D. melanogaster* genome) of the matches to the most abundant 7mer shown in Fig. 4A and 4C.

| A | | | B | | | C | | |
|------------------------|----|-----------------------|-----------------------|-----------------------|----|-------------------------------|----|--------------------|
| FBgn0001624 ...1190... | | | FBgn0000464 ...480... | | | FBgn0052672 ...500... | | |
| Dme | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | CUGUGCAUAUUUAGU 3' | 5' | CUGUGCAUAUUUAGU 3' |
| Dvi | 5' | AAGUUGGUGCAUUUGUUG 3' | 5' | CUAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Dgr | 5' | CAGUUGGUGCAUUUGUUG 3' | 5' | CUAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Dmo | 5' | AAGUUGGUGCAUUUCUUG 3' | 5' | CUAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Dsi | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | CUGUGCAUAUUUAGU 3' | 5' | CUGUGCAUAUUUAGU 3' |
| Dse | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Der | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Dya | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Dan | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Dpe | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | AUGUGCAUAUUUAGU 3' | 5' | AUGUGCAUAUUUAGU 3' |
| Dps | 5' | ACGUUGGUGCAUUUGUUG 3' | 5' | C GAUACAUAUUUGAACU 3' | 5' | AUGUGCAUAUUUAGU 3' | 5' | AUGUGCAUAUUUAGU 3' |
| Dwi | 5' | ACGCUGGUGCAUUUGUUG 3' | 5' | CUAUACAUAUUUGAACU 3' | 5' | UUGUGCAUAUUUAGU 3' | 5' | UUGUGCAUAUUUAGU 3' |
| Consensus | | acGuUGGUGCAUUUGUUG | | CgAUACAUAUUUGAACU | | uUGUGCAUAUUUAGU | | |
| GlyGCC | | 3' CCACGUA 5' | mtSerGCT | 3' GUAUAAA 5' | | GlyGCC 3' CCACGUA 5' | | |
| mir-277 | | 3' ACGUAAA 5' | | | | mtSerGCT 3' GUAUAAA 5' | | |

Figure 6. Examples of Conserved Seed Region Matches.

Grey highlights and bold text indicate seed complementarity to conserved (12 *Drosophila* genomes) 3' UTR regions. Targeted genes with overlapping coordinates in the genome are shown on top. (a) Both GlyGCC and mir-277 having a 7mer-m8 match, (b) mt:SerGCT 7mer-m8 match and (c) GlyGCC having a 7mer-1a match and mt:SerGCT having a 8mer-1a match, with additional A for the 1a matches are also highlighted (mt:SerGCT) or bolded (GlyGCC).

CHAPTER 2: DYNAMICS OF tRNA FRAGMENTS AND THEIR TARGETS IN AGING
MAMMALIAN BRAIN

This work has been published as follows:

Dynamics of tRNA fragments and their targets in aging mammalian brain

Spyros Karaiskos ¹ and Andrey Grigoriev ^{1*}

¹ Department of Biology, Center for Computational and Integrative Biology, Rutgers
University, Camden, New Jersey 08102, USA

* Correspondence: Andrey Grigoriev, Department of Biology, Center for Computational
and Integrative Biology, Rutgers University, Camden, New Jersey 08102, USA
andrey.grigoriev@rutgers.edu

Reviewers:

This article was reviewed by Tatiana Tatarinova, Laura Kamenetzky and Natalia
Macchiaroli.

Keywords:

transfer RNA, rat brain, rat cortex, tRNA fragments, aging, non-coding RNA

Contribution

My contribution to this study involves developing the pipeline to analyze the NGS data.

In particular, I generated all figures and I implemented the seed identification pipeline. I helped write the manuscript with emphasis on the results and materials and methods sections.

Abstract

Background

The progress of next-generation sequencing technologies has unveiled various non-coding RNAs that have previously been considered products of random degradation and attracted only minimal interest. Among small RNA families, microRNA (miRNAs) have traditionally been considered key post-transcriptional regulators. However, recent studies have reported evidence for widespread presence fragments of tRNA molecules (tRFs) across a range of organisms and tissues and of tRF involvement in Argonaute complexes.

Methods

We performed a meta-analysis of small RNA sequencing and RNA sequencing datasets derived from brains of young, mid-aged and old rats in our effort to elucidate potential tRF function. We adopted our computational pipeline developed to analyze *Drosophila melanogaster* tRFs to vertebrate species. Sliding 7-mer windows across the length of a tRF

we searched for putative “seed” sequences with high numbers of conserved complementary sites within 3' UTRs across 15 to 23 vertebrate genomes. We then performed Gene Ontology enrichment analysis for the predicted tRF targets. Finally, we examined the levels of transcripts potentially targeted by tRFs and miRNAs with similar abundance patterns in the context of age.

Results

We observed different behavior of two types of tRNA fragments (3' and 5' tRFs) detected in significant numbers in rat brains. These fragments showed dynamic changes with age and 3' tRFs were found to be increasing from young to mid-aged to old rats while 5' tRFs displayed less consistent patterns. Further, 3' tRFs showed a narrow range of sizes compared to 5' tRFs suggesting different biogenesis mechanisms. Putative targets of these fragments were found to be enriched in neuronal and developmental functions. Comparison of tRFs and miRNAs increasing in abundance with age revealed small but distinct changes in brain target transcript levels for the two types of small RNA, with the higher proportion of tRF targets decreasing with age.

Conclusions

Our results indicate that 3' tRF levels increase with age and their abundance patterns differ from those of 5' tRFs. In agreement with our previous results in *D. melanogaster* and experimental findings, we observed different “seed” sequence locations for different tRFs and significant enrichment of their predicted targets for neuronal and developmental functions. We also found that tRFs can have different effects on their target levels

compared to miRNAs. Finally, as a side note, we illustrate the utility of tRF analysis for annotating tRNA genes in sequenced genomes.

Introduction

Small RNA molecules derived from fragmented tRNAs form a new class of short (~16-40 nt) RNA molecules. They arise from directed cleavage of cellular tRNAs, including both tRNA precursor species as well as mature, functional tRNA molecules and have been associated with multiple infectious diseases, pathogen resistance and regulation [74, 75]. Early reports described such fragments resulting from cleavage of tRNAs in *E. coli* as a protective response to phage infection and as “biochemical warfare” directed against unrelated bacterial strains [6, 76]. Subsequent studies have expanded the known domain of these fragments to eukaryotes, including human cells [3-5, 8, 19]. Broadly, the fragments are categorized into two types based on length and biogenesis: tRNA halves and tRNA-derived fragments (tRFs) and this paper is focused on the latter. Studied and reviewed by several experimental groups [37, 38, 77, 78], tRFs are molecules of ~16-24 nt in length and can be classified into three types based on the tRNA region from which they derive: 5' tRF, 3'CCA and 3'U tRF. The latter two types originate from the 3' end of the tRNA, while the former is derived from the 5' end. The 3'CCA type is generated from the 3' end of the mature tRNA and includes the CCA that is added to all tRNAs post-transcriptionally. The 3'U type is derived from the uracil rich trailer sequence upstream of the 3' end of the precursor tRNA molecule and has multiple Us added to the 3' end. There have been various attempts to determine the biogenesis and function of these different types of tRNA-derived small RNAs, but currently most of these questions are still open.

Hypothesized to function similarly to microRNAs (miRNAs), either by regulating mRNAs (like miRNAs) or by affecting miRNA loading and processing [3, 19, 22], tRFs have also been shown to bind to Argonaute complexes in multiple species [13, 47], strengthening their likely role in RISC-mediated gene silencing. A meta-analysis of PAR-CLIP libraries found that both 5' and 3' CCA tRFs were loaded to Ago1, Ago3, and Ago4, but 3' U tRFs did not associate with Argonaute proteins in great numbers in human cells [13]. A recent study suggested a traditional miRNA-like silencing based on complementarity of the 5' seed sequence of a tRF to a sub-sequence within a 3' UTR of a transcript [22]. Yet another study has shown that the last 8-10 nts on the 3' end of the tRF are responsible for mRNA repression [79]. In our lab, using a computational approach similar to detection of miRNA seeds, we have found potential seed regions on both a 5'-and a 3' tRF end [32]. Adding to this similarity, we have also reported age-related changes of tRF abundance in *Drosophila melanogaster* [32], comparable to those detected for miRNA in the same organism [47]. Such changes with age were also detected in tRFs of *Caenorhabditis elegans* [31].

Here we report on further support for such miRNA likeness of tRFs in another experimental system, which shows that both of these types of small RNA may participate in the mechanisms of brain aging. Aging underlies cognitive decline and dementia, and is the greatest risk factor for the failure of brain functioning in adults. Analysis of aging brain can shed light on the basic neurological mechanisms and their connections with age-

related neurodegenerative conditions such as Alzheimer's and Parkinson's disease [80]. Neurological research has used rats extensively over many years as models for mammalian behavioral and neurodegeneration studies. We analyzed available RNA sequencing libraries produced from the brains of rats of different ages [81] and identified numerous tRFs, which showed consistent changes in their abundance patterns with age. We also confirmed in rat brains our previous findings on possible targeting mode of *Drosophila* tRFs and on the functional enrichment of their targets in neuronal and developmental functions [32]. Potential targets of tRF with clearly defined seeds showed higher levels of down-regulation with age compared to the rest of the brain transcriptome and to the targets of miRNAs upregulated with age. Our results strengthen the emerging consensus that tRFs are a novel class of non-coding RNA molecules, they target mRNAs in a similar manner similar to miRNAs and their abundance in the cell is dynamically regulated with regards to the aging.

Results and Discussion

tRNA fragments in rat brain

We analyzed available datasets of nine different small-RNA libraries corresponding to three replicates for three distinct time points throughout a rat lifespan. These libraries were originally produced to study miRNA in the brains of young, middle-aged and old rats [81]. We will refer to the results associated with these three time points as Y, M and O, respectively. After mapping short RNA reads from these libraries to the union of tRNA sequences obtained from two independent databases (<http://gtrnadb.ucsc.edu>, <http://trna.bioinf.uni-leipzig.de>), we observed that the vast majority of alignments localized preferentially to a 5'- or a 3'-end of a tRNA molecule. Only 1-7% of the reads amongst the nine sequencing experiments aligned elsewhere on the tRNA sequence. In the datasets we analyzed, a negligible number of reads aligned to 3' U tRFs, therefore we limited our focus to 5' and 3' tRFs, for which there was extensive evidence. The two dominant tRF classes appeared likely to be generated by different mechanisms of cleavage. For instance, there was a striking consistency regarding the cleavage site location in 3' tRFs compared to a wider distribution of those sites in 5' tRFs (Fig. 2) supporting the notion that tRFs are not byproducts of random degradation but have specific structure-dependent cleavage sites.

Age-related patterns of tRF abundance

We then analyzed age-related abundance of 3' and 5' tRFs in the brain. Interestingly, we observed a very common trend of an overall monotonous increase in the 3' tRF levels with age, $Y < M < O$ (Fig. 3). In striking contrast, the 5' tRFs displayed a much less consistent picture (Fig. 4), with several cases of monotonous increase or decrease with age but mostly with a different pattern of change, $M < Y < O$, (Fig. 5). This difference, together with the cleavage site distributions (Fig. 2) suggests that distinct processes are likely responsible for the generation of 3' and 5' tRFs and that may be relevant for their function.

Computational prediction of tRF seeds and their targets

Given the significant levels of tRFs in rat brains and their dynamic changes with age, we aimed to investigate their possible effect on the brain transcriptome. Although the mechanism of tRF action is yet to be elucidated, there is recent evidence suggesting an animal miRNA-like pathway of action. Previous reports have detected tRNA fragments in the cytoplasmic fraction of various human cells, including B-cells and A549 cells [46, 79] as well as mouse ES cells, plant cells, fission yeast cells and carcinoma cell lines including HepG2, LNCap and LNCap-derived C4-2 [5, 17, 82-84]. It has been proposed that tRFs are likely to function similar to a traditional miRNA-like mode, using perfect complementarity of the 5' seed sequence of the tRF (typically, positions 2-8 in miRNAs) to target a subsequence within a 3' UTR of a transcript [46]. Contrary to the above, an alternative mode of action for tRFs has been suggested by a study [79], which utilized luciferase

reporter assays to demonstrate that a potentially seed sequence resided in the 3' end of the tRNA fragment, ruling out a 5' and a middle segment seed binding.

A search for a near-perfect complementarity of tRF sequences against transcripts yielded very few results, further suggesting a targeting mode similar to animal miRNAs. Assuming such animal-like miRNA targeting mechanism for tRFs, we further investigated the targeting mechanism of tRFs and adjusted our computational pipeline used to find targets in *Drosophila* genomes to perform the tRF seed search in mammalian genomes [32]. This pipeline functions similarly to the approach used to identify such seed sequences for miRNAs [45]. We used a 7-nt sliding windows across the length of a tRF sequence and aligned them against conserved 3' UTR regions of 23 vertebrate species. For the purpose of this study we took into consideration the following match types: 7-mer-m8 (full 7-mer match), 7-mer-1a (perfect match of the first 6 nts followed by an A in the 3' end of the targeted transcript) and 8mer-1a (perfect 7-mer match followed by an A in the 3' end) which have been extensively confirmed for miRNAs in the past [43]. Our results demonstrate that such seed regions can be located on the 5' end and on the 3' end of the tRF (Fig. 6), concordant with the existing experimentally validated results for tRF targeting mechanisms [46, 79]. A similar arrangement of the seed regions on the 5' end and on the 3' end of the tRF has also been observed in *Drosophila* [32].

The success in finding seeds was overwhelmingly in favor of shorter 3' tRFs (6 out of 24, 25%) vs longer 5' tRFs (only 2 were found out of 30, 6.67%, not shown). Given the number of differences between these two types, we chose to focus on 3' tRFs in the remainder of this paper. In our meta-analysis in the sections below we combined the experimental results that were performed in different labs, with different brain material and at different ages (e.g., 22 months in small RNA-seq series is almost half way between 14 and 28 months in RNA-seq series). Given the small changes in gene expression, and to avoid the effects of non-monotonous changes in many 5'tRFs, we limited our subsequent analysis to six 3' RNAs (Fig. 6), which showed monotonous changes in their levels from Y to O and clearly defined seed sequences.

Gene ontology enrichment analysis of conserved predicted targets

Following our seed region identification for tRFs, we focused on their predicted targets with conserved seed matches within their 3' UTR (Supplementary file 1). We explored potential functions of targets of six tRFs that showed clearly defined seed sequences (Fig. 6). Gene Ontology (GO) enrichment analysis of conserved predicted targets of these tRFs revealed >150 significantly enriched GO terms for biological process. "Nervous system development" was found to be consistently enriched for all six tRFs except ProTGG. Additional biological process GO terms such as "central nervous system development", "neurogenesis" and "axonogenesis" were also enriched for multiple tRF targets (Supplementary file 2). Further, the same tRF targets that showed an enrichment for

nervous system functionality and development were also associated with significantly enriched neuron/axon-related cellular localization terms (Supplementary file 2). Overall, these results are in agreement with our previous work on *D. melanogaster* (Karaikos et al. 2015), where we have noted a similar enrichment for biological processes related to neuronal function and development for predicted targets of tRFs increasing with age from young to adult flies. However, in addition to these functions, ProTGG and other tRFs also appeared to target transcription and splicing regulators in rat brains (Supplementary file 1, 2).

Expression patterns of predicted tRF targets

We compared our observations of tRF abundance changes with age to the measured expression levels of their targets. We compared the profiles of all mRNAs in the rat cerebral cortex transcriptome [85] with those predicted to be targeted by miRNAs (using Targetscan [45]) and by tRFs (using perfect matching of the identified tRF seed sequence and a conserved target sequence located in the 3' UTR of a transcript). We calculated the ratios of down- to up-regulated transcripts for the whole rat cortex transcriptome and for the targets of six 3' tRFs (in which seeds could be clearly seen, Fig. 6) and five miRNAs that, similar to 3' tRFs, showed a monotonous increase with age. We observed that both tRF and miRNA targets were significantly enriched for down-regulated transcripts at three different regulation thresholds (Table 2). Interestingly, the enrichment for down-regulation in the union set of all neuron-related tRF targets was also significant (at p-value

< 0.05) for each of these three thresholds of regulation. An illustrative subset of the genes that were attributed to “nervous system development” GO term and showed a monotonous decrease with age are shown in Figure 7.

Comparing the distributions of de-regulation levels from young to old age for (i) all mRNAs detected in rat cortex, (ii) for miRNA-targeted mRNAs and (iii) for tRF targets (Figure 8), we observed consistently higher proportion of down-regulated and lower proportion of up-regulated targets in both miRNA and tRF groups of targets compared to all mRNAs. Although these proportions for mRNAs and tRF targets were generally comparable, we noted a bimodal distribution for tRF targets, whereas such bimodality was much less pronounced for miRNA targets (Figure 8). Targets for both types of small RNAs show their most prominent peaks for low levels of down-regulation with age (these range from 0 to -5% and are possibly related to targeting relevant in other cellular contexts or false positives in target predictions). However, the proportion of tRF targets down-regulated in the range of 10-22.5%, and thus more likely to be relevant in the brain, is consistently higher compared to that of miRNA targets. Such (relatively low) level of change is not surprising, given that miRNA are considered to be fine-tuning the transcriptional control by post-transcriptionally modulating the target transcript levels [86]. The age-related decrease in the mRNA levels for tRF targets is generally more pronounced than that for miRNA targets.

Finding missed tRNA genes

In our effort to identify every possible tRF present in rat brains, we took into account a union of all annotated rat tRNAs from two databases (<http://gtrnadb.ucsc.edu> and <http://trnadb.bioinf.uni-leipzig.de>). Although the latter database is rather small compared to the former, we found that it contained a handful of rat tRNA genes (to which tRF fragments did map perfectly) that were missing from the UCSC database at the time of our first analysis. Upon subsequent checking, we found that most of the missing tRNA genes have been added correctly to the most recent update of the UCSC database (not including mitochondrial tRNAs). However, there is a tRNA gene (tdbD00000658-GluCTC), which aligns perfectly to the rat genome (chr17:45,642,771-45,642,843 of rn6), and which is still absent in the latest version of the UCSC database. In our analysis we detected tRFs from all nine libraries mapping to tdbD00000658-GluCTC sequence. Together with the fact that annotating tRNAs is not a typical priority in genome sequencing projects, our observations suggest that there are potentially other tRNA genes lacking annotation in the published genomes. However, such genes appear to be sources of detectable tRFs. Hence, analysis of tRFs can have an added value of revealing unannotated tRNA genes for multiple species.

Discussion

In this study we characterized tRFs present in rat brains at three different time points, revealing that their abundance is dynamically regulated in the context of age. Previously, we have reported age-related changes in *D. melanogaster* tRFs [32]. While only two time points have been considered in that paper, it has shown the changes related to the tRF loading to Argonaute proteins and thus very likely related to the function of the RISC complex. Here, we observed two typical patterns of change in tRF levels. One was a monotonous increase with age, primarily seen in 3' tRFs. Another was a lower abundance in mid-aged rat brains and higher abundance in young and old animals, mostly observed in 5' tRFs. These patterns, together with the differences in fragment sizes suggest distinct mechanisms of cleavage for the two types of fragments, which can potentially be attributed to the different roles for these two types of tRFs. In addition to the biogenesis pathways, tRFs originating from different ends of the tRNA molecule have also been shown to localize in different sub-cellular compartments. As pointed out by Kumar et al [13], 5' tRFs were equally abundant in the nuclei and whole cell fraction of HeLa cell line [87] indicating primarily nuclear localization and consistent with large numbers of 5' tRFs in HeLa cell nucleoli [3]. On the contrary, 3' tRFs showed an enrichment in the whole cell fraction indicating their cytoplasmic localization in agreement with Haussecker et al [4]. There has been evidence of miRNAs actively loaded to Argonaute proteins in an age-dependent manner in *D. melanogaster* [47]. A very similar age-related loading pattern was also observed for *D. melanogaster* tRFs [32]. This, along with extensive evidence that Argonaute proteins are not only acting in post-transcriptional silencing but are

localized/imported to the nucleus, could imply additional unknown functions for tRFs within the nuclear compartments of the cell. Perhaps, such functions are similar to those previously described for miRNAs, which have been shown to be associated with mRNA splicing and modulation of histone epigenetic modifications [88, 89], and this is a focus of our ongoing research.

Although the mode of action for tRFs is yet to be elucidated, our results support the hypothesis that mammalian tRFs (at least, 3' tRFs) can act in a very similar way to miRNAs in post-transcriptional gene silencing. We show here that they contain 7mers, which match 3' UTR regions of transcripts at much higher rate than expected by chance, similar to the seed sequences of miRNAs. Searching for conserved matches across vertebrate genomes, we found such seeds on either end of the tRF molecules, as has been the case with 12 *Drosophila* species [32]. Previous studies have also detected both 5' and 3' seeds in different tRFs and changes in the seed sequence have been shown to affect the suppression of mRNA translation [46, 79]. It is worth noting that in miRNAs, 3'-compensatory sites [52] and central pairing sites [53] have been reported in addition to the most prevalent 5' seeds [42-45], thus finding seeds on both ends of tRFs is not unexpected. Non-traditional seed region location in miRNA is also consistent with the extensive results of Helwak et al [90], who reported that more than half of the observed miRNA-mRNA interactions do not fulfill traditional seed binding properties in HEK-293

cells. However, one cannot exclude other modes of action, for example, tRFs have been reported binding to oncogenic RNA-binding protein YBX1, displacing pro-oncogenic transcripts and acting as tumor suppressors [61].

Interestingly, for tRFs with clearly defined seed-like regions, we observed a significant and consistent enrichment for targeted genes related to neuronal function and development in Gene Ontology terms. Again, this was in agreement with a functional enrichment seen in *Drosophila* tRF targets [32]. However, in addition to these functions, rat brain tRFs also appeared to target transcription and splicing regulators, in parallel to earlier findings for rat brain miRNAs [86]. Among the tRF targets with a well-defined role in the nervous system, a netrin receptor UNC5C is related to axon guidance and neural development. A mutation in this gene has been associated with predisposition to Alzheimer's disease and has been shown to cause increased neuronal cell death in rodents [91]. Cadherin genes, which are related to development and maintenance of functional structures in the central nervous system (reviewed in [92]) were found to be targeted by tRFs (PCDH9). Fibroblast growth factor receptor-2 gene (FGFR2) was also found among the targets, suggesting that tRFs may affect key proteins involved in neural development, given that fibroblast growth factors are potent modulators of proliferation in the developing nervous system [93].

Having identified potential targets of upregulated 3' tRF, we compared age-related changes in their transcript levels with the targets of upregulated miRNAs and observed

small but significant down-regulation of such targets for both groups of ncRNAs. However, tRFs appeared to have more of their targets down-regulated to a greater extent compared to those of miRNAs. The RISC pathway functions by repressing translation and by mRNA cleavage and the exact balance of those mechanisms is not known. It has been speculated that degradation of repressed mRNAs by other mechanisms may be responsible for the observed decrease in their counts [94]. It is also unclear if the miRNA and tRF levels determined by RNA-seq correlate with their actual functional levels in the RISC complexes. Nevertheless, tRF targets appears to be more efficiently down-regulated compared to miRNA targets in aging rat brains. These findings await experimental validation and may be of relevance for human aging and neurodegeneration studies, given the comparable gross structure of the rat and human brains and the role of rat models in neurological research.

Methods

Small RNA analysis

We used publicly available datasets of small RNA from rat brain [81] with accession number ERA36511. Using the sra-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) we converted the files to fastq format using fastq-dump and removed the 3' adapter sequences with fastx-clipper. The reads of length above 16 nts were used for downstream analysis. We collapsed and mapped the reads to the rat genome (rn6, UCSC) and the union of rat tRNAs from two independent databases (<http://gtrnadb.ucsc.edu> and <http://trnadb.bioinf.uni-leipzig.de> also including mitochondrial tRNA genes from the second one) using Bowtie. Bowtie parameters were set to output only perfect matches to tRNA sequences (including the post transcriptional CCA modification). Read counts in each experiment were normalized by the total number of reads detected and averaged across three replicates for each of the three time points (ages of 6, 14 and 22 months).

Seed sequence analysis

We generated 7-mer subsequences of tRFs by applying a 7-nt sliding window and shifting by one nt from the 5' to the 3' end. We then found the counts of exact matches for each of these subsequences to the 3' UTR regions conserved in at least 15 species, including human mouse and rat (<http://www.targetscan.org/>). To estimate significance of the seed

matches we compared the observed match counts for each respective 7-mer in a tRF to (i) those expected number of matches by chance (estimated from 7-mer genomic frequency) and to (ii) average numbers of matches of all possible 7-mers with the same nucleotide composition in conserved 3'UTRs. Genes with exact matches of 7-mer and 7-mer_1a candidate seeds to the 3'UTR were considered potential targets.

RNA sequencing analysis

For target expression analysis we downloaded files with pre-computed transcript expression levels for the rat cerebral cortex transcriptome, data series with accession number GSE34272 [85]. These expression levels in each experiment were normalized by the total number of reads detected and averaged across three replicates for each of the three time points (ages of 6, 12 and 28 months).

Statistical evaluation of downregulation levels for miRNA and tRF targets

For each set of predicted targets of a tRF or a miRNA, we compared its ratio of down-regulated/up-regulated target transcripts from young to old rats with the distribution of such ratios calculated for 1,000 randomly selected transcript sets of the same size as the target set (different for each tRF and miRNA). This process was repeated three times for three different thresholds (up-regulated by >5% / downregulated by >5%, up-regulated

by >10% / downregulated by >10% and up-regulated by >20% / downregulated by >20%) using R statistical package (www.R-project.org) to find statistical significance of the difference observed.

Gene ontology enrichment analysis

The predicted targets for each tRF were used as input in order to perform GO enrichment analysis. Each set of targets was uploaded to PANTHER website (<http://pantherdb.org/>, [95]) and results were obtained using the recommended default parameters.

Author Contributions: SK participated in the design of the study, analyzed the data, and drafted the manuscript. AG conceived the study, overseeing in its design, execution and coordination, drafted and finalized the manuscript. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Grant Information: This work was in part supported by the National Science Foundation (award DBI-1458202 to A.G.).

Acknowledgments: We would like to thank Merve Ozbas for excellent technical help and Sean Smith and Joseph Kawash for critical reading of an earlier version of the manuscript.

Figure and table legends

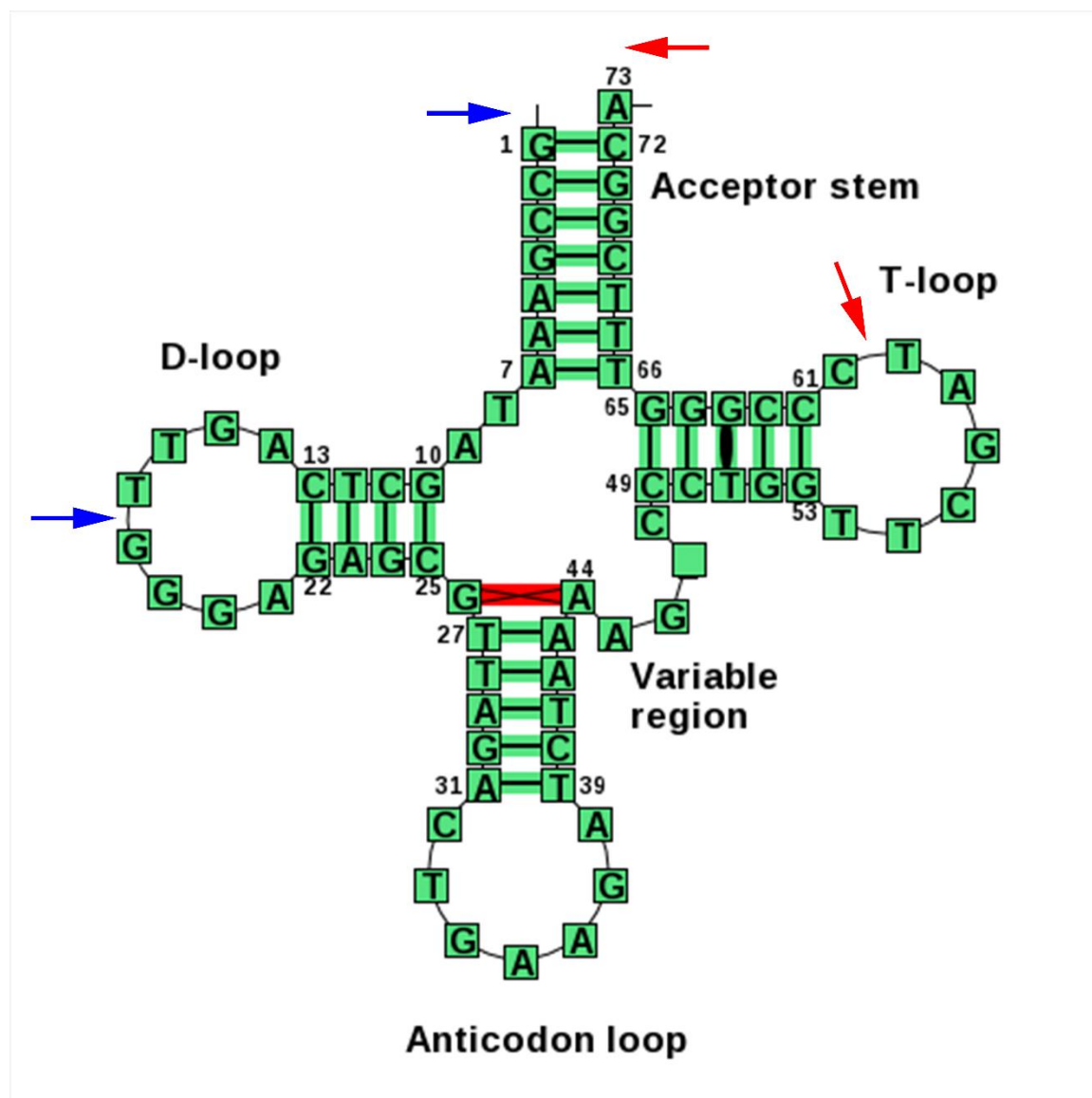


Figure 1 title. tRNA secondary structure representation

Figure 1 legend. Typical secondary structure representation for the PheGAA tRNA gene (from <http://trnadb.bioinf.uni-leipzig.de>). Blue arrows point to typical endpoints for a 5' tRF. Red arrows indicate the ends of the most frequent 3' tRF. The mature tRNA molecule also contains the post-transcriptional 3' CCA modification (as does the 3' tRF). A 3' U tRF

would derive from the uracil-rich trailer sequence downstream of the end of the tRNA gene (not shown).

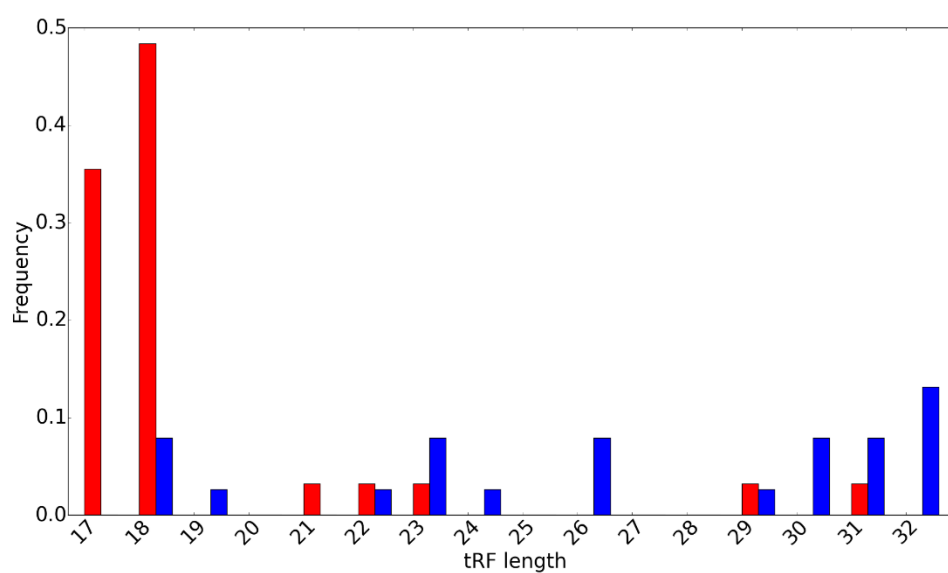


Figure 2 title. Length distribution of tRFs

Figure 2 legend. Length distributions for 5' (blue) and 3' (red) tRFs. tRF length is shown on the x-axis and the frequency on the y-axis.

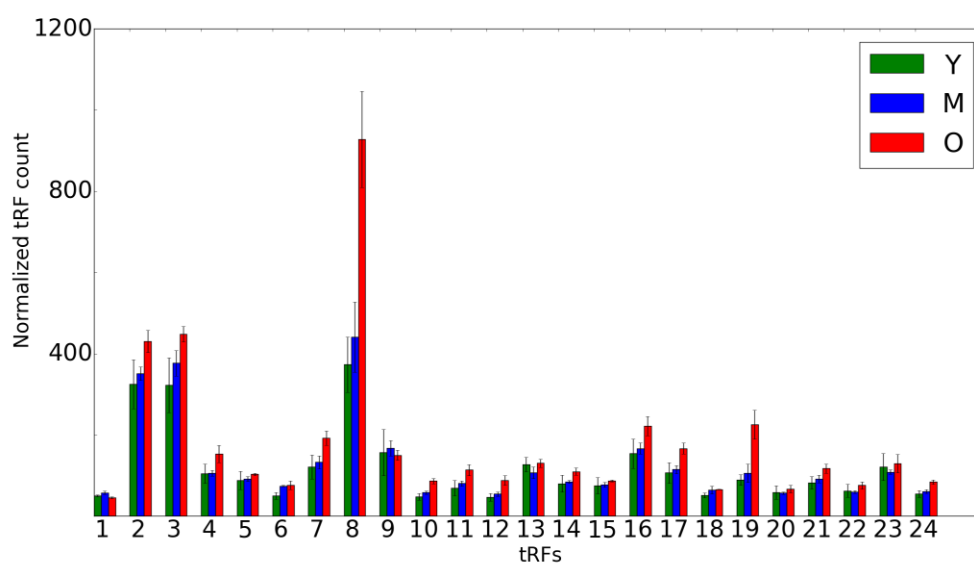


Figure 3 title. Age-related change in 3' tRF levels

Figure 3 legend. Abundance of 3' tRFs in rat brains for 3 distinct time points (Y is shown in green, M in blue and O in red). An average of 3 replicates for each tRF is shown on the x-axis for each time point, error bars indicate the range of read counts. Shown on the y-axis is the normalized tRF abundance. The numbers on the x axis correspond to tRNA genes listed in Table 1.

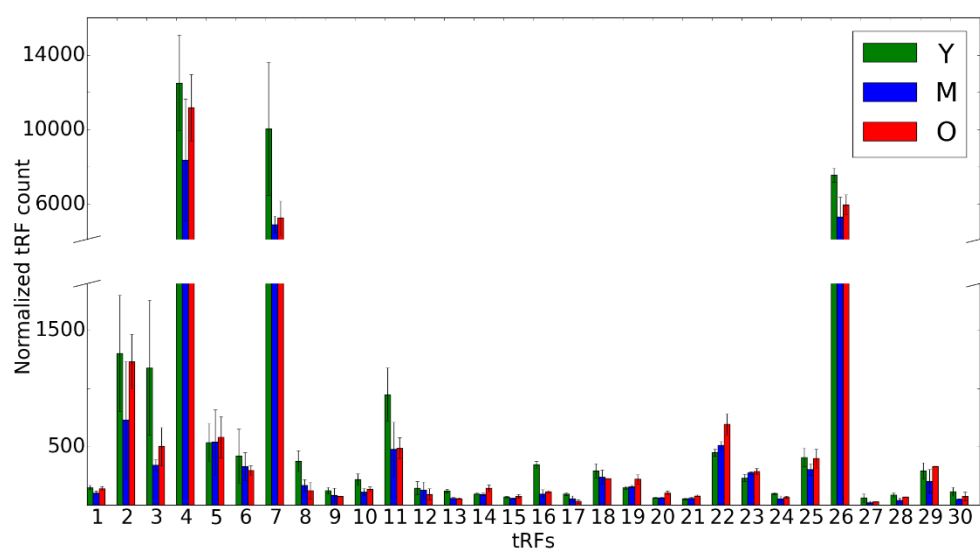


Figure 4 title. Age-related change in 3' tRF levels

Figure 4 legend. Abundance of 5' tRFs in rat brains for 3 distinct time points (Y is shown in green, M in blue and O in red). An average of 3 replicates for each tRF is shown on the x-axis for each time point, error bars indicate the range of read counts. Shown on the y-axis is the normalized tRF abundance. The numbers on the x axis correspond to tRNA genes listed in Table 1.

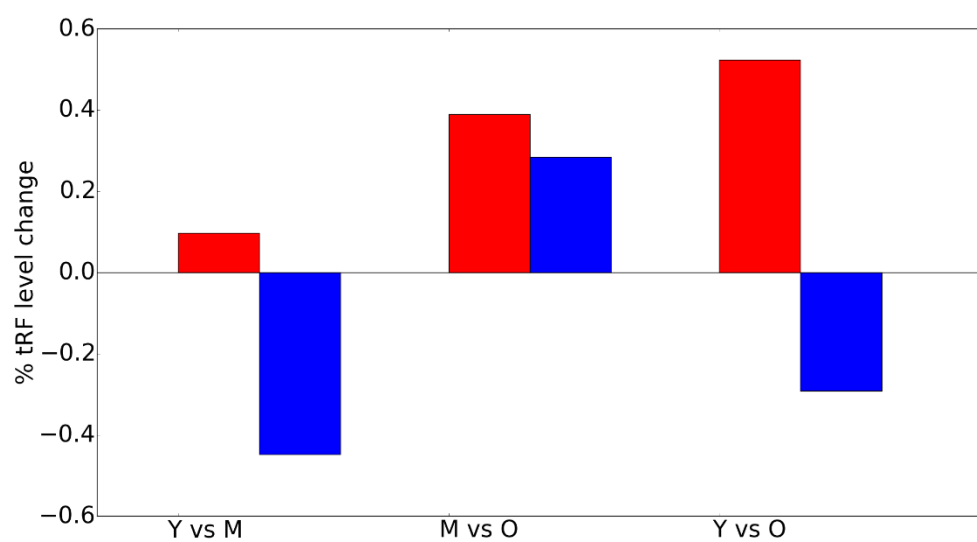


Figure 5 title. Total tRF abundance

Figure 5 legend. Change in total abundance levels with age for all 5' tRFs (blue) and 3' tRFs (red).

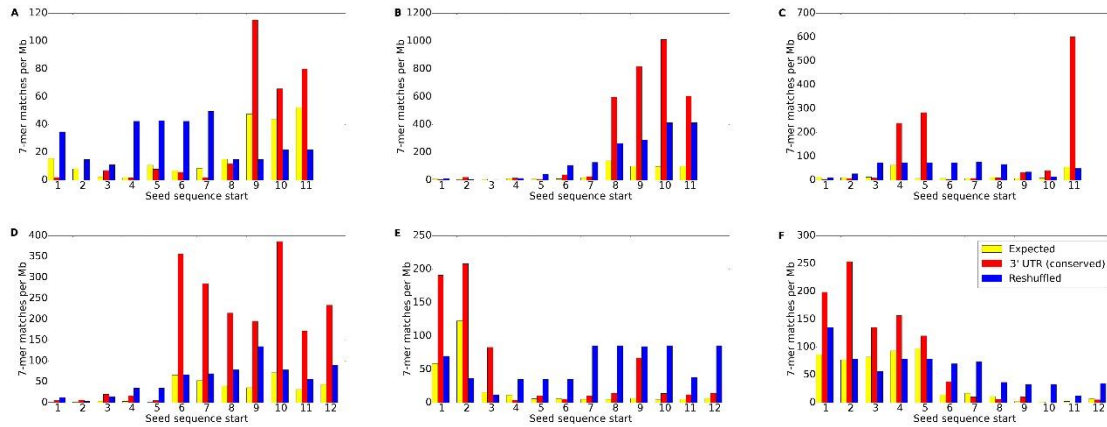


Figure 6 title. Candidate seed region locations for tRFs

Figure 6 legend. Candidate seed regions for tRFs. The numbers of exact sequence matches in the 3' UTR regions are plotted against the starting position of a 7-mer. Expected number of matches in 3' UTRs is shown in yellow, average number of conserved matches for all other 7-mers with the same nucleotide composition as the given window is shown in blue and the observed number of matches in the conserved regions of 23 vertebrates is shown in red. The letters on the top left corners of each plot correspond to individual tRFs: A) ProTGG (3' tRF), B) ValAAC (3' tRF), C) PheGAA (3' tRF), D) AlaTGC (3' tRF), E) SerAGA (3' tRF), F) SerGCT (3' tRF).

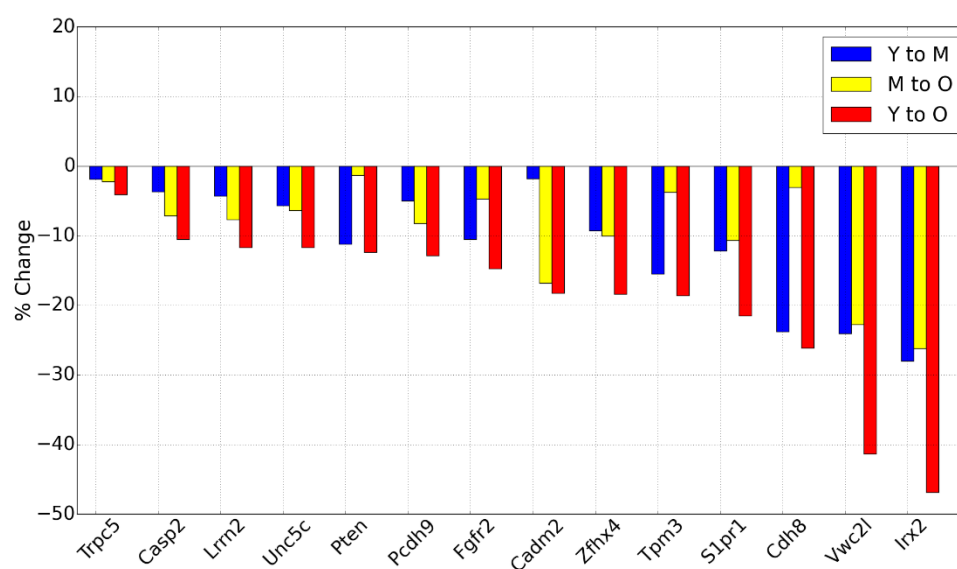


Figure 7 title. Down-regulation patterns of tRF targets

Figure 7 legend. Down-regulation patterns of tRF targeted transcripts with age. All the gene names found on the x-axis are associated with the GO term “nervous system development”, which was observed to be consistently enriched for five out of six tRFs shown in Figure 6. Down-regulation levels are shown on the y-axis.

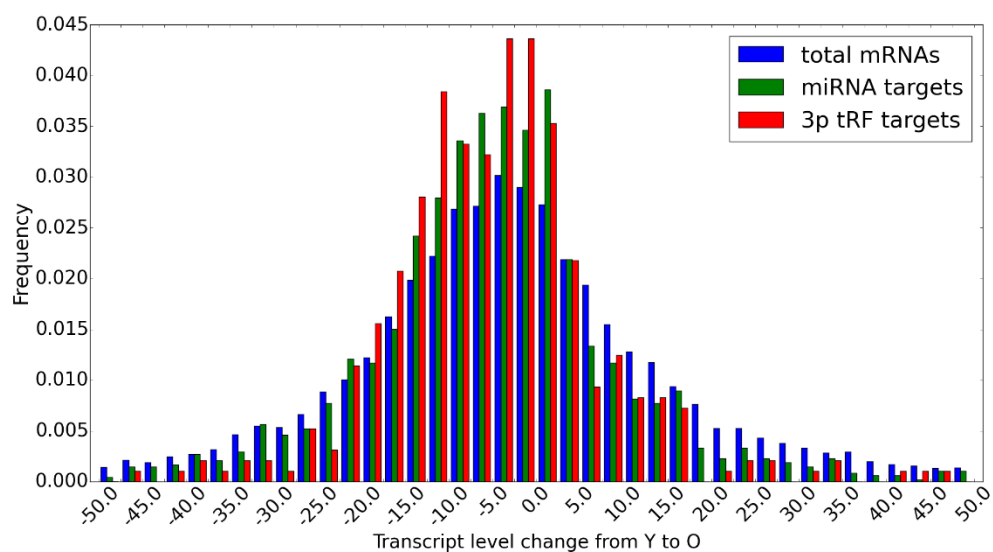


Figure 8 title. Changes in target transcript levels with age

Figure 8 legend. Transcript level changes from young to old rat brains. Distributions of changes for all detectable mRNAs in rat brains (blue), miRNA-targeted (green) and 3' tRF-targeted transcripts (red) are shown using 2.5% bins .

Table 1

| tRNA Anticodon | tRNA ID | Fig. 3 (3' tRFs) | Fig. 4 (5' tRFs) | Database | Coord. |
|----------------|----------------|------------------|------------------|--------------------|--------|
| TrpTCA | MtddbD00003370 | 1 | -- | ** (mitochondrial) | 49-69 |
| ValCAC | trna2253 | 2 | 25 | * | 59-76, |
| ValAAC | trna1605 | 3 | -- | * | 60-76 |
| TrpCCA | trna11440 | 4 | -- | * | 58-75 |
| LeuAAG | trna6448 | 5 | -- | * | 68-85 |
| LeuCAG | trna3830 | 6 | -- | * | 69-86 |
| SerGCT | trna9151 | 7 | -- | * | 68-85 |
| ProTGG | trna13310 | 8 | -- | * | 59-75 |
| SerAGA | trna455 | 9 | 20 | * | 68-85 |
| SerTGA | trna2516 | 10 | -- | * | 68-85 |
| GlnCTG | trna1868 | 11 | -- | * | 58-75 |
| GlnCTG | trna1911 | 12 | -- | * | 58-75 |
| ValTAC | MtddbD00003196 | 13 | -- | ** (mitochondrial) | 49-71 |
| LysCTT | trna186 | 14 | -- | * | 60-76 |
| PheGAA | trna3690 | 15 | -- | * | 60-76 |
| ArgACG | trna1588 | 16 | -- | * | 59-76 |
| TrpCCA | trna2467 | 17 | -- | * | 59-75 |
| AlaTGC | trna5057 | 18 | 14 | * | 58-75, |
| ProAGG | trna13311 | 19 | -- | * | 58-75 |
| GlyGCC | trna1527 | 20 | 5 | * | 58-74, |
| ArgACG | trna1600 | 21 | -- | * | 60-76 |
| GlyTCC | trna2250 | 22 | -- | * | 59-75 |
| GlyCCC | trna7585 | 23 | -- | * | 57-74 |
| GlyCCC | trna2752 | 24 | -- | * | 57-74 |
| IniCAT | MtddbD00003481 | -- | 1 | ** (mitochondrial) | 1-30 |
| GlyGCC | trna2377 | -- | 2 | * | 1-18 |
| GlyGCC | trna2376 | -- | 3 | * | 1-32 |
| GlyGCC | trna3490 | -- | 4 | * | 1-18 |
| GlyGCC | trna1897 | -- | 6 | * | 1-31 |
| GlyGCC | trna1528 | -- | 7 | * | 1-32 |
| GlyGCC | trna11254 | -- | 8 | * | 1-34 |
| LysCTT | trna7564 | -- | 9 | * | 1-26 |
| LysCTT | trna2975 | -- | 10 | * | 1-33 |
| LysCTT | trna13312 | -- | 11 | * | 1-34 |
| HisGTG | trna3497 | -- | 12 | * | 1-30 |
| HisGTG | trna4377 | -- | 13 | * | 1-33 |
| AlaTGC | trna12492 | -- | 15 | * | 1-23 |
| CysGCA | trna8932 | -- | 16 | * | 1-32 |
| CysGCA | trna8966 | -- | 17 | * | 1-33 |
| TyrGTA | MtddbD00003620 | -- | 18 | ** (mitochondrial) | 1-31 |
| SerAGA | trna3680 | -- | 19 | * | 1-23 |
| LeuCAG | trna2365 | -- | 21 | * | 1-18 |
| LeuCAG | trna2370 | -- | 22 | * | 1-26 |
| ValCAC | trna1607 | -- | 23 | * | 1-23 |
| ValAAC | trna3670 | -- | 24 | * | 1-33 |

| | | | | | |
|--------|--------------|----|----|----|------|
| ValCAC | trna1601 | -- | 26 | * | 1-33 |
| ValCAC | trna4363 | -- | 27 | * | 1-36 |
| GlnTTG | trna4 | -- | 28 | * | 1-36 |
| GluCTC | tdbD00000658 | -- | 29 | ** | 1-19 |
| GluCTC | trna2251 | -- | 30 | * | 1-34 |

Table 1 title. tRNA genes and corresponding numbers shown in Figures 3 and 4.

Table 1 legend. The tRNA anticodon is shown in column 1, the tRNA ID in column 2 and the corresponding numbers used as an x axis label for Figure 3 and Figure 4 (when applicable) are shown in columns 3 and 4, respectively. A database, which each tRNA gene was downloaded from, is shown in column 5 with * corresponding to <http://gtrnadb.ucsc.edu> and ** corresponding to <http://trnadb.bioinf.uni-leipzig.de>. Last, coordinates of tRFs on the corresponding tRNA genes are given in column 6.

Table 2

| tRF | Ratio (>5%) | Ratio (>10%) | Ratio (>20%) |
|-----------|-------------|--------------|--------------|
| ProTGG | 11.0 *** | 7.0 *** | 4.0 ** |
| AlaTGC | 2.15 * | 2.57 ** | 0.5 * |
| PheGAA | 3.33 *** | 3.7 *** | 3.25 *** |
| SerAGA | 4.0 *** | 4.0 *** | 3.0 * |
| SerGCT | 4.0 *** | 3.5 *** | 2 * |
| ValAAC | 4.1 *** | 3.5 *** | 1.5 - |
| miRNA | | | |
| mir-146b | 2.857 *** | 2.461 ** | 4.25 *** |
| mir-132 | 2.31 *** | 2.241 ** | 1.5625 * |
| mir-128-2 | 2.52 *** | 2.3 *** | 2.5 *** |
| mir-92a | 2.63 *** | 2.438 *** | 1.89 ** |
| mir-200c | 2.714 *** | 2.96 *** | 1.967 *** |

Table 2 title. Age-related down-regulation of tRF/miRNA targets

Table 2 legend. Ratios of down-regulated / up-regulated miRNA- and tRF-targeted transcripts for each of the change thresholds (>5%, >10%, >20%). Significant difference from the expected ratio is indicated by *** (p-value < 0.005), ** (p-value < 0.01) or * (p-value < 0.05); “-” indicates p-value > 0.05.

Supplementary file 1. List of predicted conserved tRF targets for the tRFs shown in Figure

6

Supplementary file 2. Gene Ontology enrichment analysis tables for the tRFs shown in

Figure 6

CHAPTER 3: ANALYSIS OF THE TARGETING MODES OF AGO1 LOADED tRFs

This work has been submitted to Biology Direct as follows:

Analysis of the targeting modes of Ago1 loaded tRFs

Spyros Karaiskos^{1#}, Lingyu Guan^{1#} and Andrey Grigoriev ^{1*}

¹ Department of Biology, Center for Computational and Integrative Biology, Rutgers

University, Camden, New Jersey 08102, USA

* Correspondence: Andrey Grigoriev, Department of Biology, Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey 08102, USA

andrey.grigoriev@rutgers.edu

These authors provided equal contribution to this work.

Keywords:

transfer RNA, tRNA fragments, tRFs, Ago, siRNA, non-coding RNA

Contribution

My contribution to this study involves developing the pipeline to analyze the CLASH sequencing data and all downstream analyses. I helped generate all figures and I helped write the manuscript.

Abstract:

Transfer RNA fragments (tRFs) are a class of small RNA molecules derived from mature or precursor tRNAs. Although characterized very recently, tRFs have been gradually attracting more attention. They are found across a wide range of organisms and tissues in cytoplasmic compartments or loaded to RISC complexes, often in numbers comparable to microRNAs.

We analyzed sequences of chimeras formed in vivo between Argonaute-loaded tRFs and their targets, corresponding to various gene types, in addition to protein-coding transcripts. In the latter, 3' UTRs were the likely primary target regions, although we observed interactions of tRFs with coding sequences and 5' UTRs. We also report a novel phenomenon – a large number of putative interactions between tRFs and introns, compatible with the role of Argonaute in the nucleus.

We clustered tRF binding patterns and identified enriched motifs that may be responsible for tRF-target interactions. Such interaction sites appear to be primarily located on the 5' end of a tRF, often involving additional binding of the 3' nucleotides of guide tRFs, similar

to microRNAs. Strikingly, our results match interaction sites detected in a recent experimental screen, confirming the validity of our approach to predict the sites and mechanisms of tRF/target interactions computationally.

Introduction

Recent advances in RNA sequencing technologies have contributed substantially towards the discovery of novel non-coding RNAs. Based on their size, non-coding RNAs can be grouped in three categories: long (>200 nts), medium (>40 nts) and small RNAs (<40 nts). In this study we focused on a particular type of small RNAs that originate from transfer RNA (tRNA) genes and are called tRNA-derived fragments (tRFs).

Mature tRNAs are usually less than 90 nts long and their secondary structure resembles a cloverleaf. Properly folded tRNAs contain four distinct arms: D arm, anticodon loop, T arm and a variable loop. Transfer RNAs are crucial components of the cell's translational machinery. They facilitate translation of mRNA codons into amino-acids through base pairing between the mRNA codon and the anti-codon tri-nucleotide located in the middle of the tRNA molecule. Despite the fact that there are only 64 codons encoding for 20 amino acids, the number of tRNA genes ranges in hundreds for multiple species. For example, the human genome contains more than 600 tRNA genes, ranking them among the most abundant RNA molecules in the human transcriptome [1, 2]. Such abundance of tRNA genes suggests that these molecules may have additional functions and properties. Hence, it is not surprising that recently there has been an explosion of reports describing abundant levels and potential novel functions of their fragments (tRFs) in cells of different species.

tRFs have been posited to arise from directed cleavage of cellular tRNAs, including both tRNA precursors and mature tRNA molecules. They are categorized into two groups based

on the length of the small RNA: tRNA halves (28~40 nts) and tRNA-derived fragments (16~24 nts). tRNA halves are considered to be a product of cleavage of mature tRNAs under stress conditions [9, 11, 96].

On the other hand, tRFs can be classified into four distinct subgroups based on their location: tRF-5, tRF-i, tRF-3 and tRF-1. tRF-5s are derived from the 5' end of the mature tRNA molecule through endonucleolytic cleavage near the D loop/arm [97]. tRF-i (internal tRFs) is the most recently identified type of tRFs, which span a variety of contiguous regions across tRNA molecules other than the very 5' and 3' ends [15, 16]. The last two subgroups originate from the 3' end of the transfer RNA molecule. 3' CCA tRFs (tRF-3) contain the post-transcriptional CCA trinucleotide addition and they are products of direct cleavage of the mature tRNA molecule (most frequently at the T arm/loop) while tRF-1 (also known as 3' U tRFs) derive from the uracil-rich sequence on the 3' end of the precursor tRNA molecule [13] .

In the past, tRFs have been excluded from small-RNA studies because they were considered to be non-functional degradation products of their parental molecules. However, there is both biochemical and computational evidence for the role of tRFs as functional molecules in multiple biological processes [32, 33, 98, 99]. tRFs have been shown to bind to Argonaute complexes in multiple species [13, 32] and they have been proposed to function similarly to microRNAs by regulating mRNAs or by affecting miRNA loading and processing [3, 19, 23]. In agreement with these similarities between miRNAs

and tRFs, a recent study showed that two miRNAs are actually tRFs deriving from the trailer sequence of tRNA genes [14].

The mode of action for tRFs with regards to RISC mediated post transcriptional RNA silencing still remains vague. It is unclear if tRFs act in a way similar to plant miRNAs (which are almost fully complementary to their target RNAs) or to animal miRNAs (which recognize their targets based on complementarity of a short “seed” region located on the 5' end of the small RNA molecule). Different results have been reported for such seed regions in tRFs. One study has demonstrated a silencing mechanism similar to miRNAs based on complementarity of the 5' seed sequence of a tRF to a 3' UTR of a reporter gene [46]. Another study has shown that such a seed region can be on the 3' end of a tRF molecule and can induce mRNA repression [79]. In our previous work in fruit fly and rat we have found that a seed region can be located on either end of a tRF molecule based on matches with conserved target sequences primarily found in 3' UTRs of mRNAs [32, 33].

The seed-driven target identification is a standard practice for miRNAs and it has been simply adopted for tRFs with some supporting computational and experimental data [13, 32, 46]. On the other hand, there is also an increasing amount of evidence for non-canonical hybridization modes for both miRNAs and tRFs [46, 79, 100]. Additionally, we have shown that seed sequences with exact match to conserved 3' UTRs are much more frequent in tRF-3 compared to tRF-5 in rat brain suggesting possible differences in binding

for different tRF types [33]. Thus in order to effectively study tRFs, there is a need to identify and understand their targeting/hybridizing modes.

In this study we investigated the targeting modes of tRFs and their post-transcriptional regulation capabilities with extensive computational meta-analyses and integration of the data produced by several independent experimental projects. In our ab initio analysis of tRF targeting modes we used the Crosslinking, Ligation, and Sequencing of Hybrids (CLASH) data series from Ago1 pulldowns in HEK293 cells [90]. We found that Ago1-loaded tRFs target a wide range of transcripts including coding and non-coding RNAs and that the target spectrum is dependent on the tRF type. We also report a novel phenomenon – a large number of putative interactions between tRFs and intronic sequences, consistent with the evidence of Ago function in the nucleus [89, 101]. Additionally, we also found that tRFs may be operating as guide molecules enabling Ago interactions with a specific group of short introns, recently identified as agotrons [102].

We analyzed sequences of chimeras formed in vivo between tRFs and their targets to identify clusters of RNA-RNA interaction signatures. We cataloged possible binding patterns between different types of tRF guides and targeted sequences and identified motifs that may be responsible for these interactions. We observed binding sites located on either end of a tRF molecule in agreement with experimentally validated results for tRFs from earlier studies [32, 33, 46, 79].

Finally, we compared our computational predictions of target interaction sites with those found in a recent experimental screen. Strikingly, for three common tRFs the predictions

matched the seed location determined in luciferase assays [20], demonstrating the predictive power of our approach. This opens the possibility of inferring the seed regions and mechanisms of tRF/target interactions computationally.

Results

1. Ago-1 loaded tRNA fragments

To investigate tRF/target RNA interactions, we analyzed a series of CLASH (Cross Linking and Sequencing of Hybrids) libraries, originally used to study miRNAs and their interactome in HEK293 cells [90]. Similar to miRNAs and their targets, CLASH captures exact tRF/target RNA interactions in vivo, and this allowed us to identify high confidence interactions. This dataset has been previously used [13] to identify a subset tRF and their targets but we found a large proportion of them missed before. Further, the sequences of tRF/target pairs allowed us to identify binding patterns and potential interaction sites.

First, we identified hybrid reads starting with a tRF sequence and then we used blast [103] in order to find the best match for the remainder of each hybrid read. Hybrid reads that passed all the quality control filters (see Materials and Methods) were considered tRF/target RNA chimeras and they were used for downstream analyses. We first examined whether tRF-5, tRF-i, tRF-3 and tRF-1 are loaded to Ago proteins. Our results show primary loading of tRF-3 to Ago1 followed by tRFs that derived from mature tRNAs (tRF-5 and tRF-i). The least frequent chimeras were formed between tRF-1 and target RNAs, in agreement with previous results [13] , (Fig. 1A).

We sought to determine whether tRFs interact with their targets in a directed manner. To test this hypothesis, we considered the degree of randomness of tRFs interacting with their Ago-loaded targets. We used RNAhybrid [104] to calculate the minimum free energy

(MFE) of hybridization for each tRF/target RNA chimera identified from CLASH data. Our results show that the observed binding of tRFs to their putative targets is significantly stronger than binding of randomly generated RNA-RNA chimeras (p-value < 10^{-16} compared to random and shuffled chimeras, Fig. 1, B).

Next, we examined whether tRFs are generated by specific cleavage sites across tRNA genes. We retrieved a large number guide tRFs (extracted from hybrid reads) mapping to the 5' end, the 3' end as well as the internal region of mature tRNAs. tRF-3 show the most prominent peak for fragments of length 18 nts whereas tRF-5 overall seem to be slightly longer fragments with the most prominent peak at 21 nts. tRF-i fragments were found to have primarily length 17 – 23nts and tRF-1 chimeras were primarily formed between a guide tRFs approximately 21 nts long, similar to tRF-5 (Fig. 1C, Supp. Fig. 1).

As another test for non-randomness of tRF generation processes, we considered potential correlation between the levels of mature tRNAs and their corresponding fragments in HEK293 cells. We compared the abundance of Ago1-loaded tRFs with tRNA abundance determined by hydro-tRNA [105] and also with tRF levels of small RNA [106] sequencing data for human kidney cells (not Ago-loaded). We found no correlation between Ago1-loaded tRFs and total cytoplasmic tRF abundance or tRNA levels (Supp. Table 1). This is consistent with the unequal loading of tRFs from the same tRNA gene to Ago1 that we observed, further indicating that all types of tRFs are not equally abundant.

Taken together, our results suggest that distinct types of tRFs are likely to be generated by different mechanisms of cleavage, in agreement with our previous findings for

divergent changes in abundance with age for different types of tRFs in *Drosophila* [32] and rat brains [33]. Our findings with tRFs detected in CLASH further support the notion that tRFs have structure-dependent cleavage sites and they are not byproducts of random degradation.

2. General features of tRF/target interactions

We analyzed all possible interactions between tRFs and their respective RNA targets, identified as distinct transcriptome fragments found within the same chimera. We observed a total of 36,140 CLASH chimeras with tRFs as guides and a variety of target RNAs. After removing all combinations of one tRF with overlapping sequences of the same target, we obtained 1,447 unique chimeras. Following the logic of CLASH experiments, we considered frequent occurrences of the same tRF/target RNA pair as evidence of interaction with a target. We did not restrict ourselves to protein coding genes and took into account every possible hybrid read formed between a tRF and its target RNA. We found that tRFs interact with a wide variety of RNAs including mRNAs, lincRNAs, rRNAs, and miRNAs (Fig. 3). Similarly, mRNA, lincRNAs and rRNAs have been earlier identified as targets of miRNAs [90].

Next, we calculated the targeting frequency of individual regions of mRNAs by tRFs. In addition to untranslated regions (UTRs) we allowed both CDS and intronic sequences as tRF targets, given the evidence that Ago proteins also localized in the nucleus [89, 101]. We found that all types of tRFs primarily target 3' UTR regions of targeted mRNAs followed by CDS and intronic regions (Fig. 3A, C). tRF-i were found to bind mostly to introns and 3' UTRs (Fig. 3B). The most frequent protein coding targets of Ago1 loaded tRFs which constitute the majority of the captured chimeras are listed in Supp. Table 2.

Our findings suggest that tRFs are likely to target RNA molecules in multiple regions, expanding beyond the canonical 3'UTR targeting mode like miRNAs [107, 108]. Strikingly, for tRF-3 type tRFs, we found that they can potentially guide Ago to a specific type of short introns called agotrons. Agotrons are identified based on length (< 150 nts) and association with Ago proteins: Ago-2 has been shown to bind some 30 nts on the 5' end (and in a few cases, on the 3' end) of the agotrons [102]. Despite the low number of known agotrons, we found several cases of tRF-3 guides forming chimeras in Ago1 precisely with the 5' ends of three agotrons (BRD4, LMNA and RRP36). Therefore, it is possible that tRFs guide Ago proteins to the borders of such agottron targets.

In addition to sense transcripts, we found that tRFs potentially also target Natural Antisense Transcripts (NATs) [109-111]. We found 293 unique interactions between tRFs and NATs. In detail, we observed that the interactions between tRFs and NATs included longer isoforms of tRF-5 and tRF-i whereas tRF-3 showed very similar size distribution between transcriptomic and NAT targets. Last, we report that overall the chimeras

between tRFs and NATs were less specific compared to chimeras between tRFs and transcripts according to the MFE of the captured interactions (Supp. Fig 2).

3. tRF/target hybridization modes

We then used specific RNA-RNA chimeras we identified from CLASH data to elucidate the binding modes of tRFs. We started by examining whether there is a particular seed region [44] which can drive the mechanism of tRF/target recognition (as is the case with miRNA). Our results show that only 9% of the CLASH captured chimeras involve a reverse complementary match of a full 7-mer seed sequence (and that includes 7-mers located anywhere in a tRF, not only at the typical seed location near the 5' end). Given such paucity of perfect 7-mer seed sequences for tRFs and the large dataset at our disposal, we analyzed the tRF/target interactome using the following *ab initio* approach.

In order to fully understand what drives tRF target recognition, we used RNAhybrid [104] to predict the hybridization patterns between tRFs and their corresponding targets. Next, we encoded each target-binding or not binding nucleotide across the length of a tRF as 1 or 0, respectively, and applied k-means clustering to reveal distinct binary signatures of interactions for tRF-3, since this type of tRFs were found to form the most unique interactions (Fig. 4). We selected tRFs of the length corresponding to the highest peak in the respective length distribution (Fig. 1C) of guide RNAs found in Ago-1 loaded chimeras. This gave us 300 unique tRF-3 guide/target RNA chimeras.

We detected five consistent and similarly shaped clusters of binding patterns between the nucleotides of guide tRFs and their target RNAs. The clusters were different in size (Fig. 4) and in their average MFE of hybridization (Fig. 4). These cluster shapes revealed several main recurrent themes in binding patterns of tRF-3 tRFs which may also be relevant for other tRFs.

Many chimeras showed binding primarily on the 5' end of the respective tRFs, often involving their middle section as well. The strongest interactions were observed for Clusters 5, 2 and 1 for which there is clear pattern of binding on the 5' end. Interestingly, clusters 4 and 5 showed consistent involvement of nucleotides on the 3' end of tRF-3s (especially nts 16-17). Cluster 3 contained chimeras with binding nucleotides located in the tRF middle yielding the least specific cluster of interactions (Fig 4).

Overall, these binding patterns include nucleotides located across the whole length of a tRF, with several dominant positions near the 5' end of the guide molecule. It appears that there is a clear preference for hybridization on the 5' of tRF-3, however, 3' nucleotides can also hybridize with the target RNAs and such cases result in the most specific interactions. These observations are similar to what has been reported for miRNAs [44].

4. Analysis of tRF/target interactions reveals major interacting sites for tRFs

We next sought to determine whether tRF/target interactions can be driven by major interaction sites/motifs in tRF guide molecules. To identify such sites, we searched for statistically over-represented sequences among the CLASH identified target RNAs of each respective major tRF isoform using MEME [112]. We selected tRFs that had a minimum of five distinct unique chimeras and we used the most abundant CLASH isoform as tRF guide sequence. Next, for every statistically over-represented sequence found using MEME, we searched whether such a sequence can match in reverse complementary orientation to a sub-sequence of the respective guide tRF using FIMO [113]. Enriched sequences that were found to match sub-sequences of guide tRFs were considered major interacting sites for tRFs.

Overall, we found one tRF-5 and six tRF-3 major tRF isoforms with an enriched motif within the set of target sequences and a reverse complement of the same motif matched to a sub-sequence of the respective guide tRF. Similar to previously reported results for miRNAs [44], most core interaction sites were found at the 5' end of the tRF molecule (five out of seven core interaction sites). For one tRF-3 and the only tRF-5 for which a core element was identifiable, we observed that the core sites were located on the 3' end of the guide tRF sequence (top two tRFs, Table 1).

Recently, the presence of a 5' core/seed element has been experimentally validated for three different tRF-3 (Leu-AAG, Leu-TAA, Cys-GCA-014) in HEK293 cells [20]. We found these tRFs to be the most abundant among Ago1-loaded ones (Fig. 2). Strikingly, motifs for all three of them are also present in Table 1. Furthermore, for all three tRFs the

location of the major interaction site was found on the 5' end of the tRF in agreement with the experiment (Fig. 5).

Evidence from two independent studies suggest that all three tRFs mentioned above recognize targets through a 5' seed. It appears that although both methods agree on the location of the seed sequence the actual length of seed sequences, as predicted by our pipeline, can vary. We report seeds ranging from 5 – 13 nts but it needs to be noted that additional nucleotides across the length of the guide tRF might be required for effective binding to target RNAs, in agreement with the clustering results (Fig 4).

Overall, there is clear evidence that core sites/seeds are primarily located close to the 5' end of tRFs. However, the possibility of 3' seed sequences and complementary binding of nucleotides located on the 3' end of tRFs should not be ruled out.

To evaluate the post-transcriptional regulatory capabilities of tRFs, we utilized prior results from an extensive screening for hundreds of post-transcriptional cis-regulatory elements (8-mers) within a reporter 3' UTR in HEK293-FLP cells [106]. We took into consideration all Ago1-captured chimeras both for tRFs and miRNAs and compared their targets for presence of 8-mers that were have been described as having activating, repressive or no effect according to the post-transcriptional cis-regulatory element screen (Supp. Table 3).

We found that overall tRFs target post-transcriptional cis-regulatory elements in a similar fashion to miRNAs when loaded to Ago1. With regards to distinct types of tRFs, we found

that tRF-5 target more repressing elements than tRF-i and tRF-3 and tRF-i share the largest portion of activating cis-regulatory elements compared to tRF-5 and tRF-3.

Discussion

In this report we characterized tRFs and their respective targets loaded to Ago1 proteins and identified putative targeting modes of tRFs using CLASH dataset from human kidney cells. We show that tRFs target a wide variety of transcripts through distinct modes of hybridization including tRF core interaction sites similar to seed sequences, previously identified for miRNAs. We show that all types of tRFs are found loaded to Ago1 proteins. tRFs that originate from mature tRNAs are vastly abundant (>98%) among Ago1 chimeras. The most frequent tRFs were 17-21 nts long, similar to the length of miRNAs.

Despite extensive similarities between tRFs and miRNAs [3, 13, 32], tRFs have been often neglected as products of random degradation of parental tRNA molecules. Here, we show that levels of tRFs loaded to Ago1 do not correlate with the expression of the parental tRNA gene. We also examined tRFs from total cell fraction and we show that there is no correlation between the amount of tRFs produced by a tRNA gene and the abundance of the same tRNA. Our findings are concordant with the results from separate studies [114-116]. Consistent lack of correlation between tRNA and tRF levels supports the view that tRFs are not products of random degradation and different types of tRFs may be produced through different mechanisms.

We considered in detail the type of tRF targets and observed among them non-coding RNAs in addition to mRNAs. We show that mRNAs are the most frequent targets for tRFs loaded to Ago1. Within mRNA targets, we observed that 3' UTRs were the most frequently targeted regions overall followed by coding sequences. 5' UTRs were found to be the least

frequently targeted region of mRNAs for all types of tRFs. Our results further support the hypothesis that tRFs can interact with UTR and CDS regions of mRNAs similar to miRNAs [107, 108]. Ribosomal protein genes RPL35A, RPS14 and RPL7L1 were among the most frequent coding targets of Ago loaded tRFs (Supp. Table 2), suggesting that tRFs may be involved in global translational regulation in addition to posttranscriptional of specific targets, similar to a hypothesis put forth in a recent *Drosophila* study [117].

Recent evidence with regards to nuclear localization of Ago proteins [101, 118] and tRF-5 [13] prompted us to consider intronic regions. Recent discoveries about miRNAs indicate that there is tremendous versatility with regards to their functions. In addition to their clear roles in the cytoplasm, miRNAs have been found also in the nucleus where they are hypothesized to regulate mRNA stability in nucleoli and alternative splicing [119]. We found that tRFs also interact with introns in pre-mRNAs (Fig. 3). It is possible that tRFs may be utilized by Ago in order to guide or regulate splicing. Ago proteins in complexes with miRNAs and siRNAs have been shown to be actively involved in transcriptional regulation and pre-mRNA splicing [101, 118, 120]. Interestingly, we found that tRFs may be involved in guiding Ago to the 5' end of short introns, recently classified as agotrons [102]. Agotrons are identified based on their length (< 150 nts) and interaction with Ago2 in the first 30 nts of their 5' end. We found that tRFs may guide Ago1 to the borders of agotrons and therefore tRFs are likely involved in agotron biogenesis with Ago1 and Ago2.

We report that tRFs also target miRNAs when loaded to Ago1. A previous study revealed that Ago-loaded miRNAs can target tRFs [90]. Taken together, these findings suggest that

tRFs and miRNAs display a similar behavior when loaded to RISC complexes. Both classes of small RNAs can regulate each other in addition to their respective coding or non-coding targets.

Aiming to provide an unbiased view of tRF targeting modes, we selected unique tRF/target RNA chimeras and performed clustering analysis for the most abundant type of tRFs. We found that tRF chimeras showed binding patterns of the guide tRF to the respective target RNAs not limited to specific nucleotides. Overall, we found that when nucleotides located on the 5' end of the guide RNA hybridize to the target RNA the interactions show the highest specificity (lowest MFE) compared to clusters with hybridizing nucleotides in the middle of the tRF. Hybridization of 3' located nucleotides should not be ruled out since it was evident for the two largest clusters of interactions (Fig.4 c4 and c5). Our findings support the notion that tRFs can recognize their targets through a variety of binding patterns in addition to previously reported seed sequence target recognition [32, 33, 46, 79].

A small fraction (9%) of Ago1 tRF chimeras showed such "classical" seed pattern, a 7-mer match between guide tRF and target RNA whereas for miRNAs it was reported that 19% of the chimeras were driven by canonical miRNA "seed" binding [90]. In light of these results, we used an ab initio approach to identify core interaction sites across tRFs. Such an approach allowed us to develop a core site/seed identification approach that is not restricted to canonical "seed" binding which represents a small fraction of the total possible tRF chimeras. We report that for seven tRFs, we were able to identify a conserved

motif likely corresponding to a core interaction site. We identified five core interaction sites close to the 5' end of the guide tRF and two sites were found near the 3' end of the tRF. Such 5' and 3' seeds have been previously reported for tRFs in both experimental and computational studies [32, 46, 79], however, seeds located on the 3' end of a tRF appear to be less frequent. And while the additional binding of 3' located nucleotides might not necessarily be required for suppression of target RNAs, it is observed quite frequently for Ago1 tRF-3 chimeras. Notably, we observed the strongest interactions between tRFs and target RNAs for chimeras with extensive 5' hybridization pattern followed by additional hybridizing nucleotides on the 3' end of the tRF (Fig. 4, c5).

Our results, based on the availability of sufficient numbers of chimeras for inferring interaction motifs, were in a good agreement with the experiment for all three tRFs, for which the 5' seeds have been recently validated [20]. However, we also observed slight variations with the 7-mer seeds proposed in that work. For distinct Cys tRF isoforms we saw a consistent core interaction site, which includes the motif GGGNACC (Table 1), but the longest motif extended further towards the 3' end of a tRF (Fig. 5). Additionally, for Leu-AAG-001 we observed that the enriched core motif was present both at the 5' and 3' end of the tRF (underlined in Fig. 5). It is worth noting that the experimental data (see Fig. 4a in Kuscu et al. [20]) also show support for binding over a part of the 3' instance of the motif. Considering these results and all available chimeras formed between tRF-3 Leu-AAG-001 and RNA targets, the binding frequency of specific nucleotides and the MFE for interactions, the 5' seed hybridization mode with contribution of some 3' nucleotides seems likely for this tRF.

In conclusion, we note that this agreement with the experimental results opens the possibility of inferring the seed regions and mechanisms of tRF/target interactions computationally.

Materials and Methods

CLASH data analysis

CLASH data for HEK293 cells were downloaded from the GEO database (GSE52996) [90]. We used `fastx_toolkit` 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) to remove barcode and adapter sequences and collapse identical reads. We used an in house developed aligner script to identify tRFs from hybrid reads, allowing no mismatches and giving preference to longer tRF isoforms. In detail, the aligner determines if a hybrid read starts with a known tRNA sequence (>16 nts) and checks if the next nucleotide can still be part of the tRF sequence, stopping at the first mismatch. This way, the longest tRF isoform is identified as the guide sequence and the remainder of the hybrid read is considered the targeted sequence. For 5' tRFs we selected tRFs that mapped to the first 5 nucleotides of a known tRNA sequence. For 3' tRFs, we selected tRFs that ended up to 5 nucleotides short of the end of a mature tRNA sequence (including the CCA tri-nucleotide at the end). All the identified tRF containing hybrids were confirmed not to be full tRNAs or pre-tRNAs by running `blastn`, word size 7, default scoring matrix against the union of tRNA sequences from two independent databases [2, 121]. The portion of the hybrid read following the tRF sequence was considered the targeted sequence and it was searched against the human genome (hg38) and the human transcriptome [122] using `blastn`, word size 7, default scoring matrix and 10 maximum hits. Reads were considered chimeras if a hit had an e-value less than or equal to 0.01 and the length of the targeted sequence and the tRF sequence was greater than or equal to 75% of the total length of the chimeric read.

Small RNA sequencing and hydro-tRNA sequencing data analysis

Hydro-tRNA sequencing data for HEK293 cells and small RNA sequencing data from whole cytoplasmic fraction of HEK293 cells were downloaded from the GEO database (GSE95683 and GSE75136) [105, 106] were downloaded. fastx_toolkit 0.0.13 was used to remove adapter sequences and collapse identical reads. Sequencing read alignments were performed using bowtie 1.1.1 aligner (<http://bowtie-bio.sourceforge.net/manual.shtml>). We aligned the sequenced reads against the human genome (version hg38) and also to the union of human tRNA sequences from two independent databases [2, 121]. For each replicate, the raw read counts were normalized by the total number of reads that mapped to the human genome. For hydro-tRNA sequences we allowed up to two mismatches.

Seed sequence and hybridization pattern analysis

We generated 7-mer sub-sequences of tRFs by applying a 7-nt sliding window and shifting by one nt from the 5' to the 3' end. We then calculated the count of exact matches for each of these sub-sequences against the targeted sequences of each tRF obtained from CLASH chimeras. We used RNAhybrid 2.1.2 [104] with default parameters to calculate minimum free energy for observed tRF/target RNA interactions and for random controls. To examine the binding mode of tRFs, we utilized the secondary structures for unique rRF/target chimeras obtained using RNAhybrid. We encoded each nucleotide across the

tRF/target RNA chimera as 0 (if it was predicted not to bind) or 1 (if it was predicted to bind with a nucleotide from the target RNA) and we performed clustering analysis for the most abundant isoforms (with regards to fragment length) for each type of tRFs originating from mature tRNAs. We used scikit-learn (<http://scikit-learn.org/>) to perform unsupervised clustering using k-means algorithm.

Motif enrichment analysis

In order to identify enriched motifs within tRF targets, we selected for every tRNA gene a representative major tRF isoform of each type (tRF-5, tRF-i and tRF-3). We took into account tRFs with at least 5 unique targets according to CLASH data. We used MEME [112] with default parameters (e-value < 0.05) and searched for enriched motifs longer than 5 nucleotides across all targeted sequences for a given tRF isoform. Next, we used FIMO [113] with default parameters (p-value < 0.05) to match such over-represented motifs back to tRF sequences potential major interaction sites.

Regulatory 8-mer analysis

To examine the post-transcriptional regulatory capabilities of tRFs, we utilized the results of a large-scale screen of short cis-regulatory elements (8-mers) in HEK293-FLP cells [106]. This study used a cell-based assay to measure the expression of a GFP-reporter as a readout for the regulatory potential of an 8-mer inserted within the human IQGAP1 3'

UTR. To reveal activating, repressing or elements with no effect, we calculated frequencies of occurrence for all such 8-mers in Ago1-loaded RNA targets targeted by tRF and miRNA guide sequences.

Figure and Table Legends

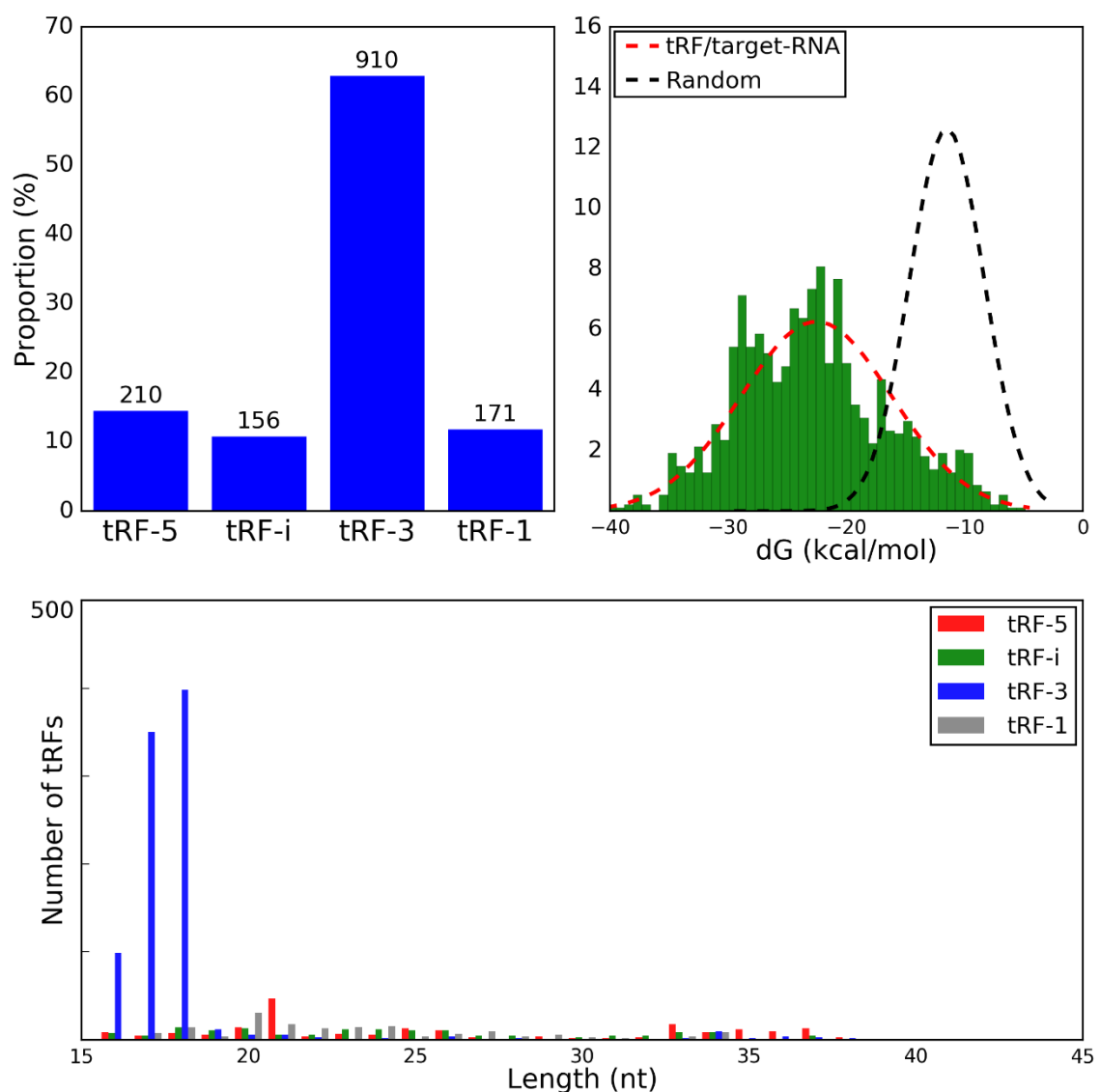


Figure 1. A) Distribution of guide tRF types identified from CLASH RNA chimeras. B) Minimum Free Energy (MFE) histogram for tRF/target RNA chimeras (green histogram and fitted red line) and for randomly generated control interactions (black line). C) tRF length distribution histograms for tRF-5 , tRF-I , tRF-3 and tRF-1 chimeras identified from CLASH data.

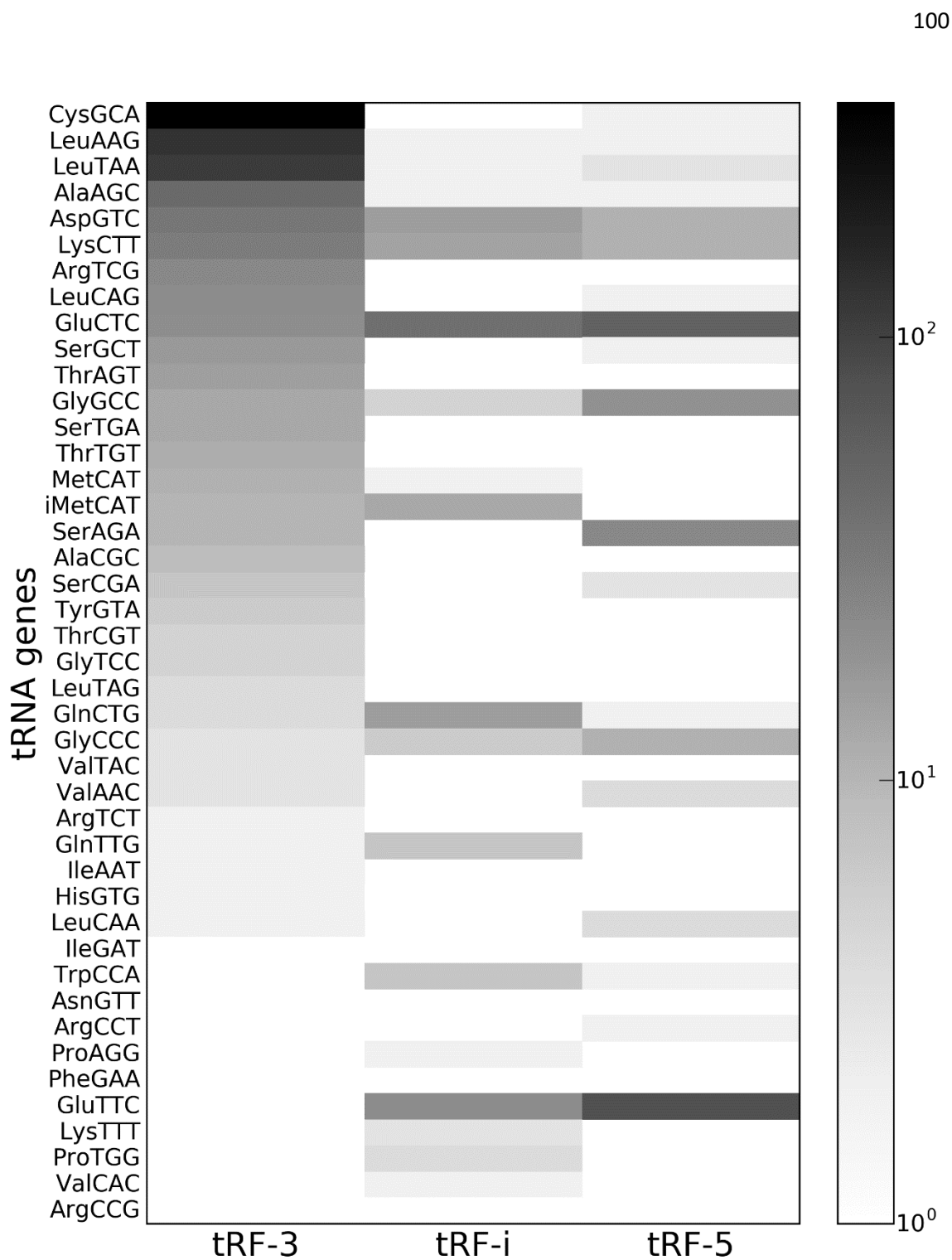


Figure 2. Heatmap of the types of tRFs generated from mature nuclear tRNAs. The scale on the right represents the count of unique chimeric reads found in CLASH data that contained each specific type of tRF as guide sequence.

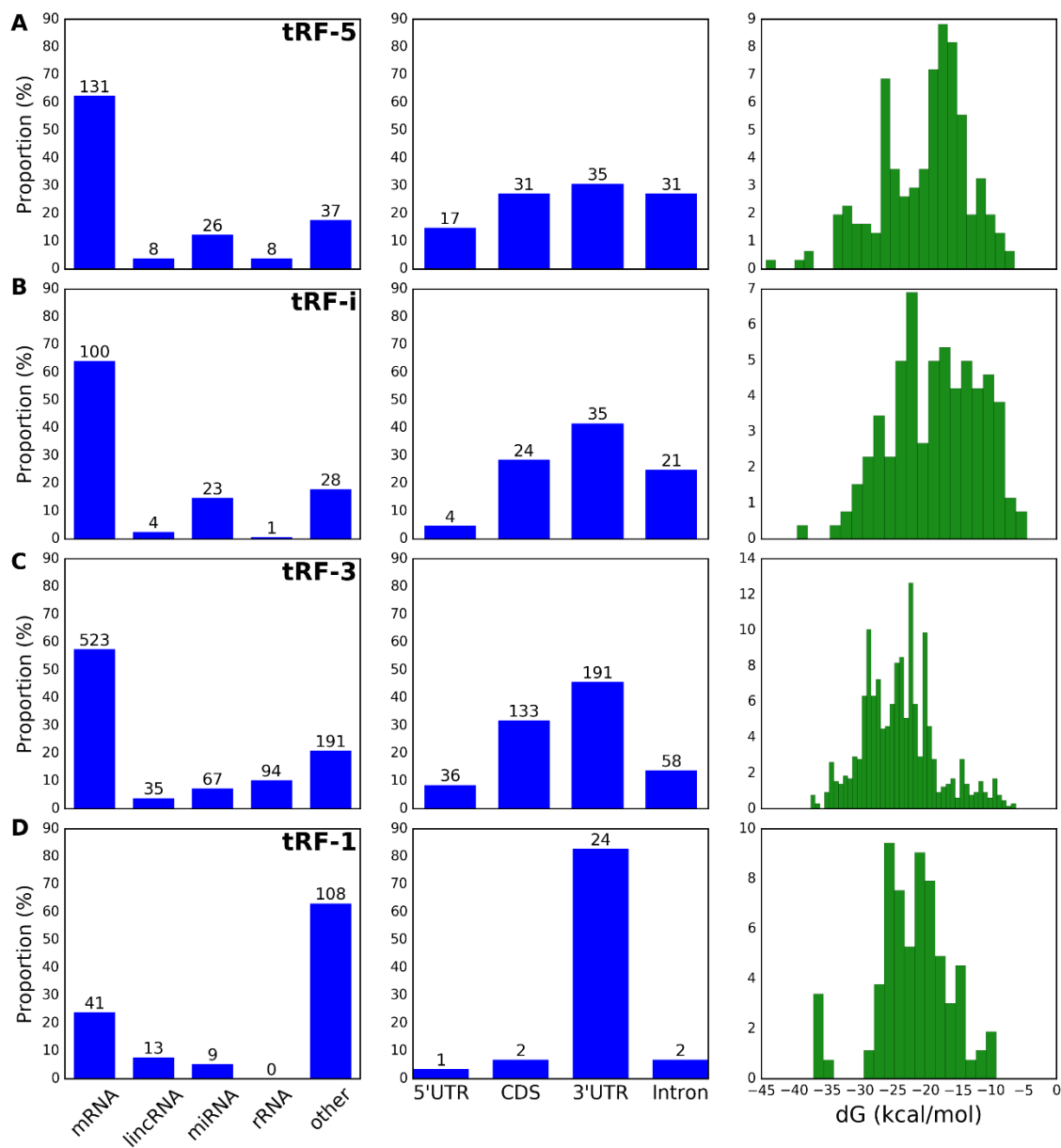


Figure 3. tRFs guide Ago1 to a variety of RNA targets. tRF target distribution plots (left, % on the y-axis and actual unique chimera counts given above the histogram bars), targeting frequency of mRNA regions (middle) and MFE histograms of tRF/target RNA unique chimeras (right). Rows depict tRF-5 (A), tRF-i (B), tRF-3 (C) and tRF-1 (D).

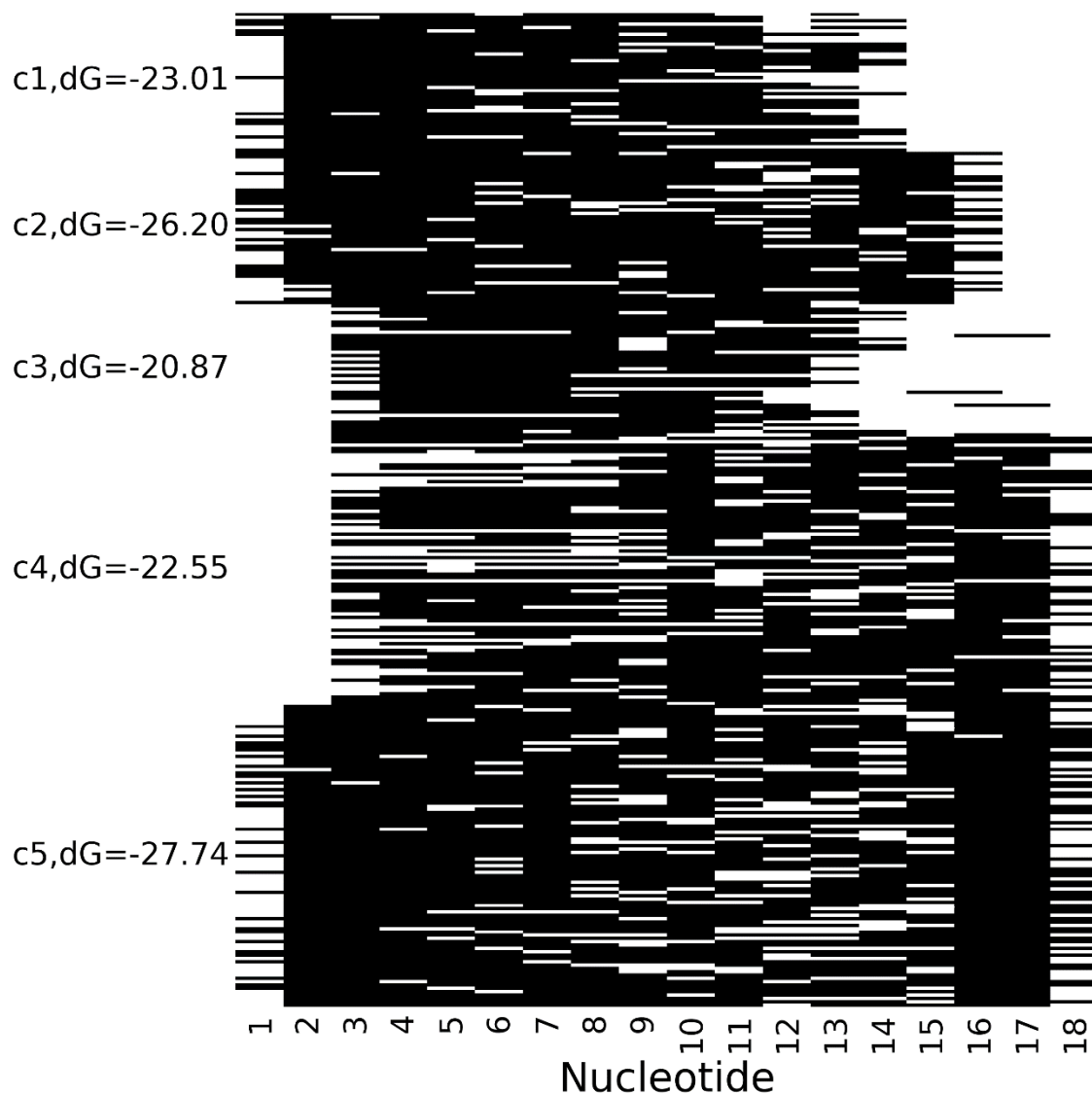


Figure 4. Base-pairing patterns for unique tRF-3/target chimeras. Each line represents a guide tRF from a unique CLASH chimera. Paired nucleotides are depicted in black and unpaired nucleotides are shown in white. The labels C1 through C5 mark the vertical center points of the five identified clusters and the average MFE for the interactions in each cluster is shown.

| | |
|-----------------------------|---------------------|
| AT CCCACC GCTGCCACCA | Leu-AAG / tRF-3001a |
| T CCGGGT GCCCCCTCCA | Cys-GCA / tRF-3003a |
| A CCCCACTCCT GGTACCA | Leu-TAA / tRF-3009a |

Figure 5. Computationally and experimentally identified interaction sites for tRF-3 type tRFs. Computationally predicted interaction site locations are shown in red boxes. Seed regions for the same tRFs proposed on the basis of luciferase assays [20] are shown in bold letters.

| Motif | Motif E-value | Major tRF ID | Start | End | p-value |
|-------|---------------|---------------------|-------|-----|----------|
| | 1.30E-02 | GlyGCC-001-5p-1-33 | 25 | 31 | 9.87E-03 |
| | 9.30E-05 | ArgTCG-002-3p-59-76 | 4 | 14 | 3.58E-04 |
| | 1.00E-06 | AspGTC-002-3p-58-75 | 10 | 17 | 8.13E-06 |
| | 1.80E-06 | CysGCA-001-3p-58-75 | 1 | 13 | 6.63E-07 |
| | 2.70E-03 | CysGCA-012-3p-59-75 | 3 | 11 | 2.23E-06 |
| | 8.10E-17 | CysGCA-014-3p-59-75 | 2 | 13 | 4.66E-08 |


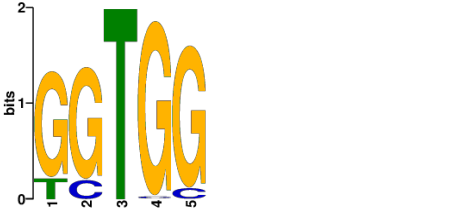
| | | | | | |
|---|----------|-----------------------|----|----|----------|
|  | 1.70E-10 | LeuTAA-001-3p-69-86 | 2 | 13 | 1.13E-07 |
|  | 6.90E-04 | LeuAAG-001-N-3p-68-85 | 4 | 8 | 7.09E-04 |
| | | | 13 | 17 | 7.09E-04 |

Table 1

Table 1. Core motifs/seeds identified for distinct tRFs. All seeds shown were found to match a tRF subsequence using FIMO (p-value < 0.05).

Supplementary Tables and Figures

Supplementary Table 1

| | tRF5 | tRFi | tRF3 | tRF1 |
|----------------|---------|---------|---------|---------|
| d0 | 0.10091 | 0.24481 | 0.01449 | 0.18009 |
| d1 | 0.08967 | 0.236 | 0.03469 | 0.25157 |
| d2 | 0.06927 | 0.21431 | 0.00088 | 0.22447 |
| RNA-seq | 0.2105 | 0.06808 | 0.13312 | 0.30939 |

Supplementary Table 1. Spearman correlation between the abundance of tRFs in CLASH data and the abundance of tRNAs in hydro-tRNA sequencing following the mismatch notation of [105, 106] and in the total cellular fractions of tRFs from [106]

Supplementary Table 2

| Target Transcript | Coding Targets | Chimera Count | Guide tRNAs |
|-------------------|----------------|---------------|---|
| ENST00000421915.5 | AGBL5 | 2858 | Tyr-GTA-7-1,tdbD00003671 Tyr GTA,Tyr-GTA-5-4,tdbD00003670 Tyr GTA |
| ENST00000369159.2 | HIST2H2AA3 | 901 | Cys-GCA-1-1,iMet-CAT-1-1,Cys-GCA-14-1,Cys-GCA-12-1 |
| ENST00000433211.6 | CTNNA3 | 832 | Cys-GCA-1-1,Thr-TGT-5-1,Cys-GCA-14-1,Cys-GCA-12-1 |
| ENST00000394654.3 | RUNDC3B | 686 | Gly-GCC-2-1,Glu-CTC-2-1,Gly-GCC-1-1 |
| ENST00000414069.2 | SPRN | 343 | Gly-CCC-1-1,Gly-GCC-1-1 |
| ENST00000558133.1 | MEX3B | 293 | Glu-CTC-2-1,Glu-TTC-3-1 |
| ENST00000416605.6 | HACE1 | 237 | Glu-TTC-3-1 |
| ENST00000600289.6 | LSM4 | 225 | Arg-TCT-2-1,Ser-GCT-1-1,Ser-GCT-2-1 |
| ENST00000373062.7 | GNL2 | 223 | Leu-CAA-1-1 |

| | | | |
|-----------|---------|-----|---|
| ENST00000 | | | |
| 329092.12 | RPL35A | 223 | Thr-AGT-6-1,Thr-AGT-5-1,Thr-TGT-1-1 |
| ENST00000 | | | Leu-TAA-1-1,Ala-AGC-1- |
| 379719.7 | IPO7 | 179 | 1,mtdbD00000532 Met CAT,Leu-AAG-1-1 |
| ENST00000 | | | |
| 255194.10 | AP3B1 | 172 | Glu-TTC-3-1 |
| ENST00000 | | | Leu-CAG-1-1,mtdbD00000532 Met CAT,Cys-GCA-14- |
| 357121.5 | OCRL | 155 | 1,Leu-AAG-1-1,Leu-TAA-1-1,Gly-GCC-1-1 |
| ENST00000 | HIST2H3 | | |
| 403683.1 | A | 141 | Gln-CTG-4-1,Leu-CAG-1-1,Ser-AGA-2-1 |
| ENST00000 | | | |
| 526277.1 | CELF1 | 137 | Gly-CCC-1-1 |
| ENST00000 | | | |
| 415136.6 | DDOST | 137 | Leu-TAA-1-1,Cys-GCA-2-4,Leu-AAG-1-1 |
| ENST00000 | | | |
| 459829.1 | RPL7L1 | 136 | Cys-GCA-17-1,Cys-GCA-14-1,Gln-CTG-6-1 |
| ENST00000 | | | |
| 630190.1 | DUSP16 | 133 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 626119.2 | COPS7A | 131 | mtdbD00000532 Met CAT |
| ENST00000 | | | |
| 306801.7 | RPTOR | 130 | Arg-TCG-2-1 |
| ENST00000 | | | |
| 478753.4 | SEPHS2 | 129 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 525587.1 | TIMM10 | 128 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 550458.1 | SRSF9 | 124 | Leu-TAA-1-1,Leu-CAG-1-1 |
| ENST00000 | GPATC | | |
| 368232.8 | H4 | 122 | Thr-AGT-6-1,Ala-AGC-1-1 |
| ENST00000 | HSP90A | | |
| 620073.4 | B1 | 120 | Asp-GTC-2-1,Gly-GCC-1-1 |
| ENST00000 | | | |
| 592456.1 | SYNGR2 | 119 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 399794.6 | RBMXL1 | 111 | His-GTG-1-1 |
| ENST00000 | | | |
| 567171.1 | TCF25 | 109 | Leu-TAA-1-1 |
| ENST00000 | | | |
| 358296.10 | ZNF100 | 104 | Glu-CTC-2-1 |
| ENST00000 | | | |
| 407193.5 | RPS14 | 101 | Asp-GTC-2-1,Glu-CTC-2-1 |
| ENST00000 | SLC25A | | |
| 451283.5 | 1 | 100 | Gly-CCC-1-1 |
| ENST00000 | | | |
| 370990.5 | SERBP1 | 99 | iMet-CAT-1-1 |
| ENST00000 | | | |
| 554636.1 | VTI1B | 98 | Leu-TAA-1-1 |
| ENST00000 | | | |
| 369839.3 | TAF5 | 98 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 261303.12 | PSMC1 | 96 | Glu-TTC-3-1 |
| ENST00000 | | | |
| 523929.1 | HSPA9 | 96 | Leu-CAA-1-1 |
| ENST00000 | | | |
| 342988.7 | SMAD4 | 95 | Lys-TTT-3-1 |

| | | | |
|-----------|---------|----|---|
| ENST00000 | | | |
| 423698.6 | ERCC1 | 87 | Glu-CTC-2-1,Glu-TTC-3-1 |
| ENST00000 | | | |
| 399627.3 | SBF1 | 86 | Arg-CCG-2-1 |
| ENST00000 | | | |
| 579190.1 | PLXDC1 | 85 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 280481.8 | FREM2 | 85 | Leu-AAG-3-1,Leu-AAG-1-1 |
| ENST00000 | | | |
| 256015.4 | BTG1 | 83 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 535269.1 | KDM5A | 82 | Leu-TAA-1-1 |
| ENST00000 | | | |
| 552516.5 | TGFBR1 | 81 | Leu-TAA-1-1 |
| ENST00000 | | | |
| 287594.7 | MPV17L | 78 | Gln-TTG-3-1,Lys-CTT-3-1,Asp-GTC-5-1,Glu-CTC-2-1 |
| ENST00000 | | | |
| 629427.2 | SHPRH | 75 | Cys-GCA-12-1 |
| ENST00000 | SLC15A | | |
| 469013.1 | 2 | 73 | Gly-CCC-1-1,Ala-AGC-2-1,Gly-TCC-3-1,Gly-GCC-1-1 |
| ENST00000 | | | |
| 541166.1 | PWP1 | 73 | Ser-TGA-1-1 |
| ENST00000 | | | |
| 521381.5 | PIK3R1 | 72 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 528746.5 | PRDM10 | 72 | Leu-AAG-1-1 |
| ENST00000 | | | |
| 521479.1 | GSR | 71 | Glu-TTC-3-1 |
| ENST00000 | | | |
| 372616.1 | CTPS1 | 71 | Leu-TAA-1-1 |
| ENST00000 | | | |
| 520817.5 | GOLGA7 | 68 | Glu-CTC-1-5 |
| ENST00000 | | | |
| 300035.8 | PCLAF | 68 | Asp-GTC-4-1 |
| ENST00000 | | | |
| 466473.1 | HILPDA | 67 | Val-CAC-10-1 |
| ENST00000 | | | |
| 378665.1 | UQCRQ | 67 | Arg-TCG-2-1 |
| ENST00000 | | | |
| 560846.1 | BAHD1 | 67 | Leu-TAA-1-1 |
| ENST00000 | ATP5J2- | | |
| 413834.5 | PTCD1 | 66 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 285928.2 | LRGUK | 66 | Lys-CTT-3-1 |
| ENST00000 | | | |
| 368320.7 | KHDC4 | 66 | Arg-TCT-4-1 |
| ENST00000 | AC0064 | | |
| 594664.1 | 86.1 | 65 | Cys-GCA-1-1,Asp-GTC-2-1 |
| ENST00000 | SELENO | | |
| 398226.7 | S | 64 | Gly-GCC-1-1 |
| ENST00000 | | | |
| 389902.7 | RNF216 | 63 | Thr-CGT-2-1 |
| ENST00000 | HIST1H2 | | |
| 359193.3 | AG | 63 | Leu-AAG-1-1 |
| ENST00000 | | | |
| 549525.1 | COX6A1 | 62 | Glu-CTC-2-1 |

| | | | |
|-----------|---------|----|-------------------------------------|
| ENST00000 | | | |
| 578339.1 | ZNF830 | 62 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 282007.7 | ZC3H13 | 62 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 395587.5 | EIF3C | 61 | Leu-TAA-1-1 |
| ENST00000 | | | |
| 490734.6 | DBNL | 61 | Ser-GCT-5-1,Glu-CTC-2-1,Gly-GCC-1-1 |
| ENST00000 | | | |
| 540357.5 | DNMT1 | 61 | Leu-TAA-1-1 |
| ENST00000 | HNRNP | | |
| 634634.1 | R | 61 | Cys-GCA-14-1 |
| ENST00000 | HNRNP | | |
| 600596.1 | UL1 | 60 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 246229.4 | PLAGL2 | 60 | Cys-GCA-12-1 |
| ENST00000 | KIAA152 | | |
| 401073.6 | 2 | 60 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 369425.5 | GPAM | 60 | Gln-TTG-3-1,Lys-CTT-3-1,Glu-CTC-2-1 |
| ENST00000 | | | |
| 377704.4 | MGME1 | 59 | mtdbD00000532 Met CAT |
| ENST00000 | | | |
| 544292.5 | PARD3 | 58 | Gly-GCC-1-1 |
| ENST00000 | | | |
| 376583.7 | MTHFR | 58 | SeC-TCA-2-1 |
| ENST00000 | | | |
| 314032.8 | UCP3 | 58 | Leu-CAA-1-1 |
| ENST00000 | | | |
| 393436.9 | RAP1B | 58 | Ala-AGC-1-1 |
| ENST00000 | | | |
| 620444.4 | DACH1 | 58 | Glu-TTC-3-1 |
| ENST00000 | | | |
| 343813.9 | ICMT | 58 | Arg-TCG-2-1 |
| ENST00000 | | | |
| 544496.5 | PCMT1 | 58 | Lys-CTT-3-1 |
| ENST00000 | | | |
| 539332.2 | DDB1 | 56 | His-GTG-1-1 |
| ENST00000 | | | |
| 598742.5 | RPS19 | 56 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 072644.6 | YIPF1 | 56 | Ala-CGC-2-1 |
| ENST00000 | ADAMT | | |
| 380548.8 | SL1 | 56 | mtdbD00000532 Met CAT |
| ENST00000 | | | |
| 262861.8 | SIPA1L2 | 56 | Cys-GCA-14-1 |
| ENST00000 | | | |
| 514697.1 | PAPD7 | 56 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 373715.10 | SRSF3 | 56 | Gly-GCC-1-1 |
| ENST00000 | | | |
| 258412.7 | TMBIM1 | 56 | Lys-CTT-3-1 |
| ENST00000 | | | |
| 588852.1 | SAFB | 55 | Cys-GCA-14-1,Leu-AAG-1-1 |
| ENST00000 | HNRNP | | |
| 618183.4 | A2B1 | 55 | Leu-TAA-1-1,Ala-AGC-1-1 |

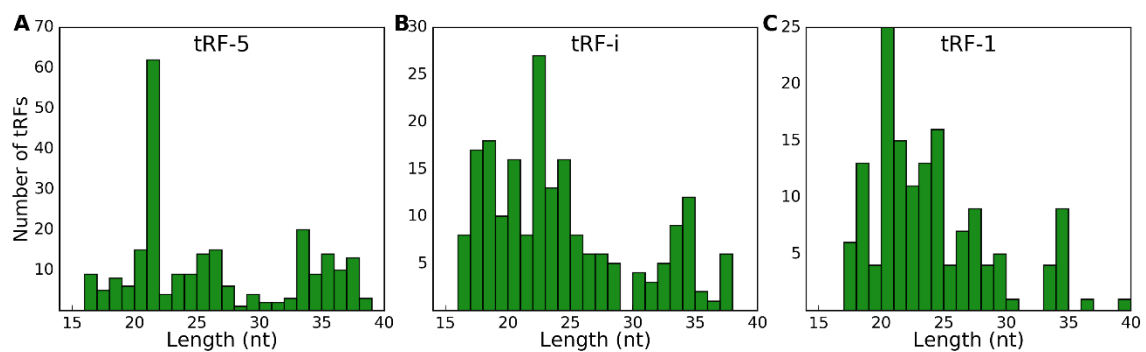
| | | | |
|-----------|---------|----|-------------|
| ENST00000 | HIST1H2 | | |
| 358739.4 | AI | 55 | Ser-TGA-1-1 |
| ENST00000 | | | |
| 291442.3 | NR2F6 | 55 | Asp-GTC-2-1 |
| ENST00000 | | | |
| 551217.1 | RPLP0 | 54 | Leu-TAG-2-1 |
| ENST00000 | | | |
| 261254.7 | CCND2 | 54 | Ser-GCT-2-1 |
| ENST00000 | | | |
| 524227.5 | ANK1 | 54 | Ser-AGA-2-1 |
| ENST00000 | HIST2H4 | | |
| 613412.1 | A | 54 | Leu-AAG-1-1 |
| ENST00000 | | | |
| 268711.3 | MED9 | 53 | Cys-GCA-1-1 |

Supplementary Table 2. List of the 100 most frequent protein coding targets (accounting for XX% of all targets) of tRFs in CLASH chimeras.

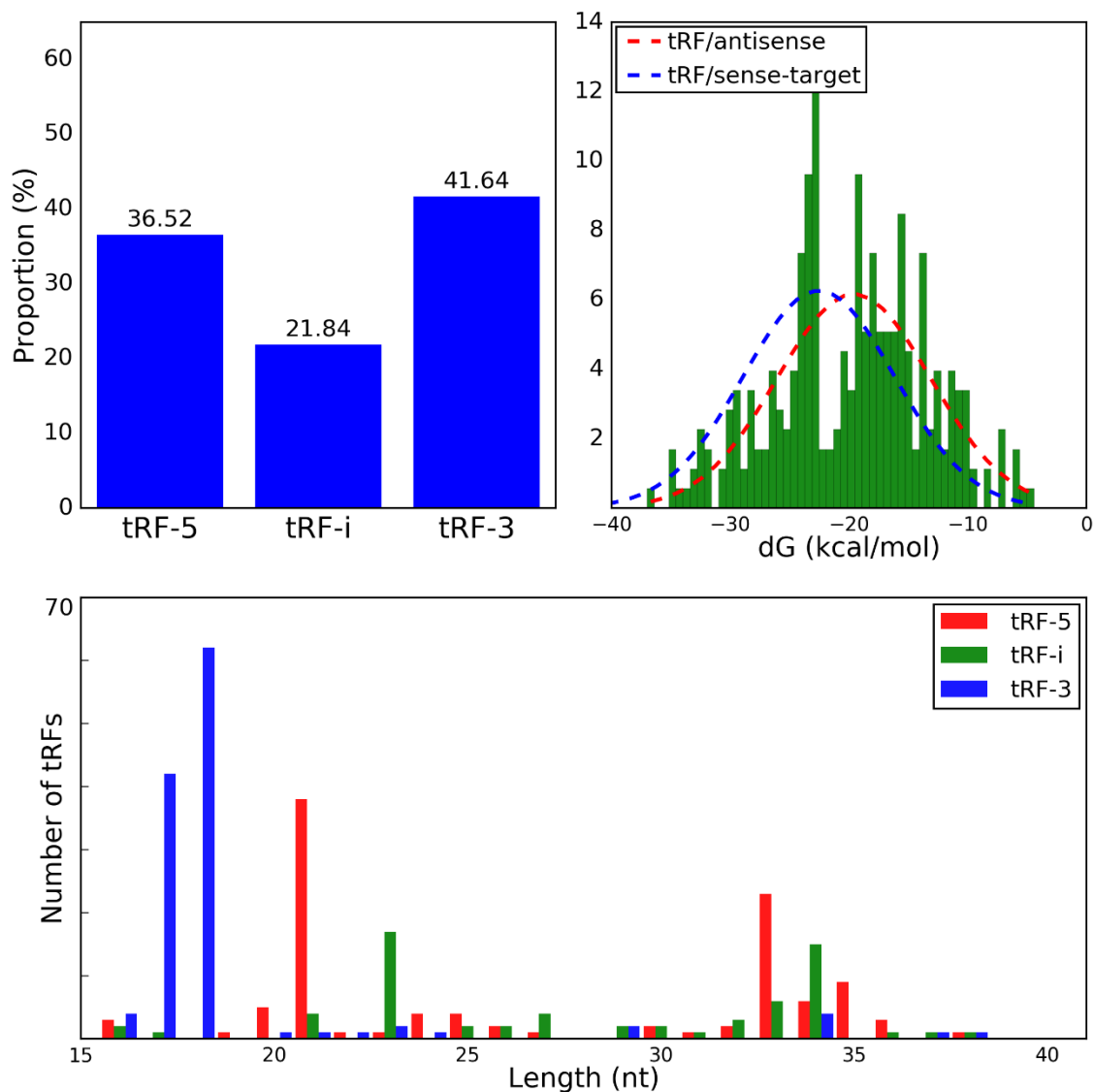
Supplementary Table 3

| Chimera guide | repressing | no effect | activating |
|---------------|--------------|---------------|--------------|
| miRNA | 3751 (19.1%) | 11205 (57.2%) | 4627 (23.6%) |
| tRF | 2689(20.7%) | 7631(58.9%) | 2640(20.4%) |
| tRF-5 | 932(51.0%) | 667(36.5%) | 229(12.5%) |
| tRF-i | 396(17.0%) | 975(41.9%) | 955(41.1%) |
| tRF-3 | 1361(15.5%) | 5989(68.0%) | 1456(16.5%) |
| tRF-1 | 90(15.6%) | 216(37.4%) | 272(47.1%) |

Supplementary Table 3. CLASH supported chimeras for miRNAs and tRFs that contain a regulatory 8-mer within the captured target RNA of the chimera.



Supplementary Figure 1. Length distribution plots for the less abundant tRF-5 (A), tRF-i (B) and tRF-1 (C) shown in Figure 1.



Supplementary Figure 2. Distribution of guide tRFs identified in tRF/antisense RNA chimeras from CLASH data. B) Minimum Free Energy (MFE) histogram for tRF/antisense RNA chimeras (green histogram and fitted red line). Interactions between tRFs and sense RNAs are shown in blue. C) tRF length distribution histograms for tRF-5, tRF-i and tRF-3 chimeras with antisense targets (no tRF-1 interactions with antisense targets were observed)

Materials And Methods

Chapter 1 Materials and methods

Mapping and quantifying tRFs

We used *Drosophila* Ago-IP libraries GSM1278635, GSM1278636, GSM1278637 and GSM1278638 available from the Gene Expression Omnibus (GEO) database, with experimental details described earlier [47]. Adaptor sequences were removed from the 3' end of the reads in the Illumina fastQ files using the fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The adapter sequences are as follows:

5' adapter = 5'- GUUCAGAGUUCUACAGUCCGACGAUC- 3'

3' adapter = 5'- TGGAATTCTCGGGTGCCAAGG- 3'

Reads were then collapsed and annotated with the number of times each was sequenced, so only unique reads were analyzed. The reads were then mapped using Bowtie to the *D. melanogaster* (dm5) genome and tRNAs obtained from FlyBase. Bowtie parameters were restricted to only output perfectly aligned matches to the tRNA sequence. The reads were aligned and mapped to the entire tRNA sequence with the CCA addition. After mapping

reads to their respective tRNAs, each library was independently normalized by the total number of reads mapped to the *D. melanogaster* genome (v. R6.03).

Differential/Preferential loading with age

We identified differential loading of tRFs with age in Ago1 and Ago2 using a ratio metric. We first identified the most abundant isoform in our 30 day libraries and used the read count numbers of that specific isoform for our ratio calculations. We calculated the ratio of 30 days to 3 days for Ago1 and Ago2 of highly abundant (1000 or more reads) tRFs. We then plotted the ratios to see loading changes that may occur with age.

To observe what was preferentially loaded (Ago1 vs Ago2) with age, we obtained a different ratio. The ratio of this measure was the ratio of reads of a particular tRF of Ago2 to Ago1 at 3 days and at 30 days.

Analysis of Seeds, Targeting and GO Terms

In order to identify a potential seed sequence in our dataset, we generated k-mer subsequences of the tRF by applying a sliding window by shifting one nt towards the 3' end after each subsequent k-mer generation. We then found exact matches for each of these subsequences to the conserved 5' UTR, 3' UTR, exon and intron regions of 12 *Drosophila* genomes provided by UCSC [72] and to those regions in the *D. melanogaster*

genome. We then compared for each k-mer in a tRF the observed number of its matches in the conserved 3' UTR regions with the expected number (based on the frequency of matches across the *D. melanogaster* genome) and with the average number of matches of all possible k-mers with the same nucleotide composition in conserved 3'UTRs to identify candidate seeds. Genes with exact matches of 7mer candidate seeds to the longest annotated 3'UTR were considered potential targets. While our approach is similar to TargetScan [43-45], we did not use its "context score" as it was unlikely to be applicable for our cases of both 3' and 5' seeds. To find the preferentially targeted regions we normalized the total match counts by the total length of each respective set of regions. AmiGO [55] was used to find enriched GO-terms in our target list for each tRF.

Chapter 2 Materials and methods

Small RNA analysis

We used publicly available datasets of small RNA from rat brain [81] with accession number ERA36511. Using the sra-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) we converted the files to fastq format using fastq-dump and removed the 3' adapter sequences with fastx-clipper. The reads of length above 16 nts were used for downstream analysis. We collapsed and mapped the reads to the rat genome (rn6, UCSC) and the union of rat tRNAs from two independent databases (<http://gtrnadb.ucsc.edu> and <http://trnadb.bioinf.uni-leipzig.de> also including mitochondrial tRNA genes from the second one) using Bowtie. Bowtie parameters were set to output only perfect matches to tRNA sequences (including the post transcriptional CCA modification). Read counts in each experiment were normalized by the total number of reads detected and averaged across three replicates for each of the three time points (ages of 6, 14 and 22 months).

Seed sequence analysis

We generated 7-mer subsequences of tRFs by applying a 7-nt sliding window and shifting by one nt from the 5' to the 3' end. We then found the counts of exact matches for each

of these subsequences to the 3' UTR regions conserved in at least 15 species, including human mouse and rat (<http://www.targetscan.org/>). To estimate significance of the seed matches we compared the observed match counts for each respective 7-mer in a tRF to (i) those expected number of matches by chance (estimated from 7-mer genomic frequency) and to (ii) average numbers of matches of all possible 7-mers with the same nucleotide composition in conserved 3'UTRs. Genes with exact matches of 7-mer and 7-mer_1a candidate seeds to the 3'UTR were considered potential targets.

RNA sequencing analysis

For target expression analysis we downloaded files with pre-computed transcript expression levels for the rat cerebral cortex transcriptome, data series with accession number GSE34272 [85]. These expression levels in each experiment were normalized by the total number of reads detected and averaged across three replicates for each of the three time points (ages of 6, 12 and 28 months).

Statistical evaluation of downregulation levels for miRNA and tRF targets

For each set of predicted targets of a tRF or a miRNA, we compared its ratio of down-regulated/up-regulated target transcripts from young to old rats with the distribution of

such ratios calculated for 1,000 randomly selected transcript sets of the same size as the target set (different for each tRF and miRNA). This process was repeated three times for three different thresholds (up-regulated by >5% / downregulated by >5%, up-regulated by >10% / downregulated by >10% and up-regulated by >20% / downregulated by >20%) using R statistical package (www.R-project.org) to find statistical significance of the difference observed.

Gene ontology enrichment analysis

The predicted targets for each tRF were used as input in order to perform GO enrichment analysis. Each set of targets was uploaded to PANTHER website (<http://pantherdb.org/>, [95]) and results were obtained using the recommended default parameters.

Chapter 3 Materials and Methods

CLASH data analysis

CLASH data for HEK293 cells were downloaded from the GEO database (GSE52996) [90]. We used *fastx_toolkit* 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) to remove barcode and adapter sequences and collapse identical reads. We used an in house developed aligner script to identify tRFs from hybrid reads, allowing no mismatches and giving preference to longer tRF isoforms. In detail, the aligner determines if a hybrid read starts with a known tRNA sequence (>16 nts) and checks if the next nucleotide can still be part of the tRF sequence, stopping at the first mismatch. This way, the longest tRF isoform is identified as the guide sequence and the remainder of the hybrid read is considered the targeted sequence. For 5' tRFs we selected tRFs that mapped to the first 5 nucleotides of a known tRNA sequence. For 3' tRFs, we selected tRFs that ended up to 5 nucleotides short of the end of a mature tRNA sequence (including the CCA tri-nucleotide at the end). All the identified tRF containing hybrids were confirmed not to be full tRNAs or pre-tRNAs by running *blastn*, word size 7, default scoring matrix against the union of tRNA sequences from two independent databases [2, 121]. The portion of the hybrid read following the tRF sequence was considered the targeted sequence and it was searched against the human genome (hg38) and the human transcriptome [122] using *blastn*, word size 7, default scoring matrix and 10 maximum hits. Reads were considered chimeras if a hit had

an e-value less than or equal to 0.01 and the length of the targeted sequence and the tRF sequence was greater than or equal to 75% of the total length of the chimeric read. Chimeras with targeted sequence that mapped in the same orientation as transcription were taken into account for downstream analyses. The same pipeline with identical parameters was applied for antisense transcripts (NATs) following the adjustment that the orientation of targets was opposite to the direction of transcription.

Small RNA sequencing and hydro-tRNA sequencing data analysis

Hydro-tRNA sequencing data for HEK293 cells and small RNA sequencing data from whole cytoplasmic fraction of HEK293 cells were downloaded from the GEO database (GSE95683 and GSE75136) [105, 106] were downloaded. *fastx_toolkit* 0.0.13 was used to remove adapter sequences and collapse identical reads. Sequencing read alignments were performed using bowtie 1.1.1 aligner (<http://bowtie-bio.sourceforge.net/manual.shtml>). We aligned the sequenced reads against the human genome (version hg38) and also to the union of human tRNA sequences from two independent databases [2, 121]. For each replicate, the raw read counts were normalized by the total number of reads that mapped to the human genome. For hydro-tRNA sequences we allowed up to two mismatches.

Seed sequence and hybridization pattern analysis

We generated 7-mer sub-sequences of tRFs by applying a 7-nt sliding window and shifting by one nt from the 5' to the 3' end. We then calculated the count of exact matches for each of these sub-sequences against the targeted sequences of each tRF obtained from CLASH chimeras. We used RNAhybrid 2.1.2 [104] with default parameters to calculate minimum free energy for observed tRF/target RNA interactions and for random controls. To examine the binding mode of tRFs, we utilized the secondary structures for unique rRF/target chimeras obtained using RNAhybrid. We encoded each nucleotide across the tRF/target RNA chimera as 0 (if it was predicted not to bind) or 1 (if it was predicted to bind with a nucleotide from the target RNA) and we performed clustering analysis for the most abundant isoforms (with regards to fragment length) for each type of tRFs originating from mature tRNAs. We used *scikit-learn* (<http://scikit-learn.org/>) to perform unsupervised clustering using k-means algorithm.

Motif enrichment analysis

In order to identify enriched motifs within tRF targets, we selected for every tRNA gene a representative major tRF isoform of each type (tRF-5, tRF-i and tRF-3). We took into account tRFs with at least 5 unique targets according to CLASH data. We used MEME [112]

with default parameters (e-value < 0.05) and searched for enriched motifs longer than 5 nucleotides across all targeted sequences for a given tRF isoform. Next, we used FIMO [113] with default parameters (p-value < 0.05) to match such over-represented motifs back to tRF sequences potential major interaction sites.

Regulatory 8-mer analysis

To examine the post-transcriptional regulatory capabilities of tRFs, we utilized the results of a large-scale screen of short cis-regulatory elements (8-mers) in HEK293-FLP cells [106]. This study used a cell-based assay to measure the expression of a GFP-reporter as a readout for the regulatory potential of an 8-mer inserted within the human IQGAP1 3' UTR. To reveal activating, repressing or elements with no effect, we calculated frequencies of occurrence for all such 8-mers in Ago1-loaded RNA targets targeted by tRF and miRNA guide sequences.

DISCUSSION

This dissertation is focused on the targeting modes of tRNA fragments. We have identified and studied tRFs across three species and we report consistent results that indicate but not limited to “seed” driven targeting mode. We initially found tRFs bound to Ago proteins in *Drosophila*. Our study (chapter 1) was one of the first to report tRFs as active components of RISC complexes and we observed a great number of similarities between miRNAs and tRFs with regards to their loading patterns with regards to the process of aging. Following our work on *Drosophila*, we sought to examine tRFs and their contribution to aging in rat brains. In chapter 2 we report our results for rat tRFs where we observed consistent results to what we previously saw in *Drosophila* for tRF-3 tRFs. There was a consistent increase with age and our “seed” identification approach based on conserved 7-mer matches between tRFs and 3’ UTRs we found “seed” sequences of tRFs that can potentially reside on the 5’ or 3’ end of a tRF. In chapter 3, we took advantage of a new sequencing method that captures in vivo small RNAs and their Ago-associated targets to utilize an ab initio approach. We found that tRFs can target a variety of RNAs both in cytoplasmic and nuclear compartments. Similar to miRNAs, the primary targets of tRFs are 3’ UTR regions of protein coding genes. With regards to the targeting modes of tRFs, we found that tRF-3 tRFs primarily utilize a 5’ located “seed” element with the 3’ showing a compensatory role. However, we report that such “seed” elements are actually not restricted to 7 nucleotides. Finally, for three out of six candidate tRF-3 from

our results there is experimental validation from an independent study showing a 5' "seed" sequence[20].

tRFs in aging *Drosophila* brain

We identified tRFs in both Ago1 and Ago2 co-immunoprecipitated libraries, indicating miRNA-like functionality of loading of these tRFs into RISC complexes. Alignment to the mature tRNA sequence revealed a high read-depth on one side of the tRNA molecule and size distributions of 16-30 base pairs in length, which suggests a similar structural motif as miRNAs. By examining age-associated patterns of tRF expression, we saw distinct isoforms changes in age-dependent manner in *Drosophila*. One possible explanation proposed for the observations of differential miRNA loading with age (which can be extended to tRFs) is that the cells are adjusting their regulatory processes for upcoming age-associated stresses [47].

Other modes of tRF-driven regulation have been proposed, from inhibiting translation initiation factors to direct interaction with ribosome, etc [7, 10, 21, 25, 37, 40, 41]. Given the base pairing in the tRNA stems, one cannot exclude potential interaction with full-length host tRNAs or their fragments. While this paper was under review, a possible role of tRFs as tumor suppressors binding to oncogenic RNA-binding protein YBX1, displacing

pro-oncogenic transcripts has been described [61]. However, the patterns of conservations we observed indicate a clear possibility of miRNA-like targeting.

Although the exact mechanism is still being unraveled, our results suggest a short seed region in tRFs that is key for recognizing potential mRNA targets. While for animal miRNAs the 5' seed location is most common, 3'-compensatory sites [52] and central pairing sites [53] have been reported. In our examples, the Gly-associated tRF in *Drosophila* has a putative 3' seed region, while the mt:SerGCT tRF has a 5' seed. Thus, in parallel to experimental data showing two possible seed locations [22, 42, 46], our results demonstrate that regions of conservation can be present at either the 5' or the 3' end in different tRFs. We also provide evidence that the 3' UTR may be where targeting occurs, allowing us to speculate that the mode of action may include translational repression or mRNA cleavage.

Alternatively, some tRFs may employ mRNA cleavage for regulation, since we observed CDS regions that also aligned to our candidate seeds [43-45]. Enrichment of seed matches in the conserved intron regions may also indicate a role of tRFs in alternative splicing and transcriptional regulation, given the evidence of Ago2 involvement in these process in the nucleus [62]. The enrichment of targets involved in development may be of particular interest in this regard as Ago2 transcriptional target genes are also bound by Polycomb group transcriptional repressor proteins and change during development [62].

Drosophila Ago1 and Ago2 employ different mechanisms to silence target mRNAs and in particular Ago2 mutants show neurodegeneration and a shortened lifespan [47]. The fact that most tRFs are loaded and/or show a dramatic change in loading with age in Ago2 suggests that these small RNAs may also be involved in such pathways. In this regard, it is notable that despite the difference in seed localization (and no common targets), putative targets of tRFs from both mt:SerGCT and GlyCTC are significantly enriched in developmental and neuronal functions. Further, we found that these target lists overlap (with up to 29 targets) with the well-studied miRNAs mir-34, mir-277, mir-190, and mir-10. All of these miRNAs impact brain function, affecting neurodegeneration, bi-polar disorder, and schizophrenia [35, 63, 64], in agreement with our predictions of tRF influence on the brain and age-related events. An overlap of the tRF seed with that of mir-277 is of importance, as it may relate one of the most abundant tRFs (GlyGCC) to brain deterioration, since mir-277 has been reported to modulate neurodegeneration [65].

tRFs in aging mammalian brain

We observed two typical patterns of change in tRF levels in aging mammalian brain. One was a monotonous increase with age, primarily seen in 3' tRFs. Another was a lower abundance in mid-aged rat brains and higher abundance in young and old animals, mostly

observed in 5' tRFs. These patterns, together with the differences in fragment sizes suggest distinct mechanisms of cleavage for the two types of fragments, which can potentially be attributed to the different roles for these two types of tRFs. In addition to the biogenesis pathways, tRFs originating from different ends of the tRNA molecule have also been shown to localize in different sub-cellular compartments. As pointed out by Kumar et al [13], 5' tRFs were equally abundant in the nuclei and whole cell fraction of HeLa cell line [87] indicating primarily nuclear localization and consistent with large numbers of 5' tRFs in HeLa cell nucleoli [3]. On the contrary, 3' tRFs showed an enrichment in the whole cell fraction indicating their cytoplasmic localization in agreement with Haussecker et al [4]. There has been evidence of miRNAs actively loaded to Argonaute proteins in an age-dependent manner in *D. melanogaster* [47]. A very similar age-related loading pattern was also observed for *D. melanogaster* tRFs [32]. This, along with extensive evidence that Argonaute proteins are not only acting in post-transcriptional silencing but are localized/imported to the nucleus, could imply additional unknown functions for tRFs within the nuclear compartments of the cell. Perhaps, such functions are similar to those previously described for miRNAs, which have been shown to be associated with mRNA splicing and modulation of histone epigenetic modifications [88, 89], and this is a focus of our ongoing research.

Although the mode of action for tRFs is yet to be elucidated, our results support the hypothesis that mammalian tRFs (at least, 3' tRFs) can act in a very similar way to miRNAs in post-transcriptional gene silencing. We show here that they contain 7mers, which

match 3' UTR regions of transcripts at much higher rate than expected by chance, similar to the seed sequences of miRNAs. Searching for conserved matches across vertebrate genomes, we found such seeds on either end of the tRF molecules, as has been the case with 12 *Drosophila* species [32]. Previous studies have also detected both 5' and 3' seeds in different tRFs and changes in the seed sequence have been shown to affect the suppression of mRNA translation [46, 79]. It is worth noting that in miRNAs, 3'-compensatory sites [52] and central pairing sites [53] have been reported in addition to the most prevalent 5' seeds [42-45], thus finding seeds on both ends of tRFs is not unexpected. Non-traditional seed region location in miRNA is also consistent with the extensive results of Helwak et al [90], who reported that more than half of the observed miRNA-mRNA interactions do not fulfill traditional seed binding properties in HEK-293 cells. However, one cannot exclude other modes of action, for example, tRFs have been reported binding to oncogenic RNA-binding protein YBX1, displacing pro-oncogenic transcripts and acting as tumor suppressors [61].

Interestingly, for tRFs with clearly defined seed-like regions, we observed a significant and consistent enrichment for targeted genes related to neuronal function and development in Gene Ontology terms. Again, this was in agreement with a functional enrichment seen in *Drosophila* tRF targets [32]. However, in addition to these functions, rat brain tRFs also appeared to target transcription and splicing regulators, in parallel to earlier findings for rat brain miRNAs [86].

Targeting modes of Ago1 loaded tRFs

In our latest work for HEK-293 cells, we show that levels of tRFs loaded to Ago1 do not correlate with the expression of the parental tRNA gene. We also examined tRFs from total cell fraction and we show that there is no correlation between the amount of tRFs produced by a tRNA gene and the abundance of the same tRNA. Our findings are concordant with the results from separate studies [114-116]. Consistent lack of correlation between tRNA and tRF levels supports the view that tRFs are not products of random degradation and different types of tRFs may be produced through different mechanisms.

We considered in detail the type of tRF targets and observed among them non-coding RNAs in addition to mRNAs. We show that mRNAs are the most frequent targets for tRFs loaded to Ago1. Within mRNA targets, we observed that 3' UTRs were the most frequently targeted regions overall followed by coding sequences. 5' UTRs were found to be the least frequently targeted region of mRNAs for all types of tRFs. Our results further support the hypothesis that tRFs can interact with UTR and CDS regions of mRNAs similar to miRNAs [107, 108]. Ribosomal protein genes RPL35A, RPS14 and RPL7L1 were among the most frequent coding targets of Ago loaded tRFs, suggesting that tRFs may be involved in global translational regulation in addition to posttranscriptional of specific targets, similar to a hypothesis put forth in a recent *Drosophila* study [117].

Recent evidence with regards to nuclear localization of Ago proteins [101, 118] and tRF-5 [13] prompted us to consider intronic regions. We found that tRFs interact with introns in pre-mRNAs (chapter 3, Fig. 3). It is possible that tRFs may be utilized by Ago in order to guide or regulate splicing. Ago proteins in complexes with miRNAs and siRNAs have been shown to be actively involved in transcriptional regulation and pre-mRNA splicing [101, 118, 120]. Interestingly, we found that tRFs may be involved in guiding Ago to the 5' end of short introns, recently classified as agotrons [102]. Agotrons are identified based on their length (< 150 nts) and interaction with Ago2 in the first 30 nts of their 5' end. We found that tRFs may guide Ago1 to the borders of agotrons and therefore tRFs are likely involved in agotron biogenesis with Ago1 and Ago2.

We report that tRFs also target miRNAs when loaded to Ago1. A previous study revealed that Ago-loaded miRNAs can target tRFs [90]. Taken together, these findings suggest that tRFs and miRNAs display a similar behavior when loaded to RISC complexes. Both classes of small RNAs can regulate each other in addition to their respective coding or non-coding targets.

A small fraction (9%) of Ago1 tRF chimeras showed a 7-mer match between guide tRF and target RNA. In light of these results, we used an *ab initio* approach to identify core interaction sites across tRFs. We report that for seven tRFs, we were able to identify a conserved motif likely corresponding to a core interaction site. We identified five core interaction sites close to the 5' end of the guide tRF and two sites were found near the 3' end of the tRF. Such 5' and 3' seeds have been previously reported for tRFs in both

experimental and computational studies [32, 46, 79], however, seeds located on the 3' end of a tRF appear to be less frequent, similar to what we observed for rat tRFs [33]. And while the additional binding of 3' located nucleotides might not necessarily be required for suppression of target RNAs, it is observed quite frequently for Ago1 tRF-3 chimeras. Notably, we observed the strongest interactions between tRFs and target RNAs for chimeras with extensive 5' hybridization pattern followed by additional hybridizing nucleotides on the 3' end of the tRF (chapter 3, Fig. 4, c5).

Our results were in agreement with the experiment for all three tRFs, for which the 5' seeds have been recently validated [20]. However, we also observed slight variations with the 7-mer seeds proposed in that work. For distinct Cys tRF isoforms we saw a consistent core interaction site, which includes the motif GGGNACC (chapter 3, Table 1), but the longest motif extended further towards the 3' end of a tRF chapter 3, (Fig. 5). Additionally, for Leu-AAG-001 we observed that the enriched core motif was present both at the 5' and 3' end of the tRF (underlined in (chapter 3, Fig. 5). It is worth noting that when taking into account all available chimeras formed between tRF-3 Leu-AAG-001 and RNA targets, the binding frequency of specific nucleotides and the MFE for interactions favor the 5'vs. 3' hybridization mode of this tRF.

Future Directions

Over the past four years we have developed a thorough pipeline to identify core/seed interaction sites for tRFs. Similar methodology can also be applied to other, less documented small RNAs that are found bound to Ago proteins. Our pipeline can be further be improved by analyzing more data and alternative sequencing methods such as PAR-CLIP. Given enough data, we could use machine learning to train on well know interactions between small RNAs and their targets and design a more comprehensive target prediction pipeline applicable to various types of small RNAs that are found loaded to RISC complexes.

References

1. Abe, T., et al., *tRNADB-CE: tRNA gene database well-timed in the era of big sequence data*. 2014. **5**(114).
2. Chan, P.P. and T.M. Lowe, *GtRNadb: a database of transfer RNA genes detected in genomic sequence*. Nucleic Acids Res, 2009. **37**(Database issue): p. D93-7.
3. Cole, C., et al., *Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs*. Rna, 2009. **15**(12): p. 2147-60.
4. Haussecker, D., et al., *Human tRNA-derived small RNAs in the global regulation of RNA silencing*. RNA, 2010. **16**(4): p. 673-95.
5. Lee, Y.S., et al., *A novel class of small RNAs: tRNA-derived RNA fragments (tRFs)*. Genes Dev, 2009. **23**(22): p. 2639-49.
6. Levitz, R., et al., *The optional E.coli prr locus encodes a latent form of phage T4-induced anticodon nuclease*. EMBO Journal, 1990. **9**(5): p. 1383-1389.
7. Li, Y., et al., *Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote Giardia lamblia*. Nucleic Acids Res, 2008. **36**(19): p. 6048-55.
8. Yeung, M.L., et al., *Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid*. Nucleic Acids Res, 2009. **37**(19): p. 6575-86.
9. Thompson, D.M. and R. Parker, *The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae*. J Cell Biol, 2009. **185**(1): p. 43-50.
10. Tuck, A.C. and D. Tollervey, *RNA in pieces*. Trends Genet, 2011. **27**(10): p. 422-32.
11. Fu, H., et al., *Stress induces tRNA cleavage by angiogenin in mammalian cells*. FEBS Letters, 2009. **583**(2): p. 437-442.
12. Emara, M.M., et al., *Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly*. J Biol Chem, 2010. **285**(14): p. 10959-68.
13. Kumar, P., et al., *Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets*. BMC Biol, 2014. **12**(1): p. 78.
14. Pekarsky, Y., et al., *Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer*. Proc Natl Acad Sci U S A, 2016. **113**(18): p. 5071-6.

15. Pliatsika, V., et al., *MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments*. Bioinformatics, 2016. **32**(16): p. 2481-9.
16. Telonis, A.G., et al., *Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies*. Oncotarget, 2015. **6**(28): p. 24797-822.
17. Babiarz, J.E., et al., *Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs*. Genes Dev, 2008. **22**(20): p. 2773-85.
18. Kumar, V., et al., *Syndapin promotes formation of a postsynaptic membrane system in Drosophila*. Mol Biol Cell, 2009. **20**(8): p. 2254-64.
19. Li, Z., et al., *Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs*. Nucleic Acids Res, 2012. **40**(14): p. 6787-99.
20. Kuscu, C., et al., *tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner*. 2018. **24**(8): p. 1093-1105.
21. Gebetsberger, J., et al., *tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii*. Archaea, 2012. **2012**: p. 260909.
22. Miyoshi, K., T. Miyoshi, and H. Siomi, *Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production*. Mol Genet Genomics, 2010. **284**(2): p. 95-103.
23. Miyoshi, K., et al., *Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production*. 2010. **284**(2): p. 95-103.
24. Haiser, H.J., et al., *Developmentally regulated cleavage of tRNAs in the bacterium Streptomyces coelicolor*. Nucleic Acids Res, 2008. **36**(3): p. 732-41.
25. Fischer, S., et al., *Regulatory RNAs in Haloferax volcanii*. Biochem Soc Trans, 2011. **39**(1): p. 159-62.
26. Peng, H., et al., *A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm*. Cell Res, 2012. **22**(11): p. 1609-12.
27. Wei, C., et al., *Transcriptome-wide analysis of small RNA expression in early zebrafish development*. RNA, 2012. **18**(5): p. 915-29.
28. Pincus, Z. and F.J. Slack, *Transcriptional (dys)regulation and aging in Caenorhabditis elegans*. Genome Biol, 2008. **9**(9): p. 233.
29. de Lencastre, A., et al., *MicroRNAs both promote and antagonize longevity in C. elegans*. Curr Biol, 2010. **20**(24): p. 2159-68.

30. Ibáñez-Ventoso, C., et al., *Modulated microRNA expression during adult lifespan in*. Vol. 5. 2006. 235-46.
31. Kato, M., et al., *Age-associated changes in expression of small, noncoding RNAs, including microRNAs, in C. elegans*. RNA, 2011. **17**(10): p. 1804-20.
32. Karaïskos, S., et al., *Age-driven modulation of tRNA-derived fragments in Drosophila and their potential targets*. Biology Direct, 2015. **10**(1): p. 51.
33. Karaïskos, S. and A. Grigoriev, *Dynamics of tRNA fragments and their targets in aging mammalian brain*. F1000Res, 2016. **5**.
34. Karp, X., et al., *Effect of life history on microRNA expression during C. elegans development*. Rna, 2011. **17**(4): p. 639-51.
35. Liu, N., et al., *The exoribonuclease Nibbler controls 3' end processing of microRNAs in Drosophila*. Curr Biol, 2011. **21**(22): p. 1888-93.
36. Gong, B., et al., *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*. BMC Infect Dis, 2013. **13**: p. 285.
37. Sobala, A. and G. Hutvagner, *Transfer RNA-derived fragments: origins, processing, and functions*. Wiley Interdiscip Rev RNA, 2007. **2**(6): p. 853-62.
38. Anderson, P. and P. Ivanov, *tRNA fragments in human health and disease*. FEBS Lett, 2014: p. 4297-4304.
39. Loss-Morais, G., P.M. Waterhouse, and R. Margis, *Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets*. Biol Direct, 2014. **8**: p. 6.
40. Garcia-Silva, M.R., et al., *Hints of tRNA-Derived Small RNAs Role in RNA Silencing Mechanisms*. Genes (Basel), 2012. **3**(4): p. 603-14.
41. Ivanov, P., et al., *Angiogenin-induced tRNA fragments inhibit translation initiation*. Mol Cell, 2011. **43**(4): p. 613-23.
42. Wang, Q., et al., *Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection*. Mol Ther, 2012. **21**(2): p. 368-79.
43. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing*. Mol Cell, 2007. **27**(1): p. 91-105.
44. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets*. Cell, 2005. **120**(1): p. 15-20.

45. Lewis, B.P., et al., *Prediction of mammalian microRNA targets*. Cell, 2003. **115**(7): p. 787-98.
46. Maute, R.L., et al., *tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma*. Proc Natl Acad Sci U S A, 2013. **110**(4): p. 1404-9.
47. Abe, M., et al., *Impact of age-associated increase in 2'-O-methylation of miRNAs on aging and neurodegeneration in Drosophila*. Genes Dev, 2014. **28**(1): p. 44-57.
48. Chen, C.J., et al., *Genome-wide discovery and analysis of microRNAs and other small RNAs from rice embryogenic callus*. RNA Biol, 2011. **8**(3): p. 538-47.
49. Ghildiyal, M., et al., *Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway*. RNA, 2009. **16**(1): p. 43-56.
50. Kim, V.N., *MicroRNA biogenesis: coordinated cropping and dicing*. Nat Rev Mol Cell Biol, 2005. **6**(5): p. 376-85.
51. Jochl, C., et al., *Small ncRNA transcriptome analysis from Aspergillus fumigatus suggests a novel mechanism for regulation of protein synthesis*. Nucleic Acids Res, 2008. **36**(8): p. 2677-89.
52. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions*. Cell, 2009. **136**(2): p. 215-233.
53. Shin, C., et al., *Expanding the microRNA targeting code: functional sites with centered pairing*. Molecular cell, 2010. **38**(6): p. 789-802.
54. Clark, A.G., et al., *Evolution of genes and genomes on the Drosophila phylogeny*. Nature, 2007. **450**(7167): p. 203-18.
55. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*. Nat Genet, 2000. **25**(1): p. 25-9.
56. Prakash, S., et al., *Complex interactions amongst N-cadherin, DLAR, and Liprin-alpha regulate Drosophila photoreceptor axon targeting*. Dev Biol, 2009. **336**(1): p. 10-9.
57. Krueger, N.X., et al., *The transmembrane tyrosine phosphatase DLAR controls motor axon guidance in Drosophila*. Cell, 1996. **84**(4): p. 611-22.
58. Smibert, P., et al., *Global patterns of tissue-specific alternative polyadenylation in Drosophila*. Cell reports, 2012. **1**(3): p. 277-289.
59. Iwasaki, S., T. Kawamata, and Y. Tomari, *Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression*. Mol Cell, 2009. **34**(1): p. 58-67.

60. Bai, H., et al., *Activin signaling targeted by insulin/dFOXO regulates aging and muscle proteostasis in Drosophila*. PLoS Genet, 2013. **9**(11): p. e1003941.
61. Goodarzi, H., et al., *Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement*. Cell, 2015. **161**(4): p. 790-802.
62. Taliaferro, J.M., et al., *Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression*. Genes & development, 2013. **27**(4): p. 378-389.
63. Moreau, M.P., et al., *Altered microRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder*. Biol Psychiatry, 2011. **69**(2): p. 188-93.
64. Perkins, D.O., et al., *microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder*. Genome Biol, 2007. **8**(2): p. R27.
65. Tan, H., et al., *MicroRNA-277 modulates the neurodegeneration caused by Fragile X premutation rCGG repeats*. PLoS Genet, 2012. **8**(5): p. e1002681.
66. Zhang, Y., et al., *PAR-1 kinase phosphorylates Dlg and regulates its postsynaptic targeting at the Drosophila neuromuscular junction*. Neuron, 2007. **53**(2): p. 201-15.
67. Azim, A.C., et al., *DLG1: chromosome location of the closest human homologue of the Drosophila discs large tumor suppressor gene*. Genomics, 1995. **30**(3): p. 613-6.
68. McIlroy, G., et al., *Toll-6 and Toll-7 function as neurotrophin receptors in the Drosophila melanogaster CNS*. Nat Neurosci, 2013. **16**(9): p. 1248-56.
69. Garber, K., et al., *Transcription, translation and fragile X syndrome*. Curr Opin Genet Dev, 2006. **16**(3): p. 270-5.
70. Jin, P., et al., *Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway*. Nat Neurosci, 2004. **7**(2): p. 113-7.
71. Ishizuka, A., M.C. Siomi, and H. Siomi, *A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins*. Genes Dev, 2002. **16**(19): p. 2497-508.
72. Rosenbloom, K.R., et al., *The UCSC Genome Browser database: 2015 update*. Nucleic Acids Res, 2014.
73. Naqvi, A., T. Cui, and A. Grigoriev, *Visualization of nucleotide substitutions in the (micro)transcriptome*. BMC Genomics, 2014. **15 Suppl 4**: p. S9.
74. Garcia-Silva, M.R., et al., *Hints of tRNA-Derived Small RNAs Role in RNA Silencing Mechanisms*. Genes (Basel), 2012. **3**(4): p. 603-14.

75. Garcia-Silva, M.R., et al., *Novel aspects of tRNA-derived small RNAs with potential impact in infectious diseases*. Advances in Bioscience and Biotechnology, 2013. **4**: p. 17-25.
76. Ogawa, T., et al., *A cytotoxic ribonuclease targeting specific transfer RNA anticodons*. Science, 1999. **283**(5410): p. 2097-100.
77. Gebetsberger, J. and N. Polacek, *Slicing tRNAs to boost functional ncRNA diversity*. RNA Biol, 2013. **10**(12): p. 1798-806.
78. Martens-Uzunova, E.S., M. Olvedy, and G. Jenster, *Beyond microRNA--novel RNAs derived from small non-coding RNA and their implication in cancer*. Cancer Lett, 2013. **340**(2): p. 201-11.
79. Wang, Q., et al., *Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection*. Mol Ther, 2013. **21**(2): p. 368-79.
80. Kirkwood, T.B., *A systematic look at an old problem*. Nature, 2008. **451**(7179): p. 644-647.
81. Yin, L., et al., *Discovering novel microRNAs and age-related nonlinear changes in rat brains using deep sequencing*. Neurobiology of aging, 2015. **36**(2): p. 1037-1044.
82. Bühler, M., et al., *TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway*. Nature structural & molecular biology, 2008. **15**(10): p. 1015-1023.
83. Hsieh, L.-C., et al., *Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing*. Plant physiology, 2009. **151**(4): p. 2120-2132.
84. Kawaji, H., et al., *Hidden layers of human small RNAs*. BMC Genomics, 2008. **9**(1): p. 157.
85. Wood, S.H., et al., *Whole transcriptome sequencing of the aging rat brain reveals dynamic RNA changes in the dark matter of the genome*. Age, 2013. **35**(3): p. 763-776.
86. Nielsen, J.A., et al., *Integrating microRNA and mRNA expression profiles of neuronal progenitors to identify regulatory networks underlying the onset of cortical neurogenesis*. BMC Neuroscience, 2009. **10**: p. 98-98.
87. Valen, E., et al., *Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes*. Nature structural & molecular biology, 2011. **18**(9): p. 1075-1082.
88. Huang, V. and L.-C. Li, *Demystifying the nuclear function of Argonaute proteins*. RNA Biol, 2014. **11**(1): p. 18-24.
89. Schraivogel, D. and G. Meister, *Import routes and nuclear functions of Argonaute and other small RNA-silencing proteins*. Trends in Biochemical Sciences, 2014. **39**(9): p. 420-431.

90. Helwak, A., et al., *Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding*. Cell, 2013. **153**(3): p. 654-65.
91. Wetzel-Smith, M.K., et al., *A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death*. Nature medicine, 2014. **20**(12): p. 1452-1457.
92. Redies, C., *Cadherins in the central nervous system*. Progress in neurobiology, 2000. **61**(6): p. 611-648.
93. Ford-Perriss, M., H. Abud, and M. Murphy, *Fibroblast growth factors in the developing central nervous system*. Clinical and Experimental Pharmacology and Physiology, 2001. **28**(7): p. 493-503.
94. Lim, L.P., et al., *Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs*. Nature, 2005. **433**(7027): p. 769-773.
95. Mi, H., et al., *Large-scale gene function analysis with the PANTHER classification system*. Nature protocols, 2013. **8**(8): p. 1551-1566.
96. Yamasaki, S., et al., *Angiogenin cleaves tRNA and promotes stress-induced translational repression*. J Cell Biol, 2009. **185**(1): p. 35-42.
97. Saikia, M., et al., *Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress*. J Biol Chem, 2012. **287**(51): p. 42708-25.
98. Shigematsu, M., S. Honda, and Y. Kirino, *Transfer RNA as a source of small functional RNA*. J Mol Biol Mol Imaging, 2014. **1**(2).
99. Anderson, P. and P. Ivanov, *tRNA fragments in human health and disease*. FEBS Lett, 2014. **588**(23): p. 4297-304.
100. Shin, C., et al., *Expanding the microRNA targeting code: functional sites with centered pairing*. Mol Cell, 2010. **38**(6): p. 789-802.
101. Taliaferro, J.M., et al., *Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression*. Genes Dev, 2013. **27**(4): p. 378-89.
102. Hansen, T.B., et al., *Argonaute-associated short introns are a novel class of gene regulators*. Nat Commun, 2016. **7**: p. 11538.
103. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
104. Krüger, J. and M. Rehmsmeier, *RNAhybrid: microRNA target prediction easy, fast and flexible*. Nucleic Acids Research, 2006. **34**(suppl_2): p. W451-W454.

105. Gogakos, T., et al., *Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP*. Cell Reports, 2017. **20**(6): p. 1463-1475.
106. Wissink, E.M., E.A. Fogarty, and A. Grimson, *High-throughput discovery of post-transcriptional cis-regulatory elements*. BMC Genomics, 2016. **17**: p. 177.
107. Hausser, J., et al., *Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation*. Genome Res, 2013. **23**(4): p. 604-15.
108. Lee, I., et al., *New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites*. Genome Res, 2009. **19**(7): p. 1175-83.
109. Khorkova, O., et al., *Natural antisense transcripts*. Hum Mol Genet, 2014. **23**(R1): p. R54-63.
110. Werner, A. and D. Swan, *What are natural antisense transcripts good for?* Biochem Soc Trans, 2010. **38**(4): p. 1144-9.
111. Wight, M. and A. Werner, *The functions of natural antisense transcripts*. Essays Biochem, 2013. **54**: p. 91-101.
112. Bailey, T.L., et al., *MEME Suite: tools for motif discovery and searching*. Nucleic Acids Research, 2009. **37**(suppl_2): p. W202-W208.
113. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif*. Bioinformatics, 2011. **27**(7): p. 1017-1018.
114. Kawaji, H., et al., *Hidden layers of human small RNAs*. BMC Genomics, 2008. **9**: p. 157.
115. Mahlab, S., T. Tuller, and M. Linial, *Conservation of the relative tRNA composition in healthy and cancerous tissues*. Rna, 2012. **18**(4): p. 640-52.
116. Martens-Uzunova, E.S., M. Olvedy, and G. Jenster, *Beyond microRNA – Novel RNAs derived from small non-coding RNA and their implication in cancer*. Cancer Letters, 2013. **340**(2): p. 201-211.
117. Luo, S., et al., *Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response*. Nucleic Acids Research, 2018. **46**(10): p. 5250-5268.
118. Ameyar-Zazoua, M., et al., *Argonaute proteins couple chromatin silencing to alternative splicing*. Nature Structural Molecular Biology, 2012. **19**: p. 998.
119. Catalanotto, C., C. Cogoni, and G. Zardo, *MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions*. International Journal of Molecular Medicine 2016. **17**(10): p. 1712.

120. Allo, M., et al., *Control of alternative splicing through siRNA-mediated transcriptional gene silencing*. Nature Structural Molecular Biology, 2009. **16** p.717-24.
121. Jühling, F., et al., *tRNADB 2009: compilation of tRNA sequences and tRNA genes*. Nucleic Acids Res, 2009. **37**(Database issue): p. D159-62.
122. Aken, B.L., et al., *Ensembl 2017*. Nucleic Acids Research, 2017. **45**(D1): p. D635-D642.