UNIVERSITY OF CALIFORNIA
RIVERSIDE

An Exploration of RNA Modification, Structure and Functionality of tRNA-Derived Small RNAs

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Cell, Molecular and Developmental Biology

by

Xudong Zhang

December 2021

Dissertation Committee:
Dr. Qi Chen, Chairperson
Dr. Xuemei Chen
Dr. Yinsheng Wang
Dr. Jernej Murn
Dr. Shenglong Zhang

The Dissertation of Xudong Zhang is approved:

_____

_____

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

First and foremost, I would like to thank the Chen lab. Dr. Qi Chen, a passionate and thoughtful explorer, not only guided me on the path to truth but also inspired me to be a scientist with persistent curiosity. Thanks for the mentorship over the years, it has been invaluable. And thanks to all the members of Chen lab, including Dr. Shichao Liu, Dr. Junchao Shi, Dr. Bailong Zhang, Dr. Yunfang Zhang, Dr. Menghong Yan, Dr. Ying Zhang, without your help I couldn't finish this great project. I also would like to thank Dr. Tong Zhou for the great help on the bioinformatic analysis of the data presented in this dissertation. In addition, thanks for the hospitality for inviting me to your house to celebrate my first Christmas holiday in the USA, that's a truly memorable occasion.

I would like to thank my committee members, Dr. Xuemei Chen, Dr. Yinsheng Wang, Dr. Jernej Murn and Dr. Shenglong Zhang. Thanks for your guidance. I benefited so much from your advice on the experimental design and the valuable comments to the data. Special thanks to Dr. Shenglong Zhang and his lab for sharing the MLC-seq method with our lab, the data in Chapter 4 was obtained by the MLC-seq most recently before the dissertation finished.

I would like to thank Julio Sosa, the student service advisor of CMDB program, who helped me to solve all the issues I meet regarding to course enrollment, credits transfer and graduation requirements *et al*.

I would like to thank all the friends I have made so far. Thanks for always being there. Thanks to Dr. Lin Chen, my high school classmate, thanks for your phone call when I am downhearted. Thanks to Yannan Hu, a truly cat lover, for sharing her

experience for cat caring and TA service. Thanks for Hasi and Chengcheng Ji, thanks for your tea from China, wish one day you two can setup a funding to support scientific research. I also would like to thank my cat Chujiu, who make me become a more responsible and reliable person.

Lastly, I would like to thank my parent, my elder brother and sister. Even though you guys didn't completely understand what I was doing and how it could be useful in the future, but you just let me do the thing I need to do to continue my Ph.D. career. Without your support, I wouldn’t have been able to make it so far. It’s so lucky for me to have you all as a family.

ABSTRACT OF THE DISSERTATION

An Exploration of RNA Modification, Structure and Functionality of tRNA-Derived Small RNAs

by

Xudong Zhang

Doctor of Philosophy, Graduate Program in Cell, Molecular and Developmental Biology
University of California, Riverside, December 2021
Dr. Qi Chen, Chairperson

tRNA-derived small RNAs (tsRNAs) are a type of widely existed small non-coding RNA (sncRNA) with their mechanisms of action little explored. tsRNAs are sophisticatedly modified, creating a challenge to study their biological function because synthetic tsRNAs may not mimic their structure and interacting potential. Here, we show examples how endogenous tsRNAs with extensive RNA modifications can have cellular functions different from synthetic ones that without modifications. By adding site-specific RNA modifications to engineer synthetic tsRNAs, we mimic the endogenous tsRNAs structure conformation and recapitulate the functional effects of endogenous tsRNAs. In addition, we applied a LC-MS based *de novo* RNA sequencing method (MLC-seq) to profile RNA modification of endogenous tRNAs or tsRNAs. Our efforts may guide future understanding and designing of modified tsRNAs for precise gene regulation with a comprehensive tsRNA modification profile identified.

Table of Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

List of Tables

Chapter 1

Introduction

1.1 Introduction to tRNA and tsRNA

Transfer RNAs (tRNAs) are small non-coding RNAs (sncRNAs) that serve as physical link between the messenger RNAs (mRNA) and amino acid sequence of peptide chains during protein synthesis, which make tRNA become the decoder of genetic information embedded in the mRNAs. Mature tRNAs are being processed from pre-tRNA which is transcribed by RNA polymerase III. Before mature tRNAs are produced, pre-tRNAs undergo serval process that includes removal of 5′ leader and 3′ trailer sequence, RNA splicing, addition of cytidine-cytidine-adenosine (CCA) sequence at the 3′ end sequence and post-transcriptional modifications (Schimmel 2018). The sequence of mature tRNA is highly conserved among different species thus mature tRNA folds into a conserved L-shape 3D structure which is curial for the translation process. Moreover, tRNAs are highly modified that each tRNA molecule has about 13 modifications on average (Su et al. 2020). For decades, tRNAs are considered relative stable and cannot be easily fragmented. In 1970s, there are studies reported that there are tRNA degradation products in the urine of cancer patients. That's the first discovery of tRNA-derived small RNAs(tsRNAs) (Borek et al. 1977, Speer et al. 1979). However, at that time, these RNAs are just considered as byproducts of tRNA degradation and didn't catch the attention of research community. Until a decade ago, with the progress in next-generation sequencing technologies, tsRNAs (also called tDRs or tRNA-derived fragments, tRFs) are increasingly recognized as an emerging class of functional sncRNAs in various fundamental biological and disease conditions (Shi et al. 2019, Su et al. 2020).

1.2 tsRNA biogenesis: from degradation to regulation

From a structural perspective, the cloverleaf-shaped secondary structure of tRNA is folded into an L-shape in 3D (Figure 1.1A) (Schimmel 2018). This L-shaped structure is overall tightly condensed but has two relatively exposed sites: the anticodon at one end of the L and the tRNA elbow at the bending site of the L, where the D-loop and the T-loop meet and interact with each other. The exposed sites of the tRNA structure could be 'points of attack' in an ancient cellular (and perhaps early proto cell) environment, being fragmented by either nonspecific stress signals such as radiation and reactive oxygen species (ROS), specific recognition by enzymes or ribozymes, or a combination of both. This simple view coincides with prevailing observations that the most abundantly detected tsRNAs are fragmented at the anticodon, and are derived from the 5′ half of the tRNA (~30 nt), whereas shorter 3′ or 5′ tsRNAs (~18–22 nt) fragmented at the T-loop or D-loop, respectively, or internal tsRNAs derived from sequences between these loops, are less abundant (Kumar et al. 2015, Shi et al. 2021) (Figure 1.1A).

This observation may support the assumption that, in early life forms, the biogenesis of tsRNAs directly originated from tRNA degradation processes starting with these loops, perhaps including multistep degradation and the generation of different intermediates (Peng et al. 2012), accompanied by regulatory elements such as RNA modifications (Figure 1.1B) and specific RNases targeting these loops (Figure 1.1C) that emerged during evolution.

1.2.1 RNases response for tsRNA biogenesis

For the most abundant 5′ half or 3′ half tsRNAs(30~40nt) which is fragmented at anticodon loop, they are processed by the PrrC nuclease in *Escherichia coli* (Levitz et al. 1990), Angiogenin (ANG, a member of RNase A family) in mammalian cell (Yamasaki et al. 2009), and Rny1 (a member of RNase T2 family) in yeast (Thompson et al. 2009). All these tRNA cleavage by RNase are triggered under the stress states such as bacteriophage infection, oxidative stress. Under normal physiological condition, the ANG doesn't cleave tRNAs into tsRNAs because this RNase is concentrated in cell nucleus (Tsuji et al. 2005) or sequestered by ANG inhibitor (RNH1) in the cytoplasm (Shapiro et al. 1987). Similar mechanism also applied to the Rny1 in yeast. Rny1 is localized to the "lysosome" of yeast under normal states but released to the cytoplasm to generate tsRNAs under stress condition (Thompson and Parker 2009). In addition to the stress-induced biogenesis of tsRNAs via RNase releasing, biogenesis of tsRNAs is also regulated by sex hormones and their receptors. It has been reported that 5′ tsRNAs from $tRNA^{Asp}$ (GUC) and $tRNA^{His}$ (GUG) are expressed in estrogen receptor (ER)-positive breast cancer and androgen receptor (AR)-positive prostate cancer cell lines in a ANG-dependent way when the cancer cells response to hormone-signaling pathways (Honda et al. 2015) (Figure.1.1B).

Shorter tsRNAs (18-22nt) that are cleaved in the D-loop (5′ tsRNA) or T-loop (3′ tsRNA) of a tRNA are similar to the well-known micro RNAs. So, early investigations are focused on the Dicer or the microprocessor complex (Drosha/DGCR8), both are required for the miRNA biogenesis, during the shorter tsRNA biogenesis process. Some

studies identified that serval shorter tsRNAs are indeed processed by Dicer in mouse embryonic stem cells (mESCs) (Babiarz et al. 2008) and human cell lines (Cole et al. 2009, Maute et al. 2013). In contrast, there also studies reported that Dicer is not essential for the biogenesis of tsRNAs (Kumar et al. 2014, Kuscu et al. 2018, Li et al. 2012). This discrepancy suggests that the biogenesis of these shorter tsRNA has various mechanism and are not evolutionary conserved, although it may be also caused by the variations of different experimental approach or platform. And there are more evidence suggesting that shorter tsRNAs are processed by RNase other than the miRNA-processing enzymes. In addition to RNases mentioned above, RNase P and RNase Z (ELAC2) can also process the tRNA into tsRNAs (Kim et al. 2020). Overall, more investigations are required to determine the detailed biogenesis mechanism of different shorter tsRNAs including the RNase involved and the location where tsRNA are processed.

1.2.2 tsRNA termini

RNA cleavage by several classes of self-cleaving ribozymes (Diener 1989) produces RNA fragments with 5′-hydroxyl (5′-OH) and 2′,3′-cyclic phosphate (2′,3′-CP) termini, whereas ancient endonucleases in the protein world, including tRNA splicing endonuclease (Trotta et al. 1997), RNase T2 (Luhtala et al. 2010), RNase L (Donovan et al. 2017), and RNase A (Donovan et al. 2017), also similarly generate 5′-OH and 2′,3′-CP RNAs. Interestingly, the RNases T2 (Andersen et al. 2012, Thompson and Parker 2009) and L (Donovan et al. 2017), and the vertebrate-specific angiogenin (RNase A family) (Fu et al. 2009, Yamasaki et al. 2009), can all cleave tRNAs at the anticodon loop,

resulting in their fragmentation into tsRNAs. Cleavage of tRNAs by these ribozymes and enzymes provides unique differences in tsRNAs from other sncRNAs such as those generated by Dicer, which bear a 5′-phosphate (5′-P) and a 3′-hydroxyl (3′-OH) (Bartel 2018), or piRNAs (and plant miRNAs) that bear 2′-O-methylation at the 3′ terminus as a result of additional enzymatic processing (Ji et al. 2012).

Notably, some tsRNA sequences can be found in the piRNA database, which could be partially due to misannotation or contamination (and not to PIWI-binding) (Tosar et al. 2018). However, some of them may represent genuine piRNAs (i.e., $tRNA^{Asp}$-GUC and $tRNA^{His}$-GUG in Bombyx) because they bind to PIWI and have 3′ terminal 2′-O-methylation but not 2′,3′-CP; these piRNAs are further processed from tsRNAs and are termed 'tRNA-derived piRNAs' (Honda et al. 2017). Although the different sncRNA termini represent the chemical properties of their processing enzymes, they may also have functional consequences. Different termini might be harnessed as intra- or intercellular sorting signals that determine protein-binding potential and subcellular compartmentalization (e.g., cytoplasmic versus nuclear), which is an exciting area that awaits further investigation.

1.2.3 Impact of tRNA modifications

Both tsRNAs and their precursor tRNAs are heavily modified. More than 150 types of modifications are found in tRNAs (Boccaletto et al. 2018), and each tRNA molecule has about 13 modifications on average. Their evolutionary origin remains an interesting question. Hypothetically, in early life forms on Earth, RNA modifications

might have been first added by random chemical reactions or by base-modifying ribozymes (Poudyal et al. 2017, Scheitl et al. 2020), which in turn expanded the functionality of ribozymes. In this regard, the RNA modifications on ancient RNAs, such as tRNAs, rRNAs, and spliceosomal small nuclear RNAs (snRNAs), may represent the vestiges of the RNA world at a time when RNA modifications were maximally exploited to increase the functional diversity of ribozymes, before being replaced by the emergence of functionally more versatile protein-based enzymes. The present-day tRNA modifications contribute to multiple aspects of tRNA function, including stability, amino acid charging, and translational accuracy, as well as to tsRNA biogenesis (Pan 2018, Suzuki 2021).

It has been demonstrated that DNMT2- and NSUN2-dependent addition of a 5-methylcytosine (m^5C) modification to several tRNAs (e.g., $tRNA^{Asp}$, $tRNA^{Val}$, $tRNA^{Gly}$, and $tRNA^{Leu}$) increases tRNA stability in flies and mice, whereas deletion of Dnmt2 and/or Nsun2 abolishes m^5C on these tRNAs, making them more likely to be cleaved into tsRNAs under stress conditions (Schaefer et al. 2010, Tuorto et al. 2012, Zhang et al. 2018). The queuosine (Q) modification by QTRT1 occurs in the wobble anticodon position of several tRNAs ($tRNA^{His}$, $tRNA^{Asn}$, $tRNA^{Tyr}$, and $tRNA^{Asp}$) and protects tRNAs against cleavage into tsRNAs in human HEK293T cells (Wang et al. 2018). Interestingly, recent reports showed that C38 Q-modified tRNA promotes DNMT2-mediated m^5C on C38 of $tRNA^{Asp}$ (Muller et al. 2015, Wang et al. 2018); these discoveries resonate with findings that the establishment of one RNA modification can depend on the existence of another (Barraud et al. 2019). Recent evidence also shows that

deletion of Alkbh1 (Rashad et al. 2020) or Alkbh3 increased the levels of N1-methyladenine (m^1A) in tRNAs (Chen et al. 2019), preventing tRNA cleavage and resulting in less tsRNA production. TRMT10A-mediated N1-methylguanine (m^1G) modification also leads to increased $tRNA^{Gln}$ stability and less production of 5′ $tsRNA^{Gln}$ (Cosentino et al. 2018). Moreover, 2′-O-methylation of C34 in human $tRNA^{Met}$ can prevent site-specific cleavage of $tRNA^{Met}$ by angiogenin and reduce tsRNA production (Vitali et al. 2019) (Figure 1.1B).

In addition to preventing tRNA cleavage, some RNA modifications can also promote tsRNA biogenesis. For example, PUS7-mediated pseudouridine (Ψ) at the U8 position has been shown to affect tsRNA biogenesis in stem cells, where deletion of Pus7 leads to a decreased levels of several types of 5′-tsRNAs with terminal oligoguanine (TOG), suggesting that (Ψ) at the U8 position increases the cleavage of these tRNAs to generate tsRNAs (Guzzi et al. 2018). In another example in yeast, 5-methoxycarbonylmethyl-2-thiouridine (mcm^5S^2) at the anticodon wobble position can promote the cleavage of tRNA into tsRNAs (Lu et al. 2008) (Figure1.1B).

In an intriguing scenario, when bacterial tRNAs ($tRNA^{Asp}$ and $tRNA^{Arg}$) are cleaved in half by ribotoxins that recognize the Q-containing anticodon of tRNA (Ogawa et al. 2006), the newly generated two tsRNA halves can be reunited/repaired into a full tRNA by a Pnkp/Hen1 heterotetramer while adding a 2′-O-methylation to the previous RNA cleavage site (Chan et al. 2009). The added 2′-O-methylation can protect the repaired tRNA from being recut by ribotoxins. This example shows how dynamic RNA modifications are harnessed to balance tRNA stability and tsRNA biogenesis and raises

the possibility that tsRNAs can gain additional modifications after being cleaved from precursor tRNAs.

Importantly, it should be noted that the changes in tRNA modifications not only affect tsRNA biogenesis but may also affect the function of the resulting tsRNAs owing to altered modification status (Zhang et al. 2016). In fact, the RNA modifications of tsRNAs have posed challenges for RNA-seq library preparation and for functional studies of tsRNAs, and this will need to be resolved to permit new waves of tsRNA research.

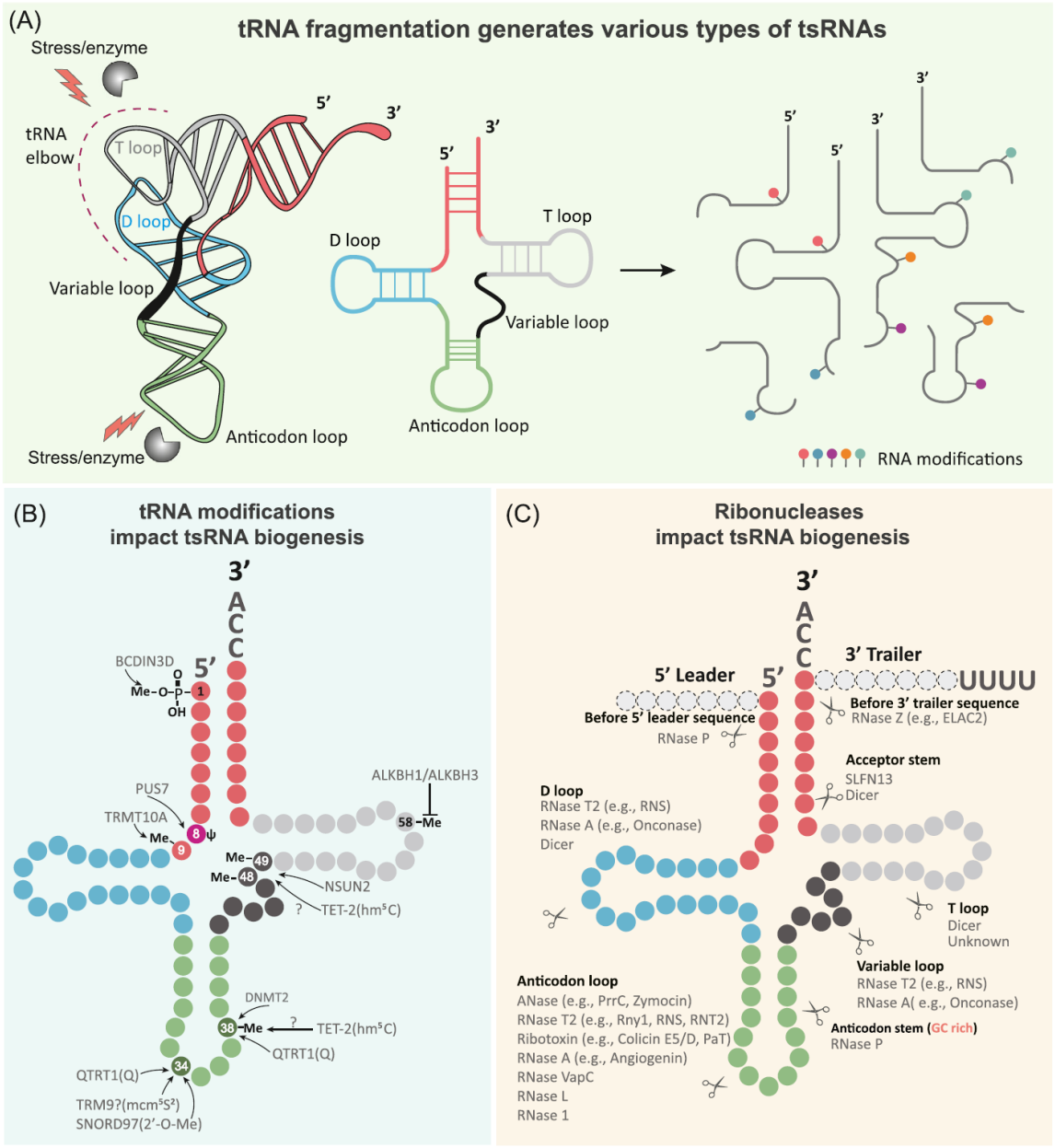

(A)
tRNA fragmentation generates various types of tsRNAs
Stress/enzyme
tRNA elbow
T loop
D loop
Variable loop
Anticodon loop
Stress/enzyme
5′
3′
RNA modifications
(B)
tRNA modifications impact tsRNA biogenesis
BCDIN3D
PUS7
TRMT10A
ALKBH1/ALKBH3
NSUN2
TET-2(hm⁵C)
DNMT2
QTRT1(Q)
TRM9?(mcm⁵S²)
SNORD97(2′-O-Me)
(C)
Ribonucleases impact tsRNA biogenesis
5′ Leader
3′ Trailer
Before 5′ leader sequence
RNase P
Before 3′ trailer sequence
RNase Z (e.g., ELAC2)
Acceptor stem
SLFN13
Dicer
D loop
RNase T2 (e.g., RNS)
RNase A (e.g., Onconase)
Dicer
T loop
Dicer
Unknown
Variable loop
RNase T2 (e.g., RNS)
RNase A(e.g., Onconase)
Anticodon loop
ANase (e.g., PrrC, Zymocin)
RNase T2 (e.g., Rny1, RNS, RNT2)
Ribotoxin (e.g., Colicin E5/D, PaT)
RNase A (e.g., Angiogenin)
RNase VapC
RNase L
RNase 1
Anticodon stem (GC rich)
RNase P

Figure 1.1 The biogenesis of tRNA-derived small RNA (tsRNA) is rooted in tRNA structure, regulated by tRNA modifications and RNases. (A) The 2D and 3D structures of a tRNA, showing the accessible sites at the anticodon loop (green) and the tRNA elbow [the junction of the D- (blue) and T-loops (gray)], which represent the preferred sites of tRNA fragmentation to generate diverse tsRNAs. (B) tRNA modifications and related enzymes that relate to the biogenesis of tsRNAs (TET2, ALKBH). Arrows indicate enzymes that add a specific RNA modification, inhibition lines indicate the effect of demethylation. (C) Currently known RNases for tsRNA biogenesis include Dicer, RNase 1, RNase P, and others. Abbreviations: hm^5C, 5-hydroxymethylcytosine; Me, methylation; Q, queuosine.

1.3 Biological functions of tsRNA

Research on the functions of tsRNAs continues to expand, and roles have been reported in regulating stem cell maintenance (Blanco et al. 2016, Krishna et al. 2019), cancer (Balatti et al. 2017, Goodarzi et al. 2015, Honda et al. 2015), epigenetic inheritance (Chen et al. 2016, Sarker et al. 2019, Zhang et al. 2018), neurological diseases (Hogg et al. 2019, Zhang et al. 2020), viral infection (Nunes et al. 2020, Wang et al. 2013) and symbiosis (Ren et al. 2019).

1.3.1 tsRNA and stem cell maintenance

In both human and mouse stem cells, tsRNAs are abundantly expressed. Stem cells are generally in a low global protein state compared to the differentiated daughter cells they give rise to. In mouse skin stem cells, it has been shown that the low rate of protein synthesis in stem cells is partially maintained by the 5′ tsRNA expressing in the cells which can inhibit cap-dependent protein synthesis (Blanco et al. 2016). The maintenance of stemness via global translation repression by tsRNAs also happened in mouse hematopoietic stem cells and human ESCs (Goncalves et al. 2016, Guzzi et al. 2018). Yet, another report has shown that some tsRNAs derived from specific tRNAs ($tRNA^{Glu}$, $tRNA^{Gly}$, $tRNA^{Gln}$, $tRNA^{Val}$ and $tRNA^{Lys}$) were upregulated during the retinoic acid (RA) induced differentiation of mouse ESCs. These specific tsRNAs can suppress pluripotency-associated genes to facilitate stem cell differentiation, such as cMyc transcripts, however, the exact targeting mechanism is unclear (Krishna et al. 2019). The different observations suggest that the function of tsRNAs is content-dependent.

1.3.2 tsRNA and cancer development

Cancer is caused by over proliferation of malignant cells, and the rapid proliferation of tumorigenic cell often results in the shortage of blood supply which leads to a hypoxic stress environment. As the biogenesis of tsRNAs is closely related to the cellular stress response, it is not surprising that the dysregulation of tsRNAs is identified in various types of cancer including breast cancer, lung cancer, prostate cancer, colorectal cancer *et al*. (Balatti et al. 2017). The role of tsRNAs in cancer development is still controversial. Some evidence suggests that tsRNAs are tumor suppressors in certain cancers by inhibiting proliferation and metastasis, but some reports support that tsRNA can promote tumorigenic cell proliferation. For example, 3′ tsRNA from $tRNA^{Leu}$ can impede colorectal cancer progression by targeting the Notch ligand JAG2 to repress Notch signaling pathway which supports the function of cancer stem-like cells (Huang et al. 2017). In a breast cancer study, it has shown that tsRNAs from the internal part of tRNA can bind to YBX1 protein by displacing 3′ untranslated regions (UTRs) of multiple oncogenic transcripts, thus destabilized theses oncogenic transcripts to suppress the breast cancer progression (Goodarzi et al. 2015). But, in the sex hormone-dependent breast and prostate cancers, short 5′ tsRNA are identified abundantly expressed and can enhance the cancer cell proliferation even though the detailed molecular mechanism is unclear (Honda et al. 2015). In addition, 3′ tsRNAs can also promote cell proliferation and cell cycle in non-small cell lung cancer (Shao et al. 2017). The ambivalent role for tsRNAs in cancer development may explained by their origins from the different tRNA

precursors and the different RNA modification on the tsRNAs, since the RNA modification can substantially impact the function of small RNAs (Zhang et al. 2016).

1.3.3 tsRNA and epigenetic inheritance

It is widely accepted that heritable traits are coded in DNA sequences, however, DNA sequence alone are not sufficient to explain the phenomenon that certain life experience of parents can induce phenotypic change in immediate offspring and sometimes progenies crossing multiple generations without DNA sequence change. This epigenetic inheritance of ancestral life experience is believed to go beyond simple protein-coding DNA sequences and need the involvement of multiple epigenetic factors such as DNA methylation, histone modifications and non-coding RNAs. These epigenetic inheritance vectors can synergistically shape the precise spatiotemporal gene expression pattern associated with phenotypic alternations (Zhang et al. 2019). Increasing evidence demonstrates that tsRNAs are one of the important epigenetic inheritance mediators, which is one fascinating function of tsRNAs (Chen et al. 2016, Sarker et al. 2019, Sharma et al. 2016, Zhang et al. 2018).

An interesting observation is that 5′ tsRNAs (~30-35nt) instead of the piRNAs, which are the most abundant small RNAs in adult testis, are highly expressed in mice mature sperm especially in sperm head (Peng et al. 2012), suggesting that these tsRNAs might play an important role in parental hereditary information. Reports showed that the injection of sperm tsRNAs, which are isolated from male mice with high-fat-diet (HFD) feeding, into normal zygotes generated offspring that partially recapitulated the paternal

metabolic disorders (Chen et al. 2016, Grandjean et al. 2015) or induced enhanced hedonic behaviors in offspring (Sarker et al. 2019). These 5′ tsRNAs are more dysregulated compared to miRNAs in the high-fat-diet mice sperm. Moreover, they are preferentially matched to the gene promoter region suggesting that 5′ tsRNA might disturb the expression of metabolic related genes in early mice embryo via transcriptional regulation (Chen et al. 2016). But how the early dysregulation of metabolic related genes in embryonic stage lasts to adult stage is still unclear and waits to be explored. In another study, it showed that low-protein-diet-induced sperm 5′ tsRNAs play a role in regulating retrotransposon activity and thus influencing specific gene expression in early embryo which might induce the metabolic disorder phenotype in adulthood (Sharma et al. 2016). Further investigation showed that methylation on the guanosine (m^2G) and cytosine (m^5C) are increased in the HFD-induced sperm 5′ tsRNAs, which might contribute to tsRNA-mediated epigenetic inheritance. Deletion in the *Dnmt2* (mouse tRNA methyltransferase) gene eliminates the increase of these two RNA modifications induced by HFD and abolishes sperm 5′ tsRNA-mediated epigenetic inheritance, suggesting that modification of sperm small RNAs is important for parental epigenetic inheritance (Zhang et al. 2018).

1.3.4 tsRNA and viral infection

Viruses rely on the host cell translation machinery for efficient synthesis of their own proteins which are necessary for their replication and propagation. tRNA, as the fundamental component of the translation machinery, are the key player in RNA virus

infections (Nunes et al. 2020). Due to the close relation between tRNA and tsRNAs, tsRNAs have also been investigated in the responses to viral infection over the past few years.

When the respiratory syncytial virus (RSV) infects epithelial cells in lung tissues, a lot of 5′ tsRNAs are produced in response to AGN signaling and promote the viral infection. RSV-induced 5′ tsRNAs can interact with the 3′ UTR of immune-activation related genes, such as *Apoer2,* to suppress host immune response thus leading to the enhancement of the viral replication (Deng et al. 2015, Wang et al. 2013). The abundant induction of 5′ tsRNAs are confirmed in other viral infection processes, for example, human T-cell leukemia virus 1 (Ruggero et al. 2014) and cytomegalovirus (Soares et al. 2015). However, in some virus infection process, the expression of 5′ tsRNAs are down-regulated (Taxis et al. 2019) or doesn't change significantly (Deng et al. 2014), suggesting the production of 5′ tsRNAs appears to be virus specific. Nevertheless, the exact role of tsRNAs underlying viral infection process remains to be revealed.

1.3.5 tsRNA and neurological diseases

ANG is an important RNase for the 5′ tRNAs biogenesis. The loss-of-function mutations of ANG are associated with increased risk for neurodegenerative diseases, such as Parkinson's disease and amyotrophic lateral sclerosis (ALS) (Su et al. 2020). It is logical to speculate that tsRNAs may play a role in the pathophysiological processes of these neurological diseases. Indeed, significant alternation of tsRNAs are identified both in the brain of Parkinson's disease mice model (Zhang et al. 2019) and the cortical tissue

of human post-mortem samples from Alzheimer's patients (Zhang et al. 2020). More important, ANG-induced 5′ tsRNA from $tRNA^{Ala}$ and $tRNA^{Cys}$ can form unique RNA G-quadruplex (G4) structure which can boost motor neuron recovery under apoptotic conditions. However, this neuroprotective activity of $tsRNA^{Ala}$ and $tsRNA^{Cys}$ can be impaired by the pathological G-quadruplex forming sequence from C9ORF72 gene, one of the leading causes of ALS (Ivanov et al. 2014). Another study demonstrated that excessive accumulation of tsRNAs in the brain of patients with Clp1 mutation, a component of the tRNA splicing machinery, can cause progressive motor neuron loss by inducing cell death (Karaca et al. 2014, Schaffer et al. 2014). These studies suggest that dysregulation of tsRNAs is potentially associated with neurological disease and need further investigation to reveal their specific role in the progression of neurological diseases.

1.3.6 tsRNA and symbiosis

Symbiosis refers to the relationship between two dissimilar organisms with close and long-term biological interactions. Micro RNAs in extracellular vesicles (EVs) are reported to be one of the key messengers in the cross-species communications which modulate the gene expression in the interacting organisms (Cai et al. 2019). Recently, tsRNAs are also demonstrated to be the communicator in the cross-species interactions (Cai et al. 2019). In a classical symbiotic relationship composing of plants legumes and nitrogen-fixing bacteria (Rhizobia), bacteria supply utilizable atmospheric nitrogen to plants in exchange for their carbohydrates. This symbiotic interaction takes place in root

nodules, and tsRNAs are identified as the modulators of root nodule development process. Rhizobia can generate both 3′ tsRNAs and 5′ tsRNAs, the latter is much less abundant, and deliver these tsRNAs to the soybean root cell via as-yet-unknown mechanism. These bacteria-derived tsRNAs can be loaded into AGO1 protein of the plant to silence host targeting genes which are critical for the early establishment of root nodulation (Baldrich et al. 2019, Ren et al. 2019). Whether the bacteria-derived tsRNAs are also transported in vesicles, like micro RNAs, from the bacterial cell to the root cell need further investigations.

1.3.7 tsRNA as biomarker for diseases

Given that tsRNA has been found abundantly in biofluids, the capability of tsRNAs to be non-invasive biomarkers is promising. As mentioned before, the early discovery of tsRNAs was in the urine of cancer patients which indicates its potential as cancer biomarker. Recently, change in serum tsRNAs has been applied to the diagnosis of breast cancer (Dhahbi et al. 2014), lung cancer (Gu et al. 2020), and renal cell carcinoma (Nientiedt et al. 2016). Additional work has shown that serum tsRNAs can be used to diagnose Parkinson's disease. Interestingly, the signature of tsRNA profile between male and female Parkinson's disease patients is different (Magee et al. 2019). Also, the urine tsRNAs recently are shown as a biomarker of chronic kidney disease (Khurana et al. 2017). With the decreasing cost of next-generation-sequencing, it is easier to expand these efforts to other biofluids in addition to serum and urine. In a large sequencing project leading by Extracellular RNA Communication Consortium,

researchers performed small-RNA sequencing across diverse biofluids including bile, plasma, cerebrospinal fluid, saliva, cell culture media, serum, and urine. This project found multiple types of small RNAs are presented in these biofluids, and tsRNAs are one of the most abundant types, especially in urine (Murillo et al. 2019, Srinivasan et al. 2019). The high abundance of tsRNA in urine will attract more attention to investigate tsRNA as non-invasion biomarker. However, it is important to note that the current founding between tsRNAs profile and specific disease are more correlative than causal, it remains unclear whether the observed changes in tsRNA profile are the cause of the disease state or result from it. Nonetheless, tsRNA expression signature could be harnessed for developing diagnostic or prognostic biomarkers.

1.4 The action mechanism of tsRNA

Unlike other canonical sncRNAs such as miRNAs, siRNAs, and piRNAs, which usually function through strict molecular mechanisms by binding to Argonaute family proteins, the mechanisms of action of tsRNAs appear to be more diverse. The functional mechanism of tsRNAs can be dissected into three possible major steps: (i) tsRNAs function via mimicry/replacement of tRNAs with sequence/structure effects (Figure 1.2 A), (ii) tsRNAs function via association with ribonucleoproteins (RNPs) (Figure 1.2B), and (iii) tsRNAs function via Argonaute proteins (i.e., AGO, PIWI) (Figure 1.2C). In addition, tsRNAs (and other mobile RNAs) can travel between cells and organisms, enabling long-distance regulation inside and between organisms, and even mediate cross-kingdom regulation between prokaryotes and eukaryotes (Figure 1.3).

1.4.1 Mimicry/displacement of tRNAs

Since the 1960s it has been reported that a tRNA fragment from $tRNA^{fMet}$ can interact with ribosomal subunits in a fashion similar to that of its precursor mature $tRNA^{fMet}$ (Rudland et al. 1969); this suggests that tsRNAs may naturally compete with the function of tRNAs under some circumstances. The tRNA:tsRNA ratio may reach an equilibrium under normal cellular conditions, whereas the dynamic change of this ratio under different environments (e.g., stress conditions that cleave more tRNAs) enables regulatory effects. tsRNA can both positively and negatively regulate the translation process. Example in Archaea (e.g., Haloferax volcanii) has shown that a stress-induced 5′ $tsRNA^{Val}$ can interact with the rRNA of the ribosomal subunit, and interferes with the loading of mRNAs into ribosomes, thus inhibiting global protein synthesis (Gebetsberger et al. 2017, Gebetsberger et al. 2012) (Figure 1.2A, top). By contrast, in another example, in Trypanosoma brucei, a unicellular protozoan, a stress-induced 3′ $tsRNA^{Thr}$ can be incorporated into ribosomes to enhance mRNA loading during stress recovery, and thus increases global translation (Fricker et al. 2019). These opposing examples suggest that tsRNAs can function flexibly in a context-dependent manner, possibly depending on where they interact with the ribosome and how they might alter the local structure of the rRNAs. In another human cell line example, 3′ $tsRNA^{Leu}$ can bind to specific ribosomal protein mRNAs (e.g., RPS28 and RPS15) to enhance their translation by changing the local secondary structure of this transcripts (Kim et al. 2017, Kim et al. 2019); depletion of 3′ $tsRNA^{Leu}$ causes decreased translation of RPS28, which impairs pre-18S ribosomal RNA processing and ribosome biogenesis (Kim et al. 2017).

In addition to tsRNA–ribosome interactions, many RNA viruses and retrotransposons (e.g., endogenous retroviruses, ERVs) can duplicate themselves by using the 3′ terminus of mature tRNA as primers for their reverse transcription (RT) (Nunes et al. 2020) (Figure 1.2A, top). In a recent study, transfection of exogenous 18 nt 3′ tsRNAs led to competition with the mature tRNAs for the primer binding sequence (PBS) of the ERVs, leading to reverse transcription blockade and thus impeding retroviral cDNA synthesis (Schorn et al. 2017). A similar 3′ tsRNA blockade effect at the PBS site may also take place in HIV-infected T cells. In this case, the 3′ tsRNA also required Ago2 proteins suggesting that RNAi machinery is involved in this silencing effect (Yeung et al. 2009). By contrast, in other scenarios, tsRNAs can be directly used as primers for reverse transcription in human T cell leukemia virus type 1 (HTLV-1) (Ruggero et al. 2014) and copia retrovirus-like particles (Kikuchi et al. 1990) . These opposite roles of tsRNAs in retroviral reverse transcription processes suggest adaptation of the viral world to specific tsRNAs that either block or mimic the function of tRNAs. In another interesting case, ATE1-mediated post-translational arginylation of proteins directly competes with protein translation involving the use of arginine-charged $tRNA^{Arg}$ (Avcilar-Kucukgoze et al. 2020). In this context, it was found that when arginine-charged $tRNA^{Arg}$ (Arg-$tRNA^{Arg}$) is fragmented into 3′ tsRNA, arginine remains bound to the 3′ $tsRNA^{Arg}$ (Arg-3′$tsRNA^{Arg}$) and the translation-incompetent Arg-3′$tsRNA^{Arg}$ can still be efficiently used as a donor for arginylation of protein by the ATE1 enzyme (Avcilar-Kucukgoze et al. 2020) (Figure 1.2A, bottom). Although this is only one piece of

evidence, these data suggest a scenario in which the balance between translation and arginylation might be regulated by the $tsRNA^{Arg}$:$tRNA^{Arg}$ ratio.

There should be more tsRNA functions can be discovered under the principle of mimicking/displacing precursor tRNAs because this scenario may represent the most ancient regulatory mode of tsRNAs – simply based on sequence/structural competition with its precursor tRNA.

1.4.2 Forming RNPs

Another well-studied mode of tsRNA functionality is through the formation of tsRNA-RNP complexes. For example, under stress conditions, angiogenin-induced tsRNAs can induce stress granule (SG) formation in a portion (~4%) of cells and the tsRNAs could be incorporated into SGs (Emara et al. 2010, Lyons et al. 2016) (Figure 1.2B, top). The ability of tsRNAs to induce SG formation requires the action of YBX1 (Lyons et al. 2016), and the role of tsRNAs in suppressing translation is dependent on their binding with eukaryotic initiation factors (eIFs) to sequester them from the translational machinery (Ivanov et al. 2011, Lyons et al. 2016). The formation of this tsRNA-RNP complex involves tsRNA secondary structures such as intermolecular RNA G-quadruplexes (RNA-G4) (Lyons et al. 2017), and RNA modifications such as PUS7-mediated Ψ modifications at U8 (Guzzi et al. 2018). Angiogenin-induced tsRNA has also been shown to form an RNP complex with cytochrome c that is released from the mitochondria, where the tsRNA-mediated sequestration of cytochrome c protects the cells from apoptosome formation and apoptosis under cell stress (Saikia et al. 2014) .

Interestingly, the binding of tsRNAs to cytochrome c to inhibit apoptosis resembles the function of some tRNAs (Saikia et al. 2014).

In addition to functioning with cytosolic RNPs, tsRNA can also interact with spliceosomal proteins (e.g., hnRNPF/H) in the nuclei, where hnRNPF/H contributes to the formation of the Cajal body (Figure 1.2B, bottom). Cajal bodies are known to regulate the production of nuclear small RNAs, including small nucleolar RNAs (snoRNAs), small Cajal body specific RNAs (scaRNAs), and snRNAs. These nuclear small RNAs, along with tsRNAs, may form other nuclear RNP complexes that contribute to the processing (Couvillion et al. 2012) and modification of rRNAs, resulting in ribosome heterogeneity that leads to translational specificity for a selected pool of mRNAs to direct the cell to a specific functional state (Genuth et al. 2018). To date, the possible nuclear function of tsRNAs remains mysterious but intriguing, especially in early embryo development. Recent evidence show that pronuclear zygotic injection of tsRNA-containing sperm RNA fractions from high-fat diet-exposed males can induce metabolic phenotypes in the resulting offspring, suggesting that tsRNAs might function through nuclear events that drive butterfly effects throughout embryo development since the zygote stage, possibly by regulating the translational specificity of the early embryo (Chen et al. 2016).

1.4.3 Binding to Argonaute

tsRNA binding to Argonaute family proteins has been reported to mediate post-transcriptional silencing (either by RNA degradation or mRNA translational inhibition) in

multiple cases (Haussecker et al. 2010, Kuscu et al. 2018, Luo et al. 2018, Maute et al. 2013, Ren et al. 2019), mostly in an RNAi-like fashion using complementary sequences binding to the targeted mRNAs (Figure 1.2C). However, the Argonaute-dependent RNAi systems only prevail in eukaryotes and are believed to have emerged primarily to defend against viruses, represented by either the AGO-centered antiviral machinery in the cytoplasm or the PIWI-centered function in the nuclei to suppress endogenous viruses (e.g., transposons) (Shabalina et al. 2008).

Unlike in eukaryotes, prokaryotic Argonaute proteins primarily use small DNA as a guide to cleave DNA, a process known as DNA interference (Swarts et al. 2014), to inhibit the propagation of foreign plasmids and infection by phages (Kuzmenko et al. 2020) (functionally similar to the CRISPR/Cas system but mechanistically analogous to RNAi). We believe that this evidence from prokaryotic systems suggests an ancient origin of Argonaute proteins that pre-dates the emergence of RNAi. Since tsRNAs are widely present in prokaryotes (Gebetsberger et al. 2017, Gebetsberger et al. 2012, Kumar et al. 2014), it is an intriguing hypothesis that tsRNAs might be used by prokaryotic Argonaute as an alternative substrate. Indeed, in a seminal study exploring the binding of bacterial Argonaute to RNAs, it was found that the 3′ halves of two types of $tsRNA^{Met}$ are abundantly found as the majority of 45 nt reads (Olovnikov et al. 2013), although their biological functions remain unresolved. It is possible that other tsRNAs might also bind to the prokaryotic Argonaute, which awaits to be confirmed by improved library construction protocols that consider the specific tsRNA termini and modifications (Shi et al. 2021).

Importantly, although the earlier mentioned tsRNA mechanisms of action have been discussed separately and from an evolutionary perspective, different mechanisms can coexist or function synergistically to achieve common outcomes. For example, the suppression of retrotransposons could be achieved by both tsRNA displacement of tRNA-mediated reverse transcription (Schorn et al. 2017) and RNAi-like post-transcriptional suppression (Martinez et al. 2017, Schorn et al. 2017). tsRNA-mediated translational regulation can be achieved by tsRNA–ribosome interaction, tsRNA-RNP-based eIF sequestration, and Argonaute-dependent RNAi effects (Shi et al. 2019). In addition, it has been shown in Tetrahymena that tsRNA can bind to an Argonaute family protein to form nuclear RNPs, which may facilitate nuclear rRNA processing and cell growth (Couvillion et al. 2012, Couvillion et al. 2010).

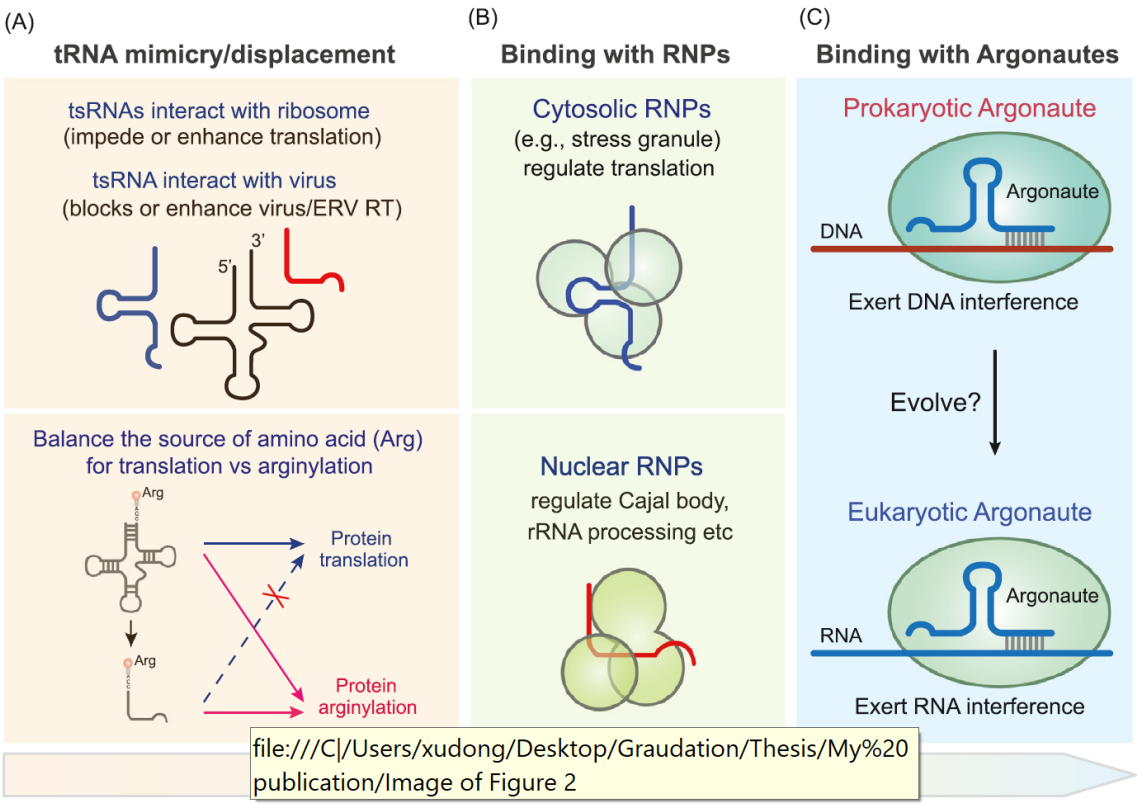


Figure 1.2. The evolving principles of tRNA-derived small RNA (tsRNA) functionality from an evolutionary perspective. Three possible major steps regarding how tsRNA function has emerged and evolved; the arrow at the bottom shows evolution of tsRNA functionality towards more specialized roles. (A) tsRNA function by tRNA mimicry/displacement and structural effects. (B) tsRNA function via cytosolic and nuclear ribonucleoproteins (RNPs). (C) tsRNA function through binding to prokaryotic or eukaryotic Argonaute proteins. Abbreviations: ERV, endogenous retrovirus; RT, reverse transcription.

1.5 tsRNAs on the move

In addition to functioning in the cytoplasm and nuclei, tsRNAs are also abundant in body fluids (Godoy et al. 2018) and can be delivered between somatic cells or between somatic and germ cells (Sharma et al. 2016), and even across species in which bacterial tsRNAs are essential for prokaryote–eukaryote communication to ensure microbiome–

plant symbiosis for soybean growth (Ren et al. 2019). Intercellular communication by tsRNAs could be either dependent or independent of extracellular vesicles (EVs) (Tosar et al. 2020) (Figure 1.3A), which resonates with the wisdom of Charles Darwin - who proposed in 1868 that each cell emits small particles or molecules, called 'gemmules', that diffuse and move between both somatic cells and germ cells (Darwin 1868), a concept that is now gaining new momentum (Liu et al. 2018).

In fact, the phenomenon of membrane vesicle release in Bacteria, Archaea, and Eukaryotes is an evolutionarily conserved mechanism (Deatherage et al. 2012) that might enable robust horizontal information exchange and facilitate the selection and propagation of traits under pressure. Whether and how tsRNAs are selectively or randomly encapsulated into EVs as cargo and specifically delivered to another cell or organism to exert their function remains largely unknown and certainly deserves intense investigation.

Importantly, tsRNAs have been observed outside the cell without being encapsulated into EVs (Dhahbi et al. 2013, Tosar et al. 2020, Zhang et al. 2014). In these cases, tsRNA could travel with extracellular RNPs or depend on specific RNA structures (e.g., dimers, RNA G4 tetraplexes (Ivanov et al. 2014, Tosar et al. 2018)) or modifications, which increase their stability. Interestingly, engineered mRNAs harboring TLSs with predicted stem–bulge–stem–loop structures are sufficient to mediate systemic mRNA transport in plants and move through graft junctions (Zhang et al. 2016). Thus, specific tRNA-related and tsRNA-mediated structures may represent RNA mobility motifs that facilitate EV-independent RNA transfer, and their entry into a target cell may

or may not rely on specific cell-membrane receptors. This type of extracellular tsRNA transfer could have significant implications because it might be involved in crosstalk between food-derived RNAs and the gut microbiome or enable the direct transport of food-derived small RNAs into animal cells (Zhou et al. 2015) (Figure 1.3B).

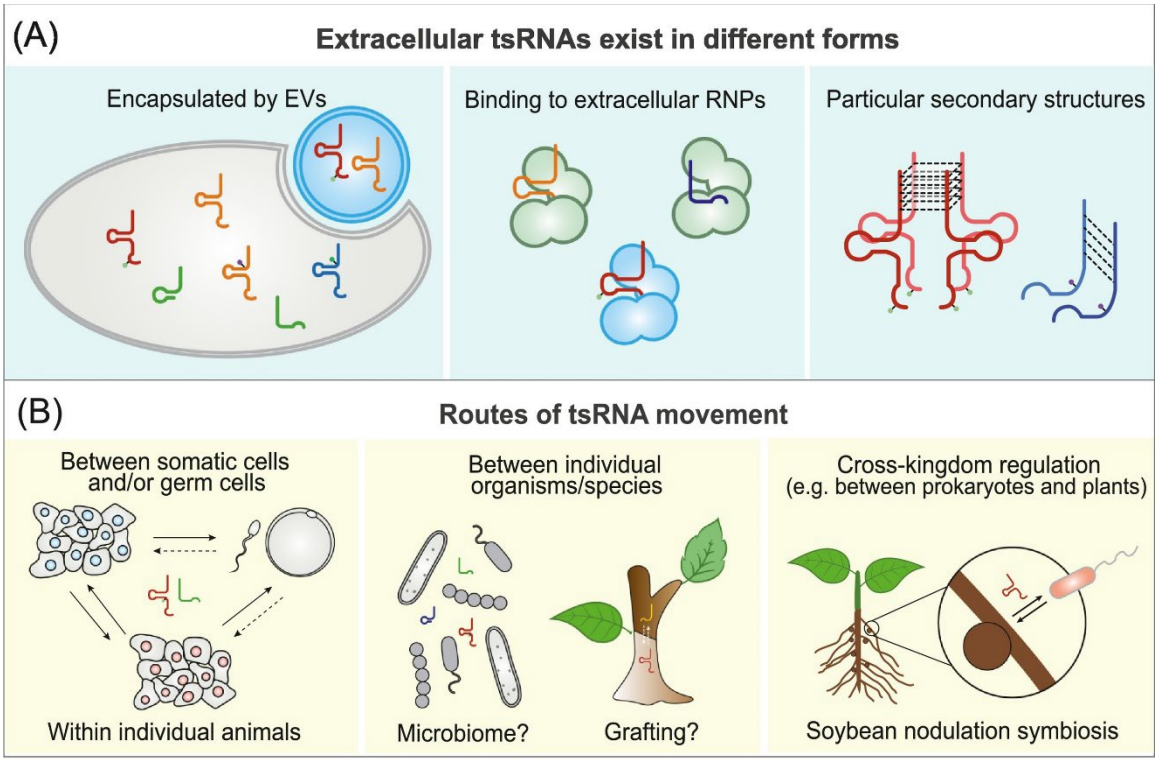


Figure 1.3. Mechanisms and routes of tRNA-derived small RNA (tsRNA) movement between cells, individual organisms, and species. (A) Extracellular tsRNAs can be encapsulated in extracellular vesicles (EVs), bind to ribonucleoproteins (RNPs), or form specific secondary/3D structures with the aid of RNA modifications. (B) Known and predicted (?) movement of tsRNA between somatic cells or between somatic and germ cells in a single individual, or between organisms/species (e.g., between microbiomes, grafted plants), where tsRNAs may mediate cross-kingdom regulation such as between bacteria and plants/animals.

1.6 Challenges on tsRNA functional study

As mentioned before, one unique feature of tsRNAs is the diverse modifications on it, and these RNA modifications closely interfere with the biogenesis and functions of tsRNAs. However, fully understanding the impact of RNA modifications to each tsRNA's function remains experimentally challenging, as synthetic tsRNAs may not fully mimic the modified tsRNAs that exist in tissues/cells. In addition, when performing

functional studies by knock-down, the blockade of tsRNAs using complementary sequences such as antisense oligonucleotides (ASOs) or locked nucleic acids (LNAs) should be carefully validated so as to not affect the levels of precursor tRNAs (Kim et al. 2017).

When profiling tsRNAs using small RNA-seq, the specific tsRNA termini (e.g., 3′-phosphate and 2′,3′-cyclic phosphate) and RNA modifications (e.g., m^1A, m^3C, m^1G, and N2-dimethylguanosine, $m^2{}_2G$) have created challenges to prevent efficient and complete conversion of tsRNAs into cDNA libraries during standard small RNA-seq protocols, resulting in biased detection and quantitation during deep sequencing (Shi et al. 2021). This biased tsRNAs profile is a big problem for the application of tsRNA as the non-invasion biomarker. Recently improved RNA-seq methods such as panoramic RNA display overcoming RNA modification aborted sequencing (PANDORA)-seq (Shi et al. 2021) and Cap-Clip acid pyrophosphatase, PNK, and AlkB (CPA)-seq (Wang et al. 2021) aim to overcome these problems by removing key RNA modifications that block adapter ligation and RT by using consecutive enzymatic treatment; these techniques await further optimization such as combination with other strategies focusing on the 5′-termini (Kugelberg et al. 2021).

1.7 Conclusion remarks

The studies on the biogenesis and functions of tsRNAs have advanced greatly for the last ten years and constituted an emerging field in small RNA biology. As a result, the biogenesis and function of tsRNAs are being well documented than before. However, the

complexity of tsRNAs in biogenesis and functional mechanism are greater than the well-studied classical small RNAs such as miRNAs, siRNAs and piRNAs. This complexity first stems from the different cleavage sites of various RNases response for the tsRNAs biogenesis that it can produce different types of tsRNAs even from same tRNA precursor by different enzyme under different cellular status, such as 5′ tsRNAs half, 3′ tsRNA half, shorter 5′ tsRNAs, shorter 3′ tsRNAs and internal tsRNAs. Another factor contribute to this great complexity is the RNA modifications on the tRNAs which impact the process of tsRNA biogenesis, moreover, the “inherited” RNA modifications on tsRNAs also impact the function of tsRNAs. So precise biogenesis network of tsRNAs and comprehensive RNA modification profile need to be established with more efforts.

1.8 Reference

Andersen, K. L. and K. Collins (2012). "Several RNase T2 enzymes function in induced tRNA and rRNA turnover in the ciliate Tetrahymena." Mol Biol Cell **23**(1): 36-44.

Avcilar-Kucukgoze, I., H. Gamper, C. Polte, Z. Ignatova, R. Kraetzner, M. Shtutman, Y. M. Hou, D. W. Dong and A. Kashina (2020). "tRNA(Arg)-Derived Fragments Can Serve as Arginine Donors for Protein Arginylation." Cell Chem Biol **27**(7): 839-849 e834.

Babiarz, J. E., J. G. Ruby, Y. Wang, D. P. Bartel and R. Blelloch (2008). "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs." Genes Dev **22**(20): 2773-2785.

Balatti, V., G. Nigita, D. Veneziano, A. Drusco, G. S. Stein, T. L. Messier, N. H. Farina, J. B. Lian, L. Tomasello, C. G. Liu, A. Palamarchuk, J. R. Hart, C. Bell, M. Carosi, E. Pescarmona, L. Perracchio, M. Diodoro, A. Russo, A. Antenucci, P. Visca, A. Ciardi, C. C. Harris, P. K. Vogt, Y. Pekarsky and C. M. Croce (2017). "tsRNA signatures in cancer." Proc Natl Acad Sci U S A **114**(30): 8071-8076.

Baldrich, P. and B. C. Meyers (2019). "Bacteria send messages to colonize plant roots." Science **365**(6456): 868-869.

Barraud, P., A. Gato, M. Heiss, M. Catala, S. Kellner and C. Tisne (2019). "Time-resolved NMR monitoring of tRNA maturation." Nat Commun **10**(1): 3373.

Bartel, D. P. (2018). "Metazoan MicroRNAs." Cell **173**(1): 20-51.

Blanco, S., R. Bandiera, M. Popis, S. Hussain, P. Lombard, J. Aleksic, A. Sajini, H. Tanna, R. Cortes-Garrido, N. Gkatza, S. Dietmann and M. Frye (2016). "Stem cell function and stress response are controlled by protein synthesis." Nature **534**(7607): 335-340.

Boccaletto, P., M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm and J. M. Bujnicki (2018). "MODOMICS: a database of RNA modification pathways. 2017 update." Nucleic Acids Res **46**(D1): D303-D307.

Borek, E., B. S. Baliga, C. W. Gehrke, C. W. Kuo, S. Belman, W. Troll and T. P. Waalkes (1977). "High turnover rate of transfer RNA in tumor tissue." Cancer Res **37**(9): 3362-3366.

Cai, Q., B. Y. He, A. Weiberg, A. H. Buck and H. L. Jin (2019). "Small RNAs and extracellular vesicles: New mechanisms of cross-species communication and innovative tools for disease control." Plos Pathogens **15**(12).

Chan, C. M., C. Zhou and R. H. Huang (2009). "Reconstituting bacterial RNA repair and modification in vitro." Science **326**(5950): 247.

Chen, Q., M. Yan, Z. Cao, X. Li, Y. Zhang, J. Shi, G. H. Feng, H. Peng, X. Zhang, Y. Zhang, J. Qian, E. Duan, Q. Zhai and Q. Zhou (2016). "Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder." Science **351**(6271): 397-400.

Chen, Z., M. Qi, B. Shen, G. Luo, Y. Wu, J. Li, Z. Lu, Z. Zheng, Q. Dai and H. Wang (2019). "Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs." Nucleic Acids Res **47**(5): 2533-2545.

Cole, C., A. Sobala, C. Lu, S. R. Thatcher, A. Bowman, J. W. Brown, P. J. Green, G. J. Barton and G. Hutvagner (2009). "Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs." RNA **15**(12): 2147-2160.

Cosentino, C., S. Toivonen, E. Diaz Villamil, M. Atta, J. L. Ravanat, S. Demine, A. A. Schiavo, N. Pachera, J. P. Deglasse, J. C. Jonas, D. Balboa, T. Otonkoski, E. R. Pearson, P. Marchetti, D. L. Eizirik, M. Cnop and M. Igoillo-Esteve (2018). "Pancreatic beta-cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes." Nucleic Acids Res **46**(19): 10302-10318.

Couvillion, M. T., G. Bounova, E. Purdom, T. P. Speed and K. Collins (2012). "A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus." Mol Cell **48**(4): 509-520.

Couvillion, M. T., R. Sachidanandam and K. Collins (2010). "A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo." Genes Dev **24**(24): 2742-2747.

Darwin, C. (1868). "The Variation of Animals and Plants under Domestication." Br Foreign Med Chir Rev **42**(83): 143-166.

Deatherage, B. L. and B. T. Cookson (2012). "Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life." Infect Immun **80**(6): 1948-1957.

Deng, J., R. N. Ptashkin, Y. Chen, Z. Cheng, G. Liu, T. Phan, X. Deng, J. Zhou, I. Lee, Y. S. Lee and X. Bao (2015). "Respiratory Syncytial Virus Utilizes a tRNA Fragment to Suppress Antiviral Responses Through a Novel Targeting Mechanism." Mol Ther **23**(10): 1622-1629.

Deng, J., R. N. Ptashkin, Q. Wang, G. Liu, G. Zhang, I. Lee, Y. S. Lee and X. Bao (2014). "Human metapneumovirus infection induces significant changes in small noncoding RNA expression in airway epithelial cells." Mol Ther Nucleic Acids **3**: e163.

Dhahbi, J. M., S. R. Spindler, H. Atamna, D. Boffelli and D. I. Martin (2014). "Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer." Biomark Cancer **6**: 37-47.

Dhahbi, J. M., S. R. Spindler, H. Atamna, A. Yamakawa, D. Boffelli, P. Mote and D. I. Martin (2013). "5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction." BMC Genomics **14**: 298.

Diener, T. O. (1989). "Circular RNAs: relics of precellular evolution?" Proc Natl Acad Sci U S A **86**(23): 9370-9374.

Donovan, J., S. Rath, D. Kolet-Mandrikov and A. Korennykh (2017). "Rapid RNase L-driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery." RNA **23**(11): 1660-1671.

Emara, M. M., P. Ivanov, T. Hickman, N. Dawra, S. Tisdale, N. Kedersha, G. F. Hu and P. Anderson (2010). "Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly." J Biol Chem **285**(14): 10959-10968.

Fricker, R., R. Brogli, H. Luidalepp, L. Wyss, M. Fasnacht, O. Joss, M. Zywicki, M. Helm, A. Schneider, M. Cristodero and N. Polacek (2019). "A tRNA half modulates translation as stress response in Trypanosoma brucei." Nat Commun **10**(1): 118.

Fu, H., J. Feng, Q. Liu, F. Sun, Y. Tie, J. Zhu, R. Xing, Z. Sun and X. Zheng (2009). "Stress induces tRNA cleavage by angiogenin in mammalian cells." FEBS Lett **583**(2): 437-442.

Gebetsberger, J., L. Wyss, A. M. Mleczko, J. Reuther and N. Polacek (2017). "A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress." RNA Biol **14**(10): 1364-1373.

Gebetsberger, J., M. Zywicki, A. Kunzi and N. Polacek (2012). "tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii." Archaea **2012**: 260909.

Genuth, N. R. and M. Barna (2018). "The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life." Mol Cell **71**(3): 364-374.

Godoy, P. M., N. R. Bhakta, A. J. Barczak, H. Cakmak, S. Fisher, T. C. MacKenzie, T. Patel, R. W. Price, J. F. Smith, P. G. Woodruff and D. J. Erle (2018). "Large Differences in Small RNA Composition Between Human Biofluids." Cell Reports **25**(5): 1346-1358.

Goncalves, K. A., L. Silberstein, S. Li, N. Severe, M. G. Hu, H. Yang, D. T. Scadden and G. F. Hu (2016). "Angiogenin Promotes Hematopoietic Regeneration by Dichotomously Regulating Quiescence of Stem and Progenitor Cells." Cell **166**(4): 894-906.

Goodarzi, H., X. Liu, H. C. Nguyen, S. Zhang, L. Fish and S. F. Tavazoie (2015). "Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement." Cell **161**(4): 790-802.

Grandjean, V., S. Fourre, D. A. De Abreu, M. A. Derieppe, J. J. Remy and M. Rassoulzadegan (2015). "RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders." Sci Rep **5**: 18193.

Gu, W. J., J. C. Shi, H. Liu, X. D. Zhang, J. J. Zhou, M. S. Li, D. D. Zhou, R. Li, J. Z. Lv, G. X. Wen, S. S. Zhu, T. Qi, W. Li, X. J. Wang, Z. H. Wang, H. Zhu, C. C. Zhou, K. S. Knox, T. Wang, Q. Chen, Z. Q. Qian and T. Zhou (2020). "Peripheral blood non-canonical small non-coding RNAs as novel biomarkers in lung cancer." Molecular Cancer **19**(1).

Guzzi, N., M. Ciesla, P. C. T. Ngoc, S. Lang, S. Arora, M. Dimitriou, K. Pimkova, M. N. E. Sommarin, R. Munita, M. Lubas, Y. Lim, K. Okuyama, S. Soneji, G. Karlsson, J. Hansson, G. Jonsson, A. H. Lund, M. Sigvardsson, E. Hellstrom-Lindberg, A. C. Hsieh and C. Bellodi (2018). "Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells." Cell **173**(5): 1204-1216 e1226.

Haussecker, D., Y. Huang, A. Lau, P. Parameswaran, A. Z. Fire and M. A. Kay (2010). "Human tRNA-derived small RNAs in the global regulation of RNA silencing." RNA **16**(4): 673-695.

Hogg, M. C., R. Raoof, H. El Naggar, N. Monsefi, N. Delanty, D. F. O'Brien, S. Bauer, F. Rosenow, D. C. Henshall and J. H. Prehn (2019). "Elevation in plasma tRNA fragments precede seizures in human epilepsy." J Clin Invest **129**(7): 2946-2951.

Honda, S., T. Kawamura, P. Loher, K. Morichika, I. Rigoutsos and Y. Kirino (2017). "The biogenesis pathway of tRNA-derived piRNAs in Bombyx germ cells." Nucleic Acids Res **45**(15): 9108-9120.

Honda, S., P. Loher, M. Shigematsu, J. P. Palazzo, R. Suzuki, I. Imoto, I. Rigoutsos and Y. Kirino (2015). "Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers." Proc Natl Acad Sci U S A **112**(29): E3816-3825.

Huang, B., H. Yang, X. Cheng, D. Wang, S. Fu, W. Shen, Q. Zhang, L. Zhang, Z. Xue, Y. Li, Y. Da, Q. Yang, Z. Li, L. Liu, L. Qiao, Y. Kong, Z. Yao, P. Zhao, M. Li and R. Zhang (2017). "tRF/miR-1280 Suppresses Stem Cell-like Cells and Metastasis in Colorectal Cancer." Cancer Res **77**(12): 3194-3206.

Ivanov, P., M. M. Emara, J. Villen, S. P. Gygi and P. Anderson (2011). "Angiogenin-induced tRNA fragments inhibit translation initiation." Mol Cell **43**(4): 613-623.

Ivanov, P., E. O'Day, M. M. Emara, G. Wagner, J. Lieberman and P. Anderson (2014). "G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments." Proc Natl Acad Sci U S A **111**(51): 18201-18206.

Ji, L. and X. Chen (2012). "Regulation of small RNA stability: methylation and beyond." Cell Res **22**(4): 624-636.

Karaca, E., S. Weitzer, D. Pehlivan, H. Shiraishi, T. Gogakos, T. Hanada, S. N. Jhangiani, W. Wiszniewski, M. Withers, I. M. Campbell, S. Erdin, S. Isikay, L. M. Franco, C. Gonzaga-Jauregui, T. Gambin, V. Gelowani, J. V. Hunter, G. Yesil, E. Koparir, S. Yilmaz, M. Brown, D. Briskin, M. Hafner, P. Morozov, T. A. Farazi, C. Bernreuther, M. Glatzel, S. Trattnig, J. Friske, C. Kronnerwetter, M. N. Bainbridge, A. Gezdirici, M. Seven, D. M. Muzny, E. Boerwinkle, M. Ozen, T. Clausen, T. Tuschl, A. Yuksel, A. Hess, R. A. Gibbs, J. Martinez, J. M. Penninger, J. R. Lupski and B. H. C. Mendelian (2014). "Human CLP1 Mutations Alter tRNA Biogenesis, Affecting Both Peripheral and Central Nervous System Function." Cell **157**(3): 636-650.

Khurana, R., G. Ranches, S. Schafferer, M. Lukasser, M. Rudnicki, G. Mayer and A. Huttenhofer (2017). "Identification of urinary exosomal noncoding RNAs as novel biomarkers in chronic kidney disease." Rna **23**(2): 142-152.

Kikuchi, Y., N. Sasaki and Y. Ando-Yamagami (1990). "Cleavage of tRNA within the mature tRNA sequence by the catalytic RNA of RNase P: implication for the formation of the primer tRNA fragment for reverse transcription in copia retrovirus-like particles." Proc Natl Acad Sci U S A **87**(20): 8105-8109.

Kim, H. K., G. Fuchs, S. Wang, W. Wei, Y. Zhang, H. Park, B. Roy-Chaudhuri, P. Li, J. Xu, K. Chu, F. Zhang, M. S. Chua, S. So, Q. C. Zhang, P. Sarnow and M. A. Kay (2017). "A transfer-RNA-derived small RNA regulates ribosome biogenesis." Nature **552**(7683): 57-62.

Kim, H. K., J. Xu, K. Chu, H. Park, H. Jang, P. Li, P. N. Valdmanis, Q. C. Zhang and M. A. Kay (2019). "A tRNA-Derived Small RNA Regulates Ribosomal Protein S28 Protein Levels after Translation Initiation in Humans and Mice." Cell Rep **29**(12): 3816-3824 e3814.

Kim, H. K., J. H. Yeom and M. A. Kay (2020). "Transfer RNA-Derived Small RNAs: Another Layer of Gene Regulation and Novel Targets for Disease Therapeutics." Molecular Therapy **28**(11): 2340-2357.

Krishna, S., D. G. Yim, V. Lakshmanan, V. Tirumalai, J. L. Koh, J. E. Park, J. K. Cheong, J. L. Low, M. J. Lim, S. K. Sze, P. Shivaprasad, A. Gulyani, S. Raghavan, D. Palakodeti and R. DasGupta (2019). "Dynamic expression of tRNA-derived small RNAs define cellular states." EMBO Rep **20**(7): e47789.

Kugelberg, U., D. Natt, S. Skog, C. Kutter and A. Ost (2021). "5 XP sRNA-seq: efficient identification of transcripts with and without 5 phosphorylation reveals evolutionary conserved small RNA." RNA Biol **18**(11): 1588-1599.

Kumar, P., J. Anaya, S. B. Mudunuri and A. Dutta (2014). "Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets." Bmc Biology **12**.

Kumar, P., S. B. Mudunuri, J. Anaya and A. Dutta (2015). "tRFdb: a database for transfer RNA fragments." Nucleic Acids Res **43**(Database issue): D141-145.

Kuscu, C., P. Kumar, M. Kiran, Z. Su, A. Malik and A. Dutta (2018). "tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner." RNA **24**(8): 1093-1105.

Kuzmenko, A., A. Oguienko, D. Esyunina, D. Yudin, M. Petrova, A. Kudinova, O. Maslova, M. Ninova, S. Ryazansky, D. Leach, A. A. Aravin and A. Kulbachinskiy (2020). "DNA targeting and interference by a bacterial Argonaute nuclease." Nature **587**(7835): 632-637.

Levitz, R., D. Chapman, M. Amitsur, R. Green, L. Snyder and G. Kaufmann (1990). "The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease." EMBO J **9**(5): 1383-1389.

Li, Z., C. Ender, G. Meister, P. S. Moore, Y. Chang and B. John (2012). "Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs." Nucleic Acids Res **40**(14): 6787-6799.

Liu, Y. and Q. Chen (2018). "150 years of Darwin's theory of intercellular flow of hereditary information." Nat Rev Mol Cell Biol **19**(12): 749-750.

Lu, J., A. Esberg, B. Huang and A. S. Bystrom (2008). "Kluyveromyces lactis gamma-toxin, a ribonuclease that recognizes the anticodon stem loop of tRNA." Nucleic Acids Res **36**(4): 1072-1080.

Luhtala, N. and R. Parker (2010). "T2 Family ribonucleases: ancient enzymes with diverse roles." Trends Biochem Sci **35**(5): 253-259.

Luo, S., F. He, J. Luo, S. Dou, Y. Wang, A. Guo and J. Lu (2018). "Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response." Nucleic Acids Res **46**(10): 5250-5268.

Lyons, S. M., C. Achorn, N. L. Kedersha, P. J. Anderson and P. Ivanov (2016). "YB-1 regulates tiRNA-induced Stress Granule formation but not translational repression." Nucleic Acids Res **44**(14): 6949-6960.

Lyons, S. M., D. Gudanis, S. M. Coyne, Z. Gdaniec and P. Ivanov (2017). "Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs." Nat Commun **8**(1): 1127.

Magee, R., E. Londin and I. Rigoutsos (2019). "TRNA-derived fragments as sex-dependent circulating candidate biomarkers for Parkinson's disease." Parkinsonism & Related Disorders **65**: 203-209.

Martinez, G., S. G. Choudury and R. K. Slotkin (2017). "tRNA-derived small RNAs target transposable element transcripts." Nucleic Acids Res **45**(9): 5142-5152.

Maute, R. L., C. Schneider, P. Sumazin, A. Holmes, A. Califano, K. Basso and R. Dalla-Favera (2013). "tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma." Proceedings of the National Academy of Sciences of the United States of America **110**(4): 1404-1409.

Maute, R. L., C. Schneider, P. Sumazin, A. Holmes, A. Califano, K. Basso and R. Dalla-Favera (2013). "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 **110**(4): 1404-1409.

Muller, M., M. Hartmann, I. Schuster, S. Bender, K. L. Thuring, M. Helm, J. R. Katze, W. Nellen, F. Lyko and A. E. Ehrenhofer-Murray (2015). "Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine." Nucleic Acids Res **43**(22): 10952-10962.

Murillo, O. D., W. Thistlethwaite, J. Rozowsky, S. L. Subramanian, R. Lucero, N. Shah, A. R. Jackson, S. Srinivasan, A. Chung, C. D. Laurent, R. R. Kitchen, T. Galeev, J. Warrell, J. A. Diao, J. A. Welsh, K. Hanspers, A. Riutta, S. Burgstaller-Muehlbacher, R. V. Shah, A. Yeri, L. M. Jenkins, M. E. Ahsen, C. Cordon-Cardo, N. Dogra, S. M. Gifford, J. T. Smith, G. Stolovitzky, A. K. Tewari, B. H. Wunsch, K. K. Yadav, K. M. Danielson, J. Filant, C. Moeller, P. Nejad, A. Paul, B. Simonson, D. K. Wong, X. Zhang, L. Balaj, R. Gandhi, A. K. Sood, R. P. Alexander, L. Wang, C. L. Wu, D. T. W. Wong,

D. J. Galas, K. van Keuren-Jensen, T. Patel, J. C. Jones, S. Das, K. H. Cheung, A. R. Pico, A. I. Su, R. L. Raffai, L. C. Laurent, M. E. Roth, M. B. Gerstein and A. Milosavljevic (2019). "exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids." Cell **177**(2): 463-+.

Nientiedt, M., M. Deng, D. Schmidt, S. Perner, S. C. Muller and J. Ellinger (2016). "Identification of aberrant tRNA-halves expression patterns in clear cell renal cell carcinoma." Scientific Reports **6**.

Nunes, A., D. R. Ribeiro, M. Marques, M. A. S. Santos, D. Ribeiro and A. R. Soares (2020). "Emerging Roles of tRNAs in RNA Virus Infections." Trends Biochem Sci **45**(9): 794-805.

Ogawa, T., S. Inoue, S. Yajima, M. Hidaka and H. Masaki (2006). "Sequence-specific recognition of colicin E5, a tRNA-targeting ribonuclease." Nucleic Acids Res **34**(21): 6065-6073.

Olovnikov, I., K. Chan, R. Sachidanandam, D. K. Newman and A. A. Aravin (2013). "Bacterial argonaute samples the transcriptome to identify foreign DNA." Mol Cell **51**(5): 594-605.

Pan, T. (2018). "Modifications and functional genomics of human transfer RNA." Cell Res **28**(4): 395-404.

Peng, H., J. Shi, Y. Zhang, H. Zhang, S. Liao, W. Li, L. Lei, C. Han, L. Ning, Y. Cao, Q. Zhou, Q. Chen and E. Duan (2012). "A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm." Cell Res **22**(11): 1609-1612.

Poudyal, R. R., P. D. Nguyen, M. P. Lokugamage, M. K. Callaway, J. V. Gavette, R. Krishnamurthy and D. H. Burke (2017). "Nucleobase modification by an RNA enzyme." Nucleic Acids Res **45**(3): 1345-1354.

Rashad, S., X. Han, K. Sato, E. Mishima, T. Abe, T. Tominaga and K. Niizuma (2020). "The stress specific impact of ALKBH1 on tRNA cleavage and tiRNA generation." RNA Biol **17**(8): 1092-1103.

Ren, B., X. T. Wang, J. B. Duan and J. X. Ma (2019). "Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation." Science **365**(6456): 919-+.

Rudland, P. S. and S. K. Dube (1969). "Specific interaction of an initiator tRNA fragment with 30 s ribosomal subunits." J Mol Biol **43**(2): 273-280.

Ruggero, K., A. Guffanti, A. Corradin, V. K. Sharma, G. De Bellis, G. Corti, A. Grassi, P. Zanovello, V. Bronte, V. Ciminale and D. M. D'Agostino (2014). "Small noncoding

RNAs in cells transformed by human T-cell leukemia virus type 1: a role for a tRNA fragment as a primer for reverse transcriptase." J Virol **88**(7): 3612-3622.

Ruggero, K., A. Guffanti, A. Corradin, V. K. Sharma, G. De Bellis, G. Corti, A. Grassi, P. Zanovello, V. Bronte, V. Ciminale and D. M. D'Agostino (2014). "Small Noncoding RNAs in Cells Transformed by Human T-Cell Leukemia Virus Type 1: a Role for a tRNA Fragment as a Primer for Reverse Transcriptase." Journal of Virology **88**(7): 3612-3622.

Saikia, M., R. Jobava, M. Parisien, A. Putnam, D. Krokowski, X. H. Gao, B. J. Guan, Y. Yuan, E. Jankowsky, Z. Feng, G. F. Hu, M. Pusztai-Carey, M. Gorla, N. B. Sepuri, T. Pan and M. Hatzoglou (2014). "Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress." Mol Cell Biol **34**(13): 2450-2463.

Sarker, G., W. F. Sun, D. Rosenkranz, P. Pelczar, L. Opitz, V. Efthymiou, C. Wolfrum and D. Peleg-Raibstein (2019). "Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs." Proceedings of the National Academy of Sciences of the United States of America **116**(21): 10547-10556.

Schaefer, M., T. Pollex, K. Hanna, F. Tuorto, M. Meusburger, M. Helm and F. Lyko (2010). "RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage." Genes Dev **24**(15): 1590-1595.

Schaffer, A. E., V. R. C. Eggens, A. O. Caglayan, M. S. Reuter, E. Scott, N. G. Coufal, J. L. Silhavy, Y. C. Xue, H. Kayserili, K. Yasuno, R. O. Rosti, M. Abdellateef, C. Caglar, P. R. Kasher, J. L. Cazemier, M. A. Weterman, V. Cantagrel, N. Cai, C. Zweier, U. Altunoglu, N. B. Satkin, F. Aktar, B. Tuysuz, C. Yalcinkaya, H. Caksen, K. Bilguvar, X. D. Fu, C. R. Trotta, S. Gabriel, A. Reis, M. Gunel, F. Baas and J. G. Gleeson (2014). "CLP1 Founder Mutation Links tRNA Splicing and Maturation to Cerebellar Development and Neurodegeneration." Cell **157**(3): 651-663.

Scheitl, C. P. M., M. Ghaem Maghami, A. K. Lenz and C. Hobartner (2020). "Site-specific RNA methylation by a methyltransferase ribozyme." Nature **587**(7835): 663-667.

Schimmel, P. (2018). "The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis." Nat Rev Mol Cell Biol **19**(1): 45-58.

Schorn, A. J., M. J. Gutbrod, C. LeBlanc and R. Martienssen (2017). "LTR-Retrotransposon Control by tRNA-Derived Small RNAs." Cell **170**(1): 61-71 e11.

Shabalina, S. A. and E. V. Koonin (2008). "Origins and evolution of eukaryotic RNA interference." Trends Ecol Evol **23**(10): 578-587.

Shao, Y., Q. L. Sun, X. M. Liu, P. Wang, R. Q. Wu and Z. L. Ma (2017). "tRF-Leu-CAG promotes cell proliferation and cell cycle in non-small cell lung cancer." Chemical Biology & Drug Design **90**(5): 730-738.

Shapiro, R. and B. L. Vallee (1987). "Human placental ribonuclease inhibitor abolishes both angiogenic and ribonucleolytic activities of angiogenin." Proc Natl Acad Sci U S A **84**(8): 2238-2241.

Sharma, U., C. C. Conine, J. M. Shea, A. Boskovic, A. G. Derr, X. Y. Bing, C. Belleannee, A. Kucukural, R. W. Serra, F. Sun, L. Song, B. R. Carone, E. P. Ricci, X. Z. Li, L. Fauquier, M. J. Moore, R. Sullivan, C. C. Mello, M. Garber and O. J. Rando (2016). "Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals." Science **351**(6271): 391-396.

Shi, J., Y. Zhang, D. Tan, X. Zhang, M. Yan, Y. Zhang, R. Franklin, M. Shahbazi, K. Mackinlay, S. Liu, B. Kuhle, E. R. James, L. Zhang, Y. Qu, Q. Zhai, W. Zhao, L. Zhao, C. Zhou, W. Gu, J. Murn, J. Guo, D. T. Carrell, Y. Wang, X. Chen, B. R. Cairns, X. L. Yang, P. Schimmel, M. Zernicka-Goetz, S. Cheloufi, Y. Zhang, T. Zhou and Q. Chen (2021). "PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications." Nat Cell Biol **23**(4): 424-436.

Shi, J., Y. Zhang, T. Zhou and Q. Chen (2019). "tsRNAs: The Swiss Army Knife for Translational Regulation." Trends Biochem Sci **44**(3): 185-189.

Soares, A. R., N. Fernandes, M. Reverendo, H. R. Araujo, J. L. Oliveira, G. M. Moura and M. A. Santos (2015). "Conserved and highly expressed tRNA derived fragments in zebrafish." BMC Mol Biol **16**: 22.

Speer, J., C. W. Gehrke, K. C. Kuo, T. P. Waalkes and E. Borek (1979). "tRNA breakdown products as markers for cancer." Cancer **44**(6): 2120-2123.

Srinivasan, S., A. Yeri, P. S. Cheah, A. Chung, K. Danielson, P. De Hoff, J. Filant, C. D. Laurent, L. D. Laurent, R. Magee, C. Moeller, V. L. Murthy, P. Nejad, A. Paul, I. Rigoutsos, R. Rodosthenous, R. V. Shah, B. Simonson, C. To, D. Wong, I. K. Yan, X. Zhang, L. Balaj, X. O. Breakefield, G. Daaboul, R. Gandhi, J. Lapidus, E. Londin, T. Patel, R. L. Raffai, A. K. Sood, R. P. Alexander, S. Das and L. C. Laurent (2019). "Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation." Cell **177**(2): 446-+.

Su, Z., B. Wilson, P. Kumar and A. Dutta (2020). "Noncanonical Roles of tRNAs: tRNA Fragments and Beyond." Annu Rev Genet **54**: 47-69.

Suzuki, T. (2021). "The expanding world of tRNA modifications and their disease relevance." Nat Rev Mol Cell Biol **22**(6): 375-392.

Swarts, D. C., M. M. Jore, E. R. Westra, Y. Zhu, J. H. Janssen, A. P. Snijders, Y. Wang, D. J. Patel, J. Berenguer, S. J. J. Brouns and J. van der Oost (2014). "DNA-guided DNA interference by a prokaryotic Argonaute." Nature **507**(7491): 258-261.

Taxis, T. M., F. V. Bauermann, J. F. Ridpath and E. Casas (2019). "Analysis of tRNA halves (tsRNAs) in serum from cattle challenged with bovine viral diarrhea virus." Genet Mol Biol **42**(2): 374-379.

Thompson, D. M. and R. Parker (2009). "The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae." J Cell Biol **185**(1): 43-50.

Tosar, J. P. and A. Cayota (2020). "Extracellular tRNAs and tRNA-derived fragments." RNA Biol **17**(8): 1149-1167.

Tosar, J. P., F. Gambaro, L. Darre, S. Pantano, E. Westhof and A. Cayota (2018). "Dimerization confers increased stability to nucleases in 5' halves from glycine and glutamic acid tRNAs." Nucleic Acids Res **46**(17): 9081-9093.

Tosar, J. P., C. Rovira and A. Cayota (2018). "Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues." Commun Biol **1**: 2.

Tosar, J. P., M. Segovia, M. Castellano, F. Gambaro, Y. Akiyama, P. Fagundez, A. Olivera, B. Costa, T. Possi, M. Hill, P. Ivanov and A. Cayota (2020). "Fragmentation of extracellular ribosomes and tRNAs shapes the extracellular RNAome." Nucleic Acids Res **48**(22): 12874-12888.

Trotta, C. R., F. Miao, E. A. Arn, S. W. Stevens, C. K. Ho, R. Rauhut and J. N. Abelson (1997). "The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases." Cell **89**(6): 849-858.

Tsuji, T., Y. Q. Sun, K. Kishimoto, K. A. Olson, S. M. Liu, S. Hirukawa and G. F. Hu (2005). "Angiogenin is translocated to the nucleus of HeLa cells and is involved in ribosomal RNA transcription and cell proliferation." Cancer Research **65**(4): 1352-1360.

Tuorto, F., R. Liebers, T. Musch, M. Schaefer, S. Hofmann, S. Kellner, M. Frye, M. Helm, G. Stoecklin and F. Lyko (2012). "RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis." Nat Struct Mol Biol **19**(9): 900-905.

Vitali, P. and T. Kiss (2019). "Cooperative 2'-O-methylation of the wobble cytidine of human elongator tRNA(Met)(CAT) by a nucleolar and a Cajal body-specific box C/D RNP." Genes Dev **33**(13-14): 741-746.

Wang, H., R. Huang, L. Li, J. Zhu, Z. Li, C. Peng, X. Zhuang, H. Lin, S. Shi and P. Huang (2021). "CPA-seq reveals small ncRNAs with methylated nucleosides and diverse termini." Cell Discov **7**(1): 25.

Wang, Q., I. Lee, J. Ren, S. S. Ajay, Y. S. Lee and X. Bao (2013). "Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection." Mol Ther **21**(2): 368-379.

Wang, X., Z. Matuszek, Y. Huang, M. Parisien, Q. Dai, W. Clark, M. H. Schwartz and T. Pan (2018). "Queuosine modification protects cognate tRNAs against ribonuclease cleavage." RNA **24**(10): 1305-1313.

Yamasaki, S., P. Ivanov, G. F. Hu and P. Anderson (2009). "Angiogenin cleaves tRNA and promotes stress-induced translational repression." J Cell Biol **185**(1): 35-42.

Yeung, M. L., Y. Bennasser, K. Watashi, S. Y. Le, L. Houzet and K. T. Jeang (2009). "Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid." Nucleic Acids Res **37**(19): 6575-6586.

Zhang, S., H. Li, L. Zheng, H. Li, C. Feng and W. Zhang (2019). "Identification of functional tRNA-derived fragments in senescence-accelerated mouse prone 8 brain." Aging (Albany NY) **11**(22): 10485-10498.

Zhang, W., C. J. Thieme, G. Kollwig, F. Apelt, L. Yang, N. Winter, N. Andresen, D. Walther and F. Kragler (2016). "tRNA-Related Sequences Trigger Systemic mRNA Transport in Plants." Plant Cell **28**(6): 1237-1249.

Zhang, X., A. E. Cozen, Y. Liu, Q. Chen and T. M. Lowe (2016). "Small RNA Modifications: Integral to Function and Disease." Trends Mol Med **22**(12): 1025-1034.

Zhang, X., F. Trebak, L. A. C. Souza, J. Shi, T. Zhou, P. G. Kehoe, Q. Chen and Y. Feng Earley (2020). "Small RNA modifications in Alzheimer's disease." Neurobiol Dis **145**: 105058.

Zhang, Y., J. Shi, M. Rassoulzadegan, F. Tuorto and Q. Chen (2019). "Sperm RNA code programmes the metabolic health of offspring." Nat Rev Endocrinol **15**(8): 489-498.

Zhang, Y., X. Zhang, J. Shi, F. Tuorto, X. Li, Y. Liu, R. Liebers, L. Zhang, Y. Qu, J. Qian, M. Pahima, Y. Liu, M. Yan, Z. Cao, X. Lei, Y. Cao, H. Peng, S. Liu, Y. Wang, H.

Zheng, R. Woolsey, D. Quilici, Q. Zhai, L. Li, T. Zhou, W. Yan, F. Lyko, Y. Zhang, Q. Zhou, E. Duan and Q. Chen (2018). "Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs." Nat Cell Biol **20**(5): 535-540.

Zhang, Y., Y. Zhang, J. Shi, H. Zhang, Z. Cao, X. Gao, W. Ren, Y. Ning, L. Ning, Y. Cao, Y. Chen, W. Ji, Z. J. Chen, Q. Chen and E. Duan (2014). "Identification and characterization of an ancient class of small RNAs enriched in serum associating with active infection." J Mol Cell Biol **6**(2): 172-174.

Zhou, Z., X. Li, J. Liu, L. Dong, Q. Chen, J. Liu, H. Kong, Q. Zhang, X. Qi, D. Hou, L. Zhang, G. Zhang, Y. Liu, Y. Zhang, J. Li, J. Wang, X. Chen, H. Wang, J. Zhang, H. Chen, K. Zen and C. Y. Zhang (2015). "Honeysuckle-encoded atypical microRNA2911 directly targets influenza A viruses." Cell Res **25**(1): 39-49.

Chapter 2

Endogenous 5′ tsRNA^{Cys} is different from synthetic 5′ tsRNA^{Cys} in translational regulation due to structure differences

2.1 Abstract:

tRNA-derived small RNAs (tsRNAs) are a type of widely existed small non-coding RNA with their mechanisms of action little explored. tsRNAs are highly modified which creates a challenge to study their biological function because synthetic tsRNAs may not mimic their structure and interacting potential. Here, we show that the endogenous tsRNAs with 5′ terminal oligoguanine (endo-5′TOG-tsRNA) extracted from mouse liver tissue are extensively modified. RNA modifications on endo-5′TOG-tsRNA reduce the formation of inter-molecular RNA G-quadruplex (inter-RG4) structure that are dominant in synthetic non-modified 5′-TOG tsRNA (syn-5′TOG-tsRNA). Transfection of endo-5′TOG-tsRNA into Hela cells shows a minor effect on overall cellular transcriptome but a strong effect in regulating the loading of specific mRNAs into ribosome machinery, whereas the syn-5′TOG-tsRNAs have minimized effect.

2.2 Introduction

It has been recently reported that synthetic tsRNAs with 5′-terminal oligoguanine motif (5′-TOG) can form the inter-molecular RNA G-quadruplex (inter-RG4) to regulate translation-related process (Lyons et al. 2017). G-quadruplexes (G4s) are non-canonical nucleic acid secondary structure formed in guanine rich sequence and composed by stacked planar G-quartets, in which four guanine residues are mutually bonded by Hoogsteen hydrogen bond (Varshney et al. 2020). Monovalent cations with adequate ionic radii can further stabilize the guanine plane due to the negatively charged core of G-quartet. For instance, physiologically relevant sodium ion (Na^+) and potassium ion (K^+) can promote G4 assembly, but lithium ion (Li^+) with smaller ionic radii cannot support G4 formation (Malgowska et al. 2016). In the early study, computational analysis has revealed the prevalence of oligoguanine motif containing sequence on the human genome (Huppert et al. 2005, Todd et al. 2005). So, much effort has been invested into the studying of DNA G4s which has been demonstrated to function in diverse biological processes including DNA recombination, genome stability, and telomere maintenance (Varshney et al. 2020). RNA is equally capable of assembling into G4s which has gained growing importance in RNA biology. The RNA G4s (RG4s) have been identified and characterized as key regulators of cellular physiology in the context of health and disease. By newly developed RG4 profiling methods, more than ten thousand RG4 forming regions are detected in human transcriptome (Kwok et al. 2016). These RG4 structure are enriched in the untranslated regions (UTRs) of mRNAs and noncoding-RNA transcripts (Kwok et al. 2018, Kwok et al. 2016, Lee et al. 2020). Even though most RG4s exist in

unfolded conformation due to active RNA helicase activities *in cellulo* (Guo et al. 2016), the existence of RG4s in living cell are also confirmed by antibody-based (Biffi et al. 2014) or small-molecule-based visualization (Chen et al. 2018). Multiple streamlines of evidence favor a model that the folding of RG4s is dynamically controlled by RNA binding proteins (RBPs). For instance, RNA helicase DHX36 first unwinds the RG4 on the mRNAs followed by the binding of hnRNP H/F (heterogeneous nuclear ribonucleoprotein H/F) which keep the RG4 unfolded (Herviou et al. 2020). This 'bind-unwind-lock' mechanism is applied to other RBPs (Benhalevy et al. 2017, Varshney et al. 2020) .

Functional analysis revealed the important role of RG4s in the control of gene expression at post-transcriptional level, such as regulating the translation efficiency of mRNA and its subcellular localization (Dumas et al. 2021). RG4-mediated translation inhibition was first reported in the mRNA of fragile X mental retardation protein (FMRP) by *in vitro* reporter assay that insertion of RG4 forming sequence from FMRP mRNA caused a 1.5-fold reduction in translation in the luciferase reporter (Schaeffer et al. 2001). The enrichment of RG4 near the beginning of 5′ UTRs indicates they have a role in translation initiation regulation (Huppert et al. 2008). The RNA helicase eukaryotic initiation factor 4A (eIF4A) facilitate the recruitments of the 43S pre-initiation complex by unfolding the structured 5′ UTRs of mRNAs (Hinnebusch 2017). Knockdown of eIF4A reduced the translation efficiency of mRNAs with RG4 forming 5′ UTRs, indicating that RNA G4s directly influence the recruitment of the ribosome to the 5′ UTR (Modelska et al. 2015, Wolfe et al. 2014). The interaction of RG4 with RBPs regulates

the subcellular localization of mRNAs that G4-containing mRNAs are more likely to be localized to the stress granule and less translated (Varshney et al. 2020).

tsRNA are naturally interfering with the mRNA translation process since their precursors (tRNA) are one of the key players during this process. tsRNAs from 5′ end of tRNA, specially the 5′-TOG tsRNA, can directly displace the translation initiation factor eIF4A from mRNAs (Lyons et al. 2017) or sequestrate eIF4A in the stress granule by interacting with YBX1 via the RG4 structure to inhibit translation globally (Lyons et al. 2016). Recently a study in human embryonic stem cells (hESCs) showed that the binding affinity of 5′ tsRNAs to translational initiation factors is depended on whether they are modified at the U8 position by pseudouridine mediating by PUS7 (Guzzi et al. 2018). In addition, 5′ tsRNAs can also repress the global translation by targeting translation-related transcripts, such as mRNAs of ribosomal proteins and eukaryotic initiation factors, which requires the AGO proteins (Luo et al. 2018, Shi et al. 2019). 5′ tsRNAs also compete with mRNAs for binding to the 16S rRNA of small ribosomal subunit (30S), impeding mRNA loading into the ribosomal machinery and thus inhibiting the global translation (Gebetsberger et al. 2017). This inhibition effect seems pervasive for the 5′ tsRNAs, however, 3′ tsRNAs are reported to help unfolding the structured coding region of ribosomal proteins mRNAs to accelerate their translation. The increased number of ribosomal proteins promote the ribosome biogenesis and thus resulting in an enhanced global protein synthesis (Kim et al. 2017, Kim et al. 2019).

Chemically synthesized 5′-TOG tsRNAs, such as $tsRNA^{Ala}$ and $tsRNA^{Cys}$, are capable of forming inter-molecular RNA G-quadruplex (inter-RG4) to regulate protein

synthesis by displacing translational initiation factors from the mRNA (Ivanov et al. 2011, Lyons et al. 2017). However, whether endogenous 5′-TOG tsRNA (endo-5′TOG-tsRNA) with natural RNA modifications can also form inter-RG4 or other structures remains an open question, and the mechanism by which they regulate cellular gene expression is still under exploration. More important, is there any other functional mechanism of endogenous 5′-TOG tsRNAs which beyond the RG4 and stress granule mediated translation regulation model. In this chapter, we set out to answer these questions and discovered that, comparing with the synthetic non-modified 5′-TOG tsRNA (syn-5′TOG-tsRNA) which assemble into inter-RG4s dominantly, the endo-5′TOG-tsRNAs are substantially prevented to form the inter-RG4s and they remain a predominant single strand state and some of the endo-5′TOG-tsRNAs also form dimmer conformation. Interestingly, the endo-5′TOG-tsRNA, but not syn-5′TOG-tsRNA, can regulate the loading of specific mRNAs into ribosome machinery, supposedly due to the single strand form facilitates the interaction between endo-5′TOG-tsRNAs and its targeted mRNAs.

2.3 Results

Synthetic RNA sequences with oligoguanine can form RG4 structure in vitro

The four-stranded inter-RG4 structure has been reported to be formed by the 5′-TOG tsRNAs (Figure 2.1a) (Lyons et al. 2017), we confirmed the previous results in the experimental system of our lab (Figure 2.1c). Since, the RG4 forming sequence is pervasive in the non-coding transcripts (Kwok et al. 2016), we also speculated that other small RNA sequences with oligoguanine sequence are capable of forming inter-molecular RG4. To test our hypothesis, we first synthesized the sequence of miR-168 and miR-155, which contain oligoguanine in or close to their 3 terminals, and determined their ability in forming RG4 by the N-Methylmesoporphyrin IX (NMM, a G4 specific ligand) staining on the native PAGE gel. NMM staining results showed that only the miR-155 can form inter-RG4 and the RG4 is the predominate form of the synthetic miR-155. miR-168 which contains three continuous guanine (G) residues in the oligoguanine cannot form RG4. However, miR-168 can get the RG4-forming ability by adding one G to the oligoguanine via an adenosine (A) to guanine (G) mutation in miR-168, as shown by the NMM signal of G4 in the miR-168 (4G) on PAGE gel (Figure 2.1b). On the other hand, disturbing the four continuous G residues in the oligoguanine of miR-155 will eliminate RG4 structure in miR-155 mutants (Figure 2.1b). Next, we checked the RG4 formation in synthetic tsRNAs including tsRNA-Ala, tsRNA-Cys and tsRNA-Pro (Figure 2.1c). The tsRNA-Cys and tsRNA-Ala are belonged to 5′-TOG tsRNAs, and tsRNA-Pro contains oligoguanine in the middle of its sequence. The NMM staining results revealed that all the synthetic tsRNAs with oligoguanine motif formed RG4 structure, but the RG4 signal

of synthetic tsRNA-Pro was weaker than the two synthetic 5′-TOG tsRNAs. From the SYBR Gold staining result, we also found that the ratio of RG4 conformation vs. single-strand conformation is smaller in synthetic tsRNA-Pro comparing to that in 5′-TOG tsRNAs including tsRNA-Cys and tsRNA-Ala. Again, two G to U mutations which disturb the oligoguanine region of synthetic tsRNA-Ala efficiently reduce the RG4 signal in the NMM staining (Figure 2.1c). Next, we performed NMM fluorescence assay with different mutations on the synthetic tsRNA-Ala locate or beyond the oligoguanine motif (Figure 2.1d). Firstly, we found that the interruption of the terminal oligoguanine motif dramatically reduced the signal of NMM fluorescence. The increased uridine contents promoted the signal intensity of NMM fluorescence as shown by the tsRNA-Ala (UUC)/(CUU)/(UUU) mutants, but the increase of cytidine content will downturn the NMM fluorescence signal when the RNA concentration is lower than 30 μM (Figure 2.1d). From these *in vitro* results, it demonstrated that oligoguanine motif which contains at least four guanine residues is necessary for the RG4 formation in small RNAs. The position of the oligoguanine motif may influence the potential of the RG4 forming ability but it’s not decisive.

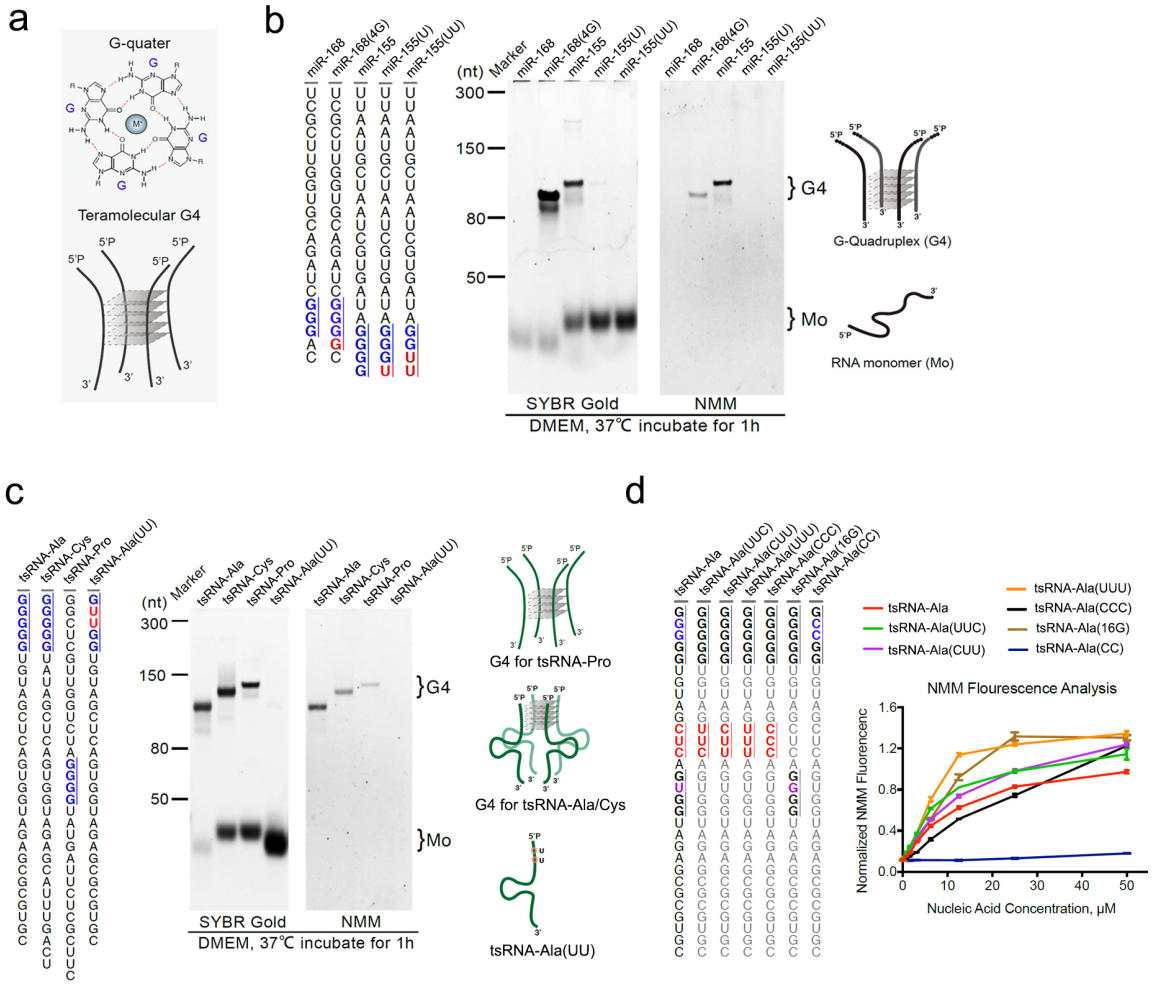


Figure 2.1. **RG4 detection by NMM staining in different oligoguanine containing small RNA sequences**. a: The schematic of inter-molecular four-stranded RG4 structure b: RG4 formation in the miRNA-155, miRNA168 as well as their mutants; c: RG4 formation in TOG containing tsRNAs including tsRNA-Ala and tsRNA-Cys, tsRNA-Pro and the G to U mutant of tsRNA-Ala; d: NMM fluorescence analysis of different mutant of 5′TOG tsRNA^{Ala} under different concentrations.

Endogenous 5′-TOG tsRNA from mouse tissue are highly modified and less RG4s formed

RNA modifications are important both for the biogenesis and function of tsRNAs (Zhang et al. 2016). To obtain insights into the chemical and biological impacts of RNA

modifications on endogenous 5′ tsRNAs, we first set out to develop an affinity-based purification protocol to isolate single endogenous tsRNA from mouse liver tissue based on previous reports (Akiyama et al. 2020, Drino et al. 2020). Using a biotin-labeled DNA probe which is complementary to 5′ tsRNA-Ala/Cys (both belong to classic 5′-TOG tsRNA) and Streptavidin Agarose Beads, we got a mixture of tRNAs and tsRNAs after elution (Figure 2.2a). To identify which RNA band contains the targeted tsRNAs, we performed Northern blot assay with probes specific to 5′ tsRNA^{Ala} and 5′ tsRNA^{Cys} respectively. The Northern blot results showed that the upper band in the tsRNA region is a mixture of endo-tsRNA-Ala and endo-tsRNA-Cys, and the lower band only contains endo-tsRNA-Cys (Figure 2.2b). Next, we retrieved the endogenous tsRNAs from these two bands respectively followed by cDNA synthesis, plasmid ligation and Sanger sequencing. The Sanger sequencing results further confirmed our Northern blot results that the fifteen clones originated from lower band only contain the sequence of endo-tsRNA-Cys. For the 32 sequenced clones from the endo-tsRNA-Ala/Cys mixture, 25 clones contain the sequence of endo-tsRNA-Cys, and 7 clones contain the sequence of endo-tsRNA-Ala with different length (Figure 2.2c).

Next, the purified endo-tsRNA-Cys and the endo-tsRNA-Ala/Cys mixture were processed for RNA modifications quantification using a high-throughput platform based on LC-MS/MS as we previous described (Zhang et al. 2018) (Figure 2.3a). The data from LC-MS/MS has revealed that both endo-tsRNA-Cys and the mixture of endo-tsRNA-Ala/Cys mixture have multiple RNA modifications (Figure 2.3b, c, d, e), including 1-methylguanine (m^1G), 2-methylguanine (m^2G), 1-methyladenosine (m^1A), 5-

methylcytidine (m^5C) and pseudouridine (Ψ) *et.al*. Moreover, these two bands contained quite distinctive pattern of RNA modifications that endo-tsRNA-Cys have higher ratio of modified adenosine (A) and cytidine (C) compared to the endo-tsRNA-Ala/Cys mixture (Figure 2.3c, e). We also observed a very high percentage of inosine, about 30% in the modified A of endo-tsRNA-Cys, which is much higher than what we observed in small RNA fractions from mice liver in our previous study (Shi et al. 2021). Inosine is the deamination product of adenosine catalyzed by the adenosine deaminase acting on RNA (ADAR) enzyme *in vivo* (Roth et al. 2019). This high percentage of inosine may result from the A-to-I editing event during the endogenous 5′ $tsRNA^{Cys}$ biogenesis. However, we could not exclude the possibility that some inosines were the by-products of the RNA digestion process before the LC-MS/MS assay. This is because the enzymes used to digest tsRNAs may contain some deaminase contamination during their purification process (Figure 2.3c). For the modification on C base, endo-tsRNA-Cys contained more m^5C and 2-O-methycytidine (Cm), but less 3-methylcytidine (m^3C) (Figure 2.3e). We found that there are 12.5% G residues are modified in the endo-tsRNA-Cys, and 27.6% in the endo-tsRNA-Ala/Cys mixture. Among the modified G, m^2G is more abundant than the others including m^1G and 2-O-methylguanine (Gm) (Figure 2.3b). For the U bases in the sequence of endo-tsRNA-Cys, they were less modified than the endo-tsRNA-Ala/Cys mixture. The pseduouridine and 2-O-methyluridine (Um) are both lower in the endo-tsRNA-Cys (Figure 2.3d).

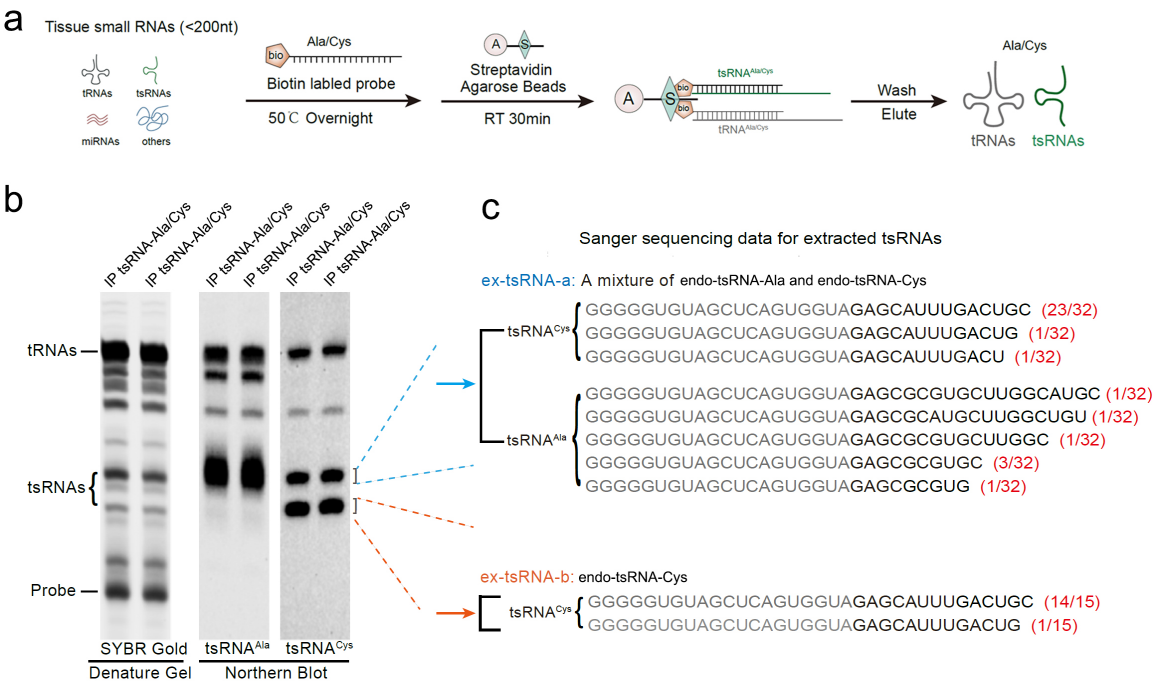
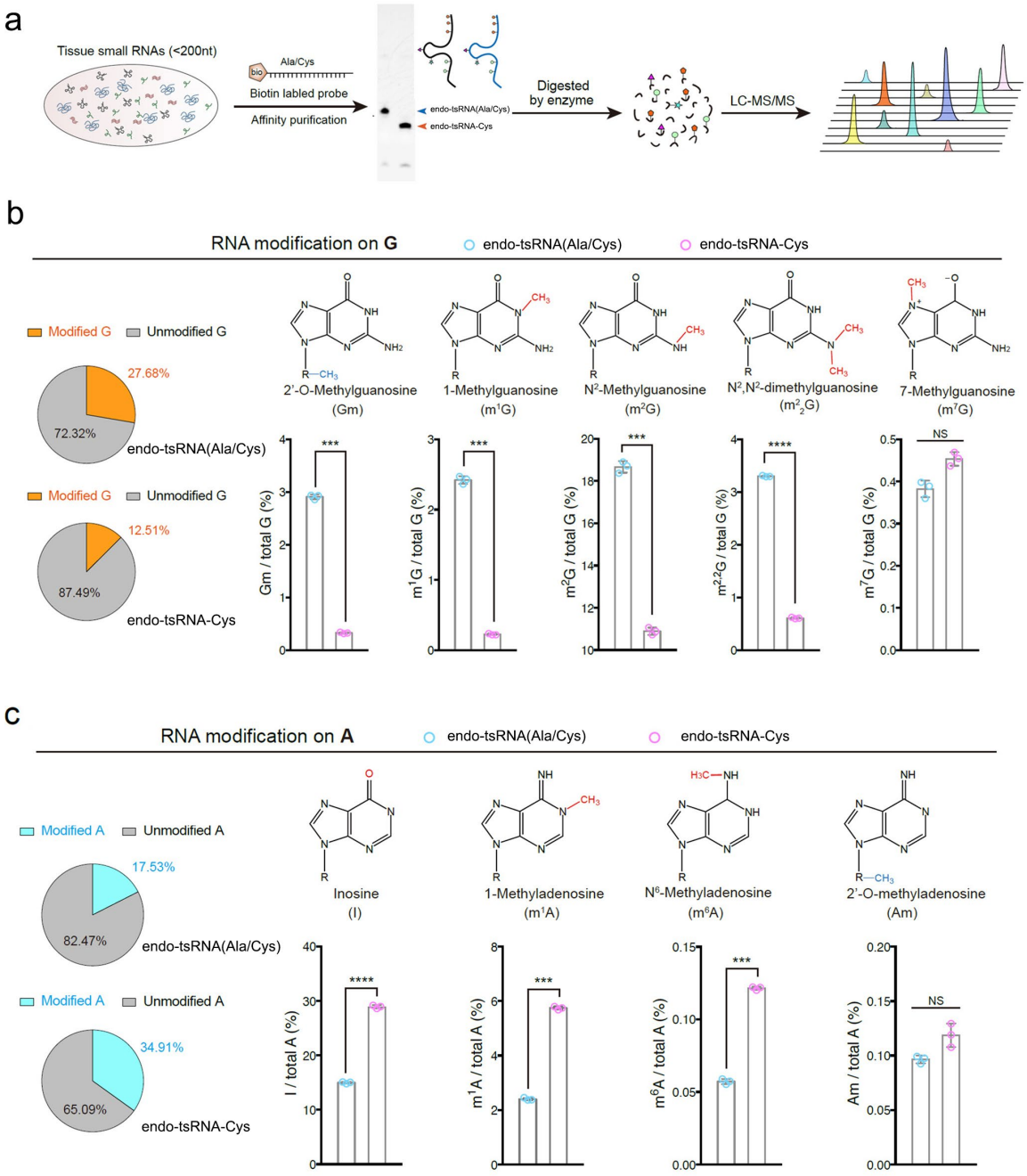


Figure 2.2 Purification of endogenous 5′-TOG tsRNA^{Cys}. a: The schematic of affinity-purification based protocol to purify endogenous tsRNAs. b: Norther blot of the RNA mixture pull-down by biotin-labeled probe from mouse liver tissue. The lower bands of tsRNAs were identified as endo-tsRNA-Cys, the upper band are the mixture of tsRNA-Ala/Cys; c: Sanger sequencing results show that all 15 clones from lower tsRNAs band contain the same sequence of endo-tsRNA-Cys, the clone come from upper tsRNA band, contain sequence of endo-tsRNA-Ala (7/32) and endo-tsRNA-Cys (25/32).


a
Tissue small RNAs (<200nt)
Ala/Cys
Biotin labled probe
Affinity purification
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
Digested by enzyme
LC-MS/MS
b
RNA modification on G
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
Modified G
Unmodified G
27.68%
72.32%
endo-tsRNA(Ala/Cys)
12.51%
87.49%
endo-tsRNA-Cys
2'-O-Methylguanosine (Gm)
1-Methylguanosine (m1G)
N2-Methylguanosine (m2G)
N2,N2-dimethylguanosine (m22G)
7-Methylguanosine (m7G)
Gm / total G (%)
m1G / total G (%)
m2G / total G (%)
m22G / total G (%)
m7G / total G (%)
***
***
***
****
NS
c
RNA modification on A
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
Modified A
Unmodified A
17.53%
82.47%
endo-tsRNA(Ala/Cys)
34.91%
65.09%
endo-tsRNA-Cys
Inosine (I)
1-Methyladenosine (m1A)
N6-Methyladenosine (m6A)
2'-O-methyladenosine (Am)
I / total A (%)
m1A / total A (%)
m6A / total A (%)
Am / total A (%)
****
***
***
NS

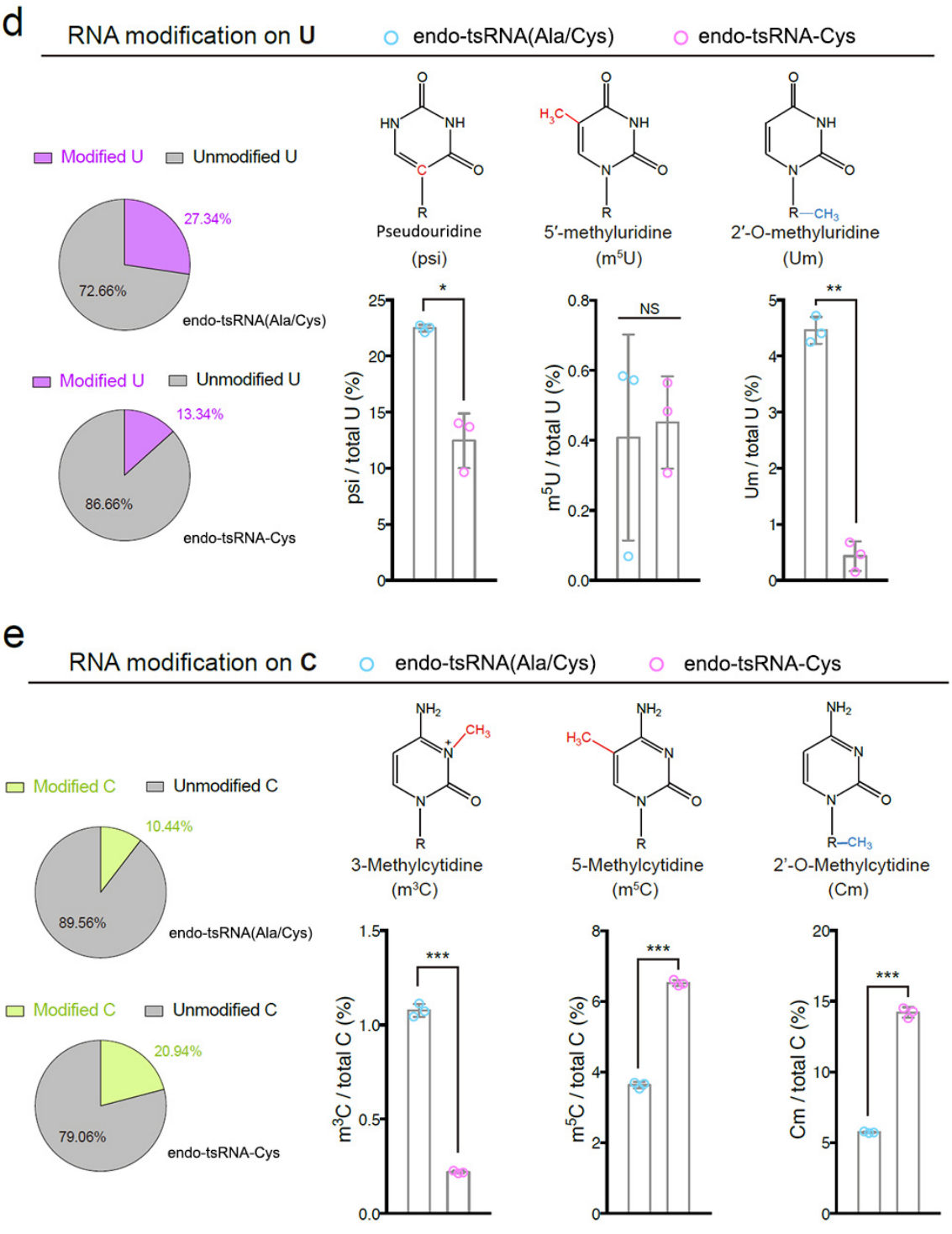

d
RNA modification on U
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
Modified U
Unmodified U
27.34%
72.66%
endo-tsRNA(Ala/Cys)
Modified U
Unmodified U
13.34%
86.66%
endo-tsRNA-Cys
Pseudouridine
(psi)
5'-methyluridine
(m5U)
2'-O-methyluridine
(Um)
psi / total U (%)
m5U / total U (%)
Um / total U (%)
*
NS
**
e
RNA modification on C
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
Modified C
Unmodified C
10.44%
89.56%
endo-tsRNA(Ala/Cys)
Modified C
Unmodified C
20.94%
79.06%
endo-tsRNA-Cys
3-Methylcytidine
(m3C)
5-Methylcytidine
(m5C)
2'-O-Methylcytidine
(Cm)
m3C / total C (%)
m5C / total C (%)
Cm / total C (%)
***
***
***

Figure 2.3: Comprehensive quantification of RNA modifications on endogenous 5′ tsRNA^{Cys} and 5′ tsRNA$^{Ala/Cys}$ mixture. a: The schematic of purify endogenous tsRNA followed by LC-MS/MS; b: The types of modified G and their ratio in the endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys. There are 26.68% and 12.51% modified G in the endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys respectively, among them m^2G are the most abundant modification; c: The types of modified A and their ratio in in the endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys. There are 17.53% and 34.91%% modified A in the endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys respectively; d: The types of modified U and their ratio in the endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys ; There are 27.34% and 17.34% modified U in endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys respectively; e: The types of modified C and their ratio in the endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys; endo-tsRNA-Cys contains more m^5C and Cm but less m^3C compared to endo-tsRNA-Ala/Cys mixture. (All data are plotted as mean ± SEM (n = number of biologically independent experiments). All statistical analysis was performed by two-tailed, one-way Anova, uncorrected Fisher's LSD. NS: not significant; *: $p < 0.05$; **$p<0.01$; ***: $p<0.001$; ****: $p<0.0001$;)

Endogenous 5′-TOG tsRNAs show have reduced ability of forming intermolecular RG4

Although it has been demonstrated that some chemically synthesized tsRNAs, such as 5′ tsRNA-Ala or 5′ tsRNA-Cys can form inter-molecular RNA G-quadruplex (inter-RG4), whether endogenous 5′-TOG tsRNA with various RNA modifications can also form inter-RG4 remains unknown, as no specific endogenous tsRNAs has been extracted for extensive study regarding their RNA modifications, structure, and function. We have successfully purified one endogenous 5′-TOG tsRNA (endo-tsRNA-Cys) via the biotin-based affinity method and quantified multiple types of RNA modifications on it as well as their relative amount. Even though, the exact meanings of these RNA modifications are unknown yet, we speculate these modifications may influence the RG4 forming ability according to the previous experience in our lab that different RNA modifications will lead to different secondary structures of tsRNAs (Chen et al. 2016, Zhang et al. 2018). Next, we explored the RG4 forming ability of endo-tsRNA-Cys purified from mouse liver tissue by both *in vitro* and *in vivo* assay.

As we mentioned above, only suitable cations with adequate ionic radii can support the formation of G4 structure (Malgowska et al. 2016), so we incubated the purified endo-tsRNA-Cys or endo-tsRNA-Ala/Cys mixture in the buffer either containing 100 mM KCl or 100 mM LiCl to test their RG4 formation potential. The results showed that both the endo-tsRNA-Cys and endo-tsRNA-Ala/Cys mixture can form RG4 when the potassium ions (K^+) were added to the incubation buffer, but the endo-tsRNA-Ala/Cys mixture formed much less RG4s compared to the purified endo-tsRNA-Cys that

the intensity of RG4 band of the tsRNA-Ala/Cys mixture is much weaker than the purified endo-tsRNA-Cys (Figure 2.4a). As expected, when lithium ions (Li^+) with smaller ionic radii were added to the incubation buffer, no RG4 band was observed on the native PAGE gel indicates that there were no RG4 formed both in the endo-tsRNA-Ala/Cys mixture and purified endo-tsRNA-Cys. In addition to the RG4 structure, there were dimmer conformation and single-strand conformation co-exist in the endo-tsRNA-Cys, and the single-strand monomer was the predominant existing form of endo-tsRNA-Cys (Figure 2.4a). We also compared the potential of RG4 formation between the synthetic 5′-TOG tsRNAs and purified endogenous 5′-TOG tsRNAs under the same concentration of potassium ion (K^+). For synthetic tsRNA-Ala and tsRNA-Cys, RG4s were the predominate structure conformation that only very few single-strand form existed, and no dimmer conformation were observed in the synthetic 5′-TOG tsRNAs. Also, the synthetic tsRNA-Ala has a higher potential of RG4 forming than the synthetic tsRNA-Cys. However, RG4s formed by endo-tsRNA-Cys are much less than that of synthetic tsRNA-Ala and tsRNA-Cys (Figure 2.4b). These results suggested that the endogenous 5′ $tsRNA^{Cys}$ are quite different from the synthetic 5′ $tsRNA^{Cys}$ in aspect to the RG4 forming ability and structure conformation *in vitro*.

To explore whether the difference of RG4 formation potential between the endogenous 5′ $tsRNA^{Cys}$ and synthetic 5′ $tsRNA^{Cys}$ *in vitro* can also be observed *in vivo*, the synthetic 5′-TOG tsRNAs, including synthetic tsRNA-Cys and synthetic tsRNA-Ala, and purified endo-tsRNA-Cys were transfected into the Hela cells at the concentration of 50 nM by the Lipofectamine™ 3000 reagents followed by an *in vivo* RG4 detection

assay. After 24 hours of transfection, RNA G-quadruplexes formed by tsRNAs in the cytoplasm of Hela cell were detected by an RG4-specific fluorescent probe (QUMA-1) which only bind to the G-quadplex formed by RNA rather than DNA (Chen et al. 2018). In the control groups which only contain the transfection reagents, no obvious RG4 foci were observed. Hela cell transfected by endo-tsRNA-Cys had serval cytoplasmic RG4 foci, but the number is less than that in the synthetic tsRNA-Cys transfected Hela cells (Figure 2.4c). Hela cell transfecting with synthetic tsRNA-Ala has most RG4 foci in the cytoplasm (Figure 2.4c). After statistical analysis, synthetic tsRNA-Cys formed more cytoplasmic RG4 foci than the endo-tsRNA-Cys (5.1 *vs*. 2.7, p=0.016) in the cytoplasm of Hela cell. About 10 cytoplasmic RG4 foci per cell were formed in the Hela cell which were transfected with the synthetic tsRNA-Ala (Figure 2.4c). The difference of RG4 formation potential between synthetic 5′-TOG tsRNAs and endogenous 5′-TOG tsRNAs in Hela cell was consistent with what had been observed on the native PAGE gel *in vitro* (Figure 2.4b).

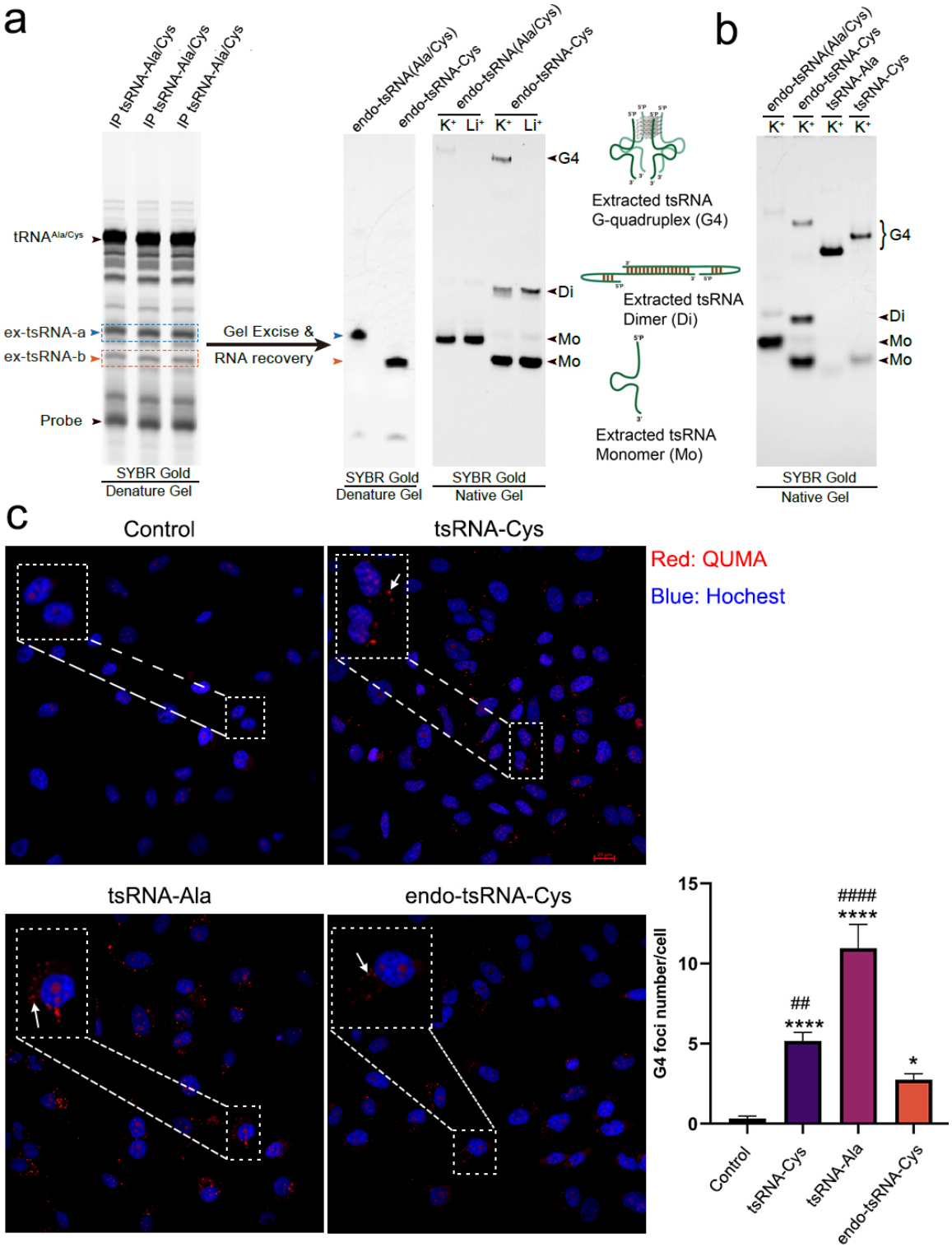

a
IP tsRNA-Ala/Cys
IP tsRNA-Ala/Cys
IP tsRNA-Ala/Cys
tRNA Ala/Cys
ex-tsRNA-a
ex-tsRNA-b
Probe
SYBR Gold
Denature Gel
Gel Excise &
RNA recovery
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
K+ Li+ K+ Li+
G4
Di
Mo
Mo
SYBR Gold
Denature Gel
SYBR Gold
Native Gel
Extracted tsRNA
G-quadruplex (G4)
Extracted tsRNA
Dimer (Di)
Extracted tsRNA
Monomer (Mo)
b
endo-tsRNA(Ala/Cys)
endo-tsRNA-Cys
tsRNA-Ala
tsRNA-Cys
K+ K+ K+ K+
G4
Di
Mo
Mo
SYBR Gold
Native Gel
c
Control
tsRNA-Cys
Red: QUMA
Blue: Hochest
tsRNA-Ala
endo-tsRNA-Cys
G4 foci number/cell
15
10
5
0
##
****
####
****
*
Control
tsRNA-Cys
tsRNA-Ala
endo-tsRNA-Cys

Figure 2.4: Comparison of RG4 formation between the synthetic 5′-TOG tsRNA and endogenous 5′ tsRNA^{Cys} *in vitro* and *in vivo*. a: RG4 formation of endo-tsRNA-Ala/Cys mixture and endo-tsRNA-Cys in different incubation buffers containing K^+ or Li^+. When K^+ are presented, endo-tsRNA-Cys has an obvious RG4 signal on the native PAGE gel, but the endogenous tsRNA-Ala/Cys mixture has a very weak RG4 band. In the Li^+ containing incubation buffer, no RG4 bands were observed on both groups. endo-tsRNA-Cys has bands on the single-stand position and dimmer position. b: RG4 formation of synthetic 5′TOG-tsRNAs and endogenous 5′TOG-tsRNAs in the G4 supporting buffer (K^+). c: RG4 foci detected in the cytoplasm of Hela cell by the QUMA-1, 24h after tsRNA transfection. Control group is empty transfection that no tsRNAs was added to the transfection reagents; the other group were transfected at the concentration of 50nM tsRNAs. Red color represents the signal of QUMA-1; Blue color is the signal of Hochest. Arrowhead indicates the cytoplasmic RG4 foci in the Hela cell. The magnifying view indicated by rectangle is 200% of original view. Note that, there are also strong QUMA-1 signal in the nucleolus which is not background signal. The RG4 foci number were also analyzed statistically. Data are plotted as mean ± SEM (n = 3 independent experiments). All statistical analysis was performed by one-way Anova, uncorrected Fisher's LSD. * mean compared with control group, *: $p < 0.05$; ****: $p<0.0001$; # means compared with endo-tsRNA-Cys group; ##: $p<0.01$; ####: $p<0.0001$;)

Endogenous 5′ tsRNA^{Cys} strongly regulate the loading of specific mRNAs into ribosomes in polysome binding fraction

Because the function of tsRNAs are context dependent and their biological effects range from regulating mRNA cleavage to overall translation efficiency, and to impacting ribosome biogenesis in versatile ways (Chen et al. 2021), we next set out to explore the biological effects of endogenous 5′ tsRNA^{Cys} by transfecting purified endo-tsRNA-Cys into Hela cells, and comparing with Hela cells transfecting with synthetic unmodified tsRNA-Cys. Hela cells with blank transfection will be used as a control.

After transfection, we first performed the full cellular transcriptome (FC) sequencing; then we carried out the polysome profiling assay followed by transcriptomic analyses of the mRNAs from different fractions, namely ribosome-free (RF) fraction, monosome-binding (MB) fraction, and polysome-binding (PB) fraction (Figure 2.5a). The principal component analysis (PCA) on the gene expression data of individual samples showed that mRNAs in the same cellular category/fraction trend to cluster together regardless of the type of tsRNAs they were transfected, which demonstrated the distinct signature of mRNAs from different cellular categories (FC, RF, MB, and PB) (Figure 2.5b). Next, we analyzed the differentially expressed genes in all cellular categories respectively and found that transfection of endo-tsRNA-Cys induced much more dysregulated genes than the transfection of synthetic tRNA-Cys in the sub-fractions including RF, MB and PB fractions. However, the dysregulated genes in the full cellular transcriptome are much less than those in sub-fractions, and there is no significant difference of between the endo-tsRNA-Cys and syn-tsRNA-Cys groups (Figure 2.5c).

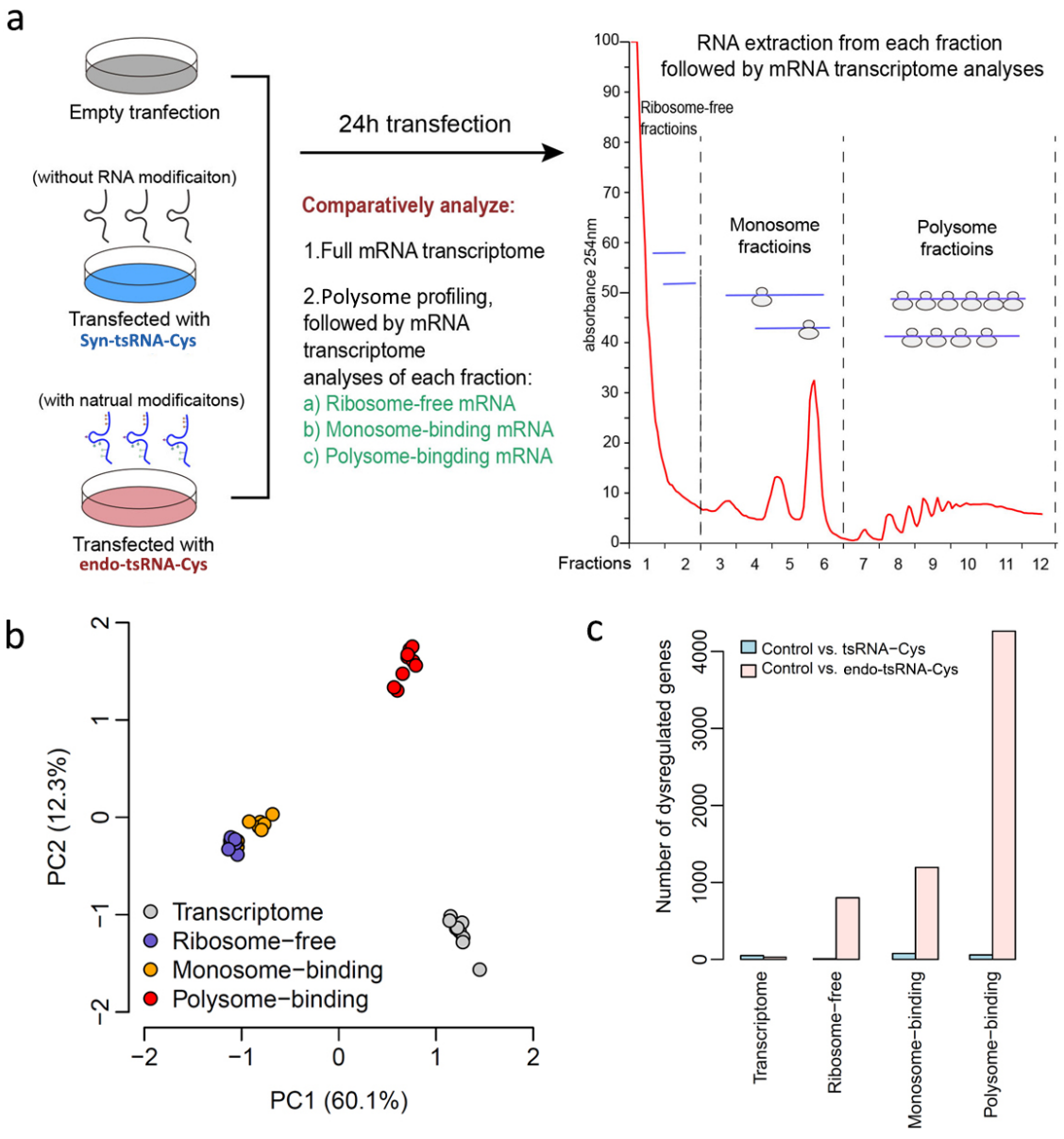


Figure 2.5: Bioinformatic analysis of mRNA from different fraction after endogenous 5′ tsRNA^{Cys} and synthetic 5′ tsRNA^{Cys} transfection. a: Schematic of the transfection of tsRNAs into cell line, followed by either full mRNA transcriptome analysis, or mRNA transcriptome analyses of each fraction (ribosome-free, monosome-binding and polysome-binding). b: Principal component analysis on gene expression of induvial samples; PC1: the first principal component; PC2: the second principal component. c: The number of dysregulated genes induced by tsRNA transfection in each sub-fraction from polysome profiling and full transcriptome.

Since most dysregulated genes, which were induced by endo-tsRNA-Cys transfection, appeared in the polysome binding fraction, we focused on this part in our next analysis and clustered these dysregulated genes by heatmap analysis. The heatmap analysis showed that endo-tsRNA-Cys caused a dramatic change of gene expression pattern in the polysome fraction where mRNAs are being actively translated (Figure 2.6a). Even though synthetic tsRNA-Cys also caused some mRNAs deregulated in the polysome binding fraction (Figure 2.5c), the gene expression pattern induced by synthetic tsRNA-Cys transfection is very similar to the control transfection group overall (Figure 2.6a). To further investigate the mRNAs which were regulated by the transfection of endo-tsRNA-Cys, we performed gene ontology analysis and found that these dysregulated mRNAs were involved in diverse biological processes from molecular regulation to cellular process (Figure 2.6b, c). Among the top 20 pathways which the upregulated mRNAs are preferentially enriched in, we found molecular regulation process related to translation initiation, ribosomal RNA (rRNA) processing and translation (Figure 2.6b). In the downregulated pathways, pathways related to mRNA processing showed up such as mRNA splicing, mRNA transporting and mRNA stability regulation (Figure 2.6c).

These results suggest that endo-tsRNA-Cys has minor effect to regulating gene expression at transcriptional level, but strong effect in regulating the loading of mRNAs into ribosome, thus modulating the sub-pool of mRNAs which will be translated in the polysome fraction. The syn-tsRNA-Cys has no such effect, demonstrating the essential role of RNA modifications for proper tsRNA's structure thus the proper function.

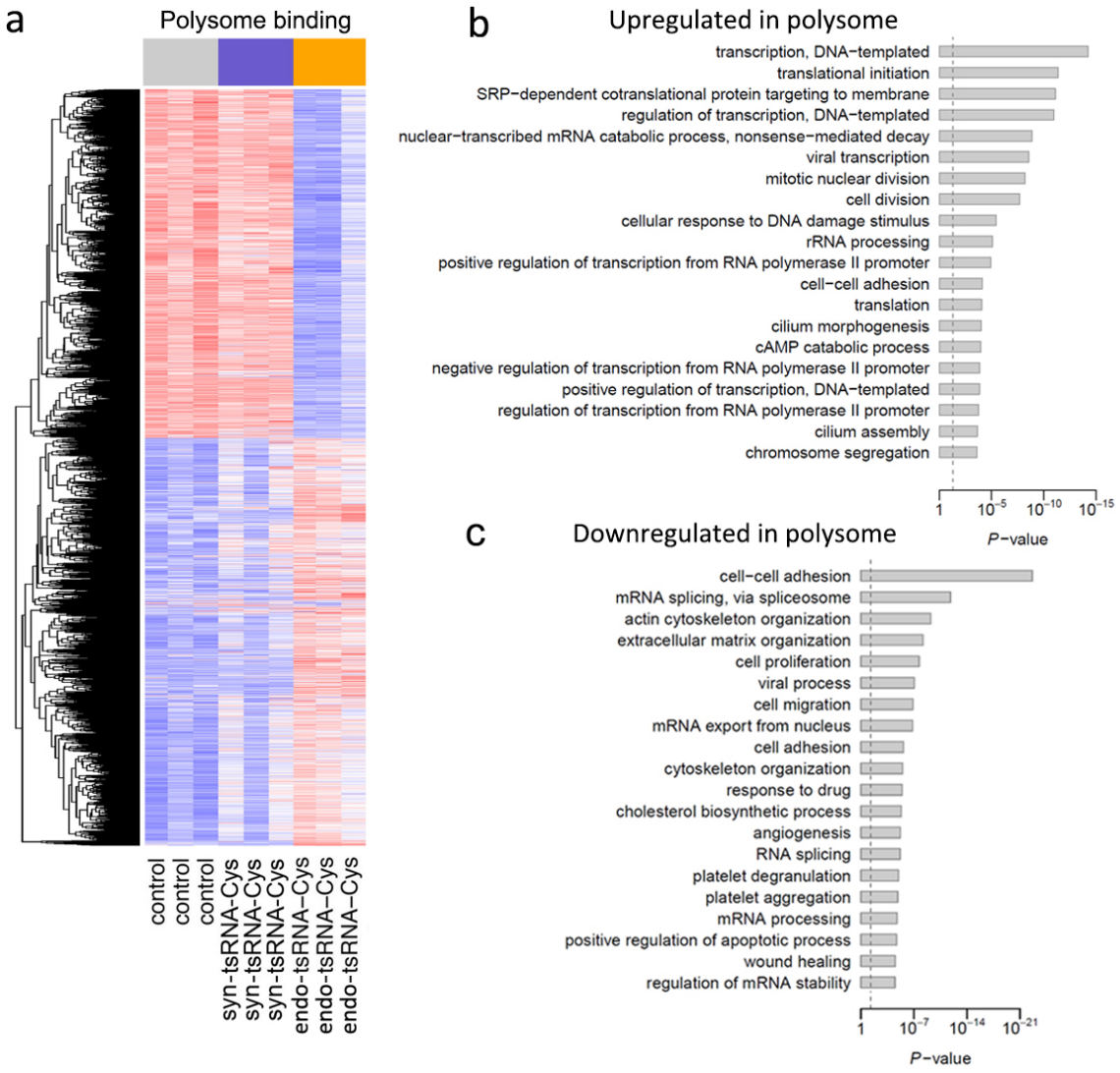


Figure 2.6: Heatmap and gene ontology analysis of dysregulated genes in polysome binding fraction. a: Integrative heatmap of dysregulated genes induced by endo-tsRNA-Cys and synthetic tsRNA-Cys transfection. b: Geno ontology analyses: the top twenty enriched pathways of upregulated mRNAs induced by endo-RNA-Cys in the polysome binding fraction; c: Geno ontology analyses: the top twenty enriched pathways of downregulated mRNAs induced by endo-tsRNA-Cys in the polysome binding fraction.

2.4 Discussion

Our RG4 formation assay *in vitro* showed that oligoguanine which has at least four continuous G residues are necessary for the RNA G-quadruplex formation in the small RNA sequence in regardless of its position on the sequence (Figure 2.1). This will expand the population of small RNAs which are capable of forming inter-molecule RG4, indicating that small RNAs regulating the cellular processes by dynamically controlled RG4 structure may be more general than we thought before (Dumas et al. 2021). The RG4 structure can be unfolded by G4 helicase *in vivo* (Guo and Bartel 2016), and our results showed that only a small part of the endogenous 5′ tsRNA^{Cys} is structured in the RG4 conformation both *in vitro* and *in vivo* (Figure 2.4), thus endogenous RG4-forming small RNAs may need additional factors to maintain the G-quadruplex, such as G4 specific RNA binding proteins, to function (Dumas et al. 2021). We found the primary structure conformation of endo-tsRNA-Cys is single-strand form and there are some dimmer conformation co-existed, which is totally different from the synthetic tsRNA-Cys. Recent study has shown that 5′ tsRNA^{Gly} and 5′ tsRNA^{Glu} existed as dimmer in the cells and the dimerization conferred increased stability to the nuclease in 5′ tsRNAs (Tosar et al. 2018). The dimmer we observed in the endogenous 5′ tsRNA^{Cys} may serve as a reservoir of functional monomer or may have its own function different from the monomer which is unknown yet and needed more investigation.

Due to the great difference of structure between the endo-tsRNA-Cys and synthetic tsRNA-Cys, we preformed tsRNA transfection experiment followed by mRNA transcriptome analysis in different cellular categories (FC, RB, MB, PB). The results

showed that the effects induced by endo-tsRNA-Cys and synthetic tsRNA-Cys are quite different that endogenous tsRNAs have a strong effect on regulating the loading of mRNAs into ribosome, but synthetic tsRNAs do not have such effects (Figure 2.5, 2.6). This functional difference may originate from the structure difference of them, that the single-stranded endo-tsRNA-Cys may directly interact with mRNAs by sequence complementary or interact with the component of ribosome machinery to regulate the mRNA loading process (Chen et al. 2021, Shi et al. 2019). Moreover, when we further analyzed the dysregulated genes induced by endo-tsRNA-Cys, we found some pathways associated with translation initiation, rRNA processing and translation are upregulated (Figure 2.6b), and mRNA processing pathways are downregulated (Figure 2.6c). The enriched upregulated pathways relating to translational processes may not be a direct result induced by endo-tsRNA-Cys. One of the possible scenarios is that the unoccupied ribosomes will load the mRNAs according to their abundance after tsRNAs impede some mRNAs loading to ribosomes. That is, more abundant mRNAs in the cells get more chance to be loaded into the ribosome if they are not targeted by endogenous single stand 5′ $tsRNA^{Cys}$. This may could explain why the pathways relating to the transcription and translation process showed up in the upregulated pathways (Figure 2.6b), because these two processes are the fundamental activities for cell maintenance that mRNAs involved in these two processes are quite abundant in the cells.

Recent studies reported that ribosomes are heterogenous and different ribosomes can preferentially translate specific mRNAs, namely ribosome-heterogeneity, to control some important cellular processes and to play a role in human disease (Gay et al. 2021).

Ribosome heterogeneity can be resulted by the different ribosome proteins or rRNA components (Genuth et al. 2018, Simsek et al. 2017). tsRNA were reported to influence the translation of some specific ribosomal proteins (Kim et al. 2017) and the 18s rRNA processing (Kim et al. 2019), so it is possible that specific tsRNA can regulate cellular process via inducing ribosome heterogeneity in addition to the translation initiation regulation or stress granules formation (Chen et al. 2021). This will be a new ear waits to explore. We also have an interesting finding that there were strong RG4 signal in the nucleolus where the rRNA are actively transcribed which strongly indicates that RG4 structure may be actively involved in the rRNA transcription and processing. Supporting this idea, 28s rRNA were reported to form RG4 structure *in vitro* recently (Mestre-Fos et al. 2019). Moreover, RG4 formed by 28sRNA *in vivo* may regulate heme bioavailability in HEK293 cell (Mestre-Fos et al. 2020). However, more investigation is needed to elucidate the role of RG4s in the rRNA biogenesis and function.

2.5 Materials and Methods

Animals

Animal experiments were conducted under the protocol and approval of the institutional animal care and use committees of the University of California, Riverside. Mice were given access to food and water ad libitum and were maintained on a 12 h light/12 h dark artificial lighting cycle. Mice were housed in cages at a temperature of 22–25 °C, with 40–60% humidity.

Detect RNA G4 structure by Native PAGE gel

100 ng synthesized TOG RNAs or endo-TOG RNA were added into the incubation buffer, which was made up either by DMEM (Invitrogen™) or by the mixture of 100 mM Tris-HCl (pH 7.5, Invitrogen™), 100 mM KCl or 100 mM LiCl. Then tsRNAs/miRNAs were incubated at 37 ℃ for 1h. After the incubation, the RNA mixture was then gently mixed well with 10× RNA loading buffer (Invitrogen™) and 2.5 μl sample was loaded into 10 % native PAGE gel in each well. Run the PAGE gel in the 1× TBE buffer (Invitrogen™) for 70 min under 70V at 4 ℃ by using Bio-Rad Mini Gel Running system. To detect the G4 specific bands on the gel, firstly stain the gel with 50 μM N-methyl mesoporphyrin IX (NMM, Santa Cruz Biotechnology), and image the gel by ChemiDoc™ MP Imaging System under chemiluminescence working mode. After that, stain the gel with 1× SYBR Gold Nucleic Acid Gel Stain (Invitrogen™) and image the gel again to show all the RNA bands on the native gel. All synthetic RNA sequences are in **Table 2.1**.

Table 2.1 Sequence of all synthetic RNAs

Name	Sequence
tsRNA-Ala	5'P-GGGGGUGUAGCUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala (UU)	5'-UUGGGUGUAGCUCAGUGGUAGAGCG CGUGC-3'OH
tsRNA-Ala (1,m^1G)	5'P-(m1G)GGGGUGUAGCUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala (3,m1G)	5'P-GG(m1G)GGUGUAGCUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(1,3,m1G)	5'P-(m1G)G(m1G)GGUGUAGCUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(4m1G)	5'P-(m1Gm1Gm1Gm1G)GUGUAGCUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala (UUC)	5'P-GGGGGUGUAGUUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(CUU)	5'P-GGGGGUGUAGCUUAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(UUU)	5'P-GGGGGUGUAGUUUAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(UUU)	5'P-GGGGGUGUAGCCCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(16G)	5'P-GGGGGUGUAGCUCAGGGGUAGAGCGCGUGC-3'OH
tsRNA-Ala(CC)	5'P-GCCGGUGUAGCUCAGUGGUAGAGCGCGUGC-3'OH
tsRNA-Cys	5'P-GGGGGUGUAGUUCAGUGGUAGAGCAUUUGACU-3'OH
tsRNA-Cys(UU)	5'P-GUUGGUGUAGCUCAGUGGUAGAGCAUUUGACU-3'OH
tsRNA-Cys(4m1G)	5'P-(m1Gm1Gm1Gm1G)GUGUAGCUCAGUGGUAGAGCAUUUGACU-3'OH
tsRNA-Pro	5'P-GGCUCGUUGGUCUAGGGGUAUGAUUCUCGCUUC-3'OH
miRNA-155	5'P-UUAAUGCUAAUCGUGAUAGGGG-3'OH
miRNA-155(U)	5'P-UUAAUGCUAAUCGUGAUAGGGU-3'OH
miRNA-155(2U)	5'P-UUAAUGCUAAUCGUGAUAGGGU-3'OH
miRNA-168	5'P-UCGCUUGGUGCAGAUCGGGAC-3'OH
miRNA-168(4G)	5'P-UCGCUUGGUGCAGAUCGGGGC-3'OH

NMM Fluorescence analysis.

Fluorescence assays were performed in 30 μL of 10 mM sodium phosphate buffer (pH 6.4), 100 mM KCl, 4 mM $MgCl_2$, and 5 μM NMM. The oligonucleotides concentration ranged from 0 to 50 μM. All fluorescence experiments were performed using a FlexStation III (Molecular Devices) plate reader with excitation and emission wavelengths of 399 nm and 614 nm, respectively. Fluorescence measurements were repeated three times for each sample, and the intensities were averaged and corrected by running a buffer control without RNA before each series of experiments.

Isolation and validation of endo-5′tsRNA^{Cys}

The endo-5′ tsRNA-Cys was purified by affinity pulldown assay combined with gel recovery, with modified protocols from the previous report (Akiyama et al. 2020, Drino et al. 2020). The total RNAs of mouse liver were harvest by TRIzol^{TM} reagent (Invitrogen^{TM} 15596026) as per manufacturer instructions. The concentration of total RNAs solution was adjusted to 2mg/ml by RNase-free water. Then small RNAs (<200nt) were separated by the buffer contains 50%(w/v) PEG 8000 and 0.5M NaCl solution via centrifuging at 12000rpm at 4 °C for 20mins. The supernatant was collected followed by adding 1/10 volume NaAc solution (Invitrogen^{TM}). 1 ml supernatant was added with 3 ml Ethanol, and 5 μl Linear Acrylamide (Invitrogen^{TM}) to precipitate small RNAs (<200 nt) with -20 °C overnight incubation followed by centrifugation at 12000rpm at 4 °C for 20 min. Small RNA (<200 nt) solution was adjusted to 1mg/ml, 1 ml small RNA solution with 6 μl biotinylated probe(100 μM), 26 μl 20× SSC solution (Invitrogen^{TM}) and 15 μl

RNase inhibitor (NEB) were incubated at 50 °C overnight. 200 µl Streptavidin Sepharose (Cytiva 17511301) was added to the hybridization solution to enrich the biotin labeled 5′ tsRNA^{Cys} probe that captured with the targeted RNA. After incubation at room temperature for 30mins, the Streptavidin Sepharose was transferred to the 1.5 ml Ultrafree-MC tube (Millipore, UFC30GV0S) and washed by 0.5× saline-sodium citrate (SSC) buffer, the washing step was repeated five times. 500 µl nuclease-free water was added to the MC tube and incubated at 70°C for 15 min followed by centrifugation at 2500g at room temperature for 1 min to elute the RNAs that are complementary to the biotinylated probe. Repeat this step for three times and get 1.5ml eluent followed by adding 150 µl NaAc solution, 4.5ml ethanol and 20 µl linear acrylamide to precipitate the enriched RNAs with -20 °C overnight incubation followed by centrifugation at 12000rpm at 4 °C for 20 min. Add 30 µl nuclease-free water to dissolve the RNA pallets. Take 1µl RNA solution to perform the Norther blot to identify the band of purified endo-5′tsRNA^{Cys}. After Northern blot validation, run the 7M Urea-PAGE gel and cut the main tsRNA bind and perform the gel recovery to get purified endo-5′tsRNA^{Cys} or endo-tsRNA-Ala/Cys mixture. To further confirm the purity of endo-5′tsRNA^{Cys}, we also applied sanger sequencing. Briefly, the extracted endo-5′tsRNA^{Cys} was reverse-transcribed into cDNAs and ligated into the plasmids by pGEM®-T Easy Vector Systems (Promega Cat# A1360). The plasmids contained the target sequence were then transformed into E. coli for amplification. After that, the harvest plasmids were sent to sanger sequencing. The data were analyzed by SnapGene Viewer software.

Northern blot

RNA was separated by 10% Urea-PAGE gel and stained with SYBR Gold for 10 min at room temperature. The gel was immediately imaged and then transferred to positively charged Nylon membranes (Roche) followed by UV crosslink with an energy of 0.12 J. Membranes were pre-hybridized with DIG Easy Hyb solution (Roche) for 1h at 42 °C. To detect tsRNAs, membranes were incubated overnight (12–16h) at 42 °C with DIG-labelled oligonucleotides probes synthesized by Integrated DNA Technologies (IDT). The membranes were washed twice with low stringent buffer (2× SSC with 0.1% (wt/vol) SDS) at 42 °C for 15 min each, rinsed twice with high stringent buffer (0.1× SSC with 0.1% (wt/vol) SDS) for 5 min each, and rinsed in washing buffer (1× SSC) for 10 min. Following the washes, the membranes were transferred into 1× blocking buffer (Roche) and incubated at room temperature for 3 h, after which the Anti-Digoxigenin-AP Fab fragments (Roche,) was added into the blocking buffer at a ratio of 1:10,000 and incubated for an additional 30mins at room temperature. The membranes were washed four times with DIG washing buffer (1× maleic acid buffer, 0.3% Tween-20) for 15 min each, followed by incubation in DIG detection buffer (0.1 M TrisHCl, 0.1 M NaCl, pH 9.5) for 5 min, and then coated with CSPD ready-to-use reagent (Roche), incubated in the dark for 30 min at 37 °C before imaging with ChemiDoc™ MP Imaging System (Bio-Rad). Digoxigenin labeled Northern blot probe for tsRNA detection were synthesized by IDT and the sequence was listed below:

DIG-tsRNA^{Cys} probe: 5′-DIG-AGTCAAATGCTCTACCAC

DIG-tsRNA^{Ala} probe: 5′-DIG-GCACGCGCTCTACCACTG

LC-MS/MS based RNA modification analysis for in endo-5′tsRNA^{Cys}

Hydrolysis of RNA samples were performed as previously described with some modification(Su et al. 2014). Purified small RNAs (500-1000ng) were digested with 9.1 U Benzonase (Sigma-Aldrich), 0.045 U phosphodiesterase I (USB) and 0.45 U alkaline phosphatase (Sigma-Aldrich) in 37 ℃ for 3h. And then, the enzymes in digestion mixtures were removed by centrifugation through Nanosep 3K device with Omega membrane (Pall). The samples were analyzed on Ultra 3000 liquid chromatography system (AB, Rockwell) equipped with a triple Quadrupole mass spectrometer. A Waters ACQUITY HSS T3 1.8 μm column (150 × 2.1 mm) was used at 30 °C. Water with 0.1% (v/v) formic acid as mobile phase A, acetonitrile containing 0.1% (v/v) formic acid as mobile phase B, the flow rate is 0.40 ml/min and the gradient is as follows:
0 → 6 min: 0% B; 6 → 6.5 min: 0 → 0.5% B; 6.5 → 9.0 min: 0.5% → 0.8% B; 9.0 → 11.0 min: 0.8% B; 11.0 → 11.5 min: 0.8% → 2.0% B; 11.5 → 30 min: 2.0% → 4.0% B; 30 → 31 min: 4.0% → 0% B; 31 → 36 min: 0% B.

The total run time for an LC–MS/MS analysis was 31 minutes. Nucleosides were detected by a TSQ Vantage triple-quadrupole mass spectrometer (ThermoFisher) equipped with an electrospray ionization (ESI) source. The ESI source was set in a positive ionization mode, and an ion-spray voltage of +5000 V was applied. The heated capillary temperature was set at 282 °C. Nitrogen was applied as sheath and auxiliary gas at a pressure of 60 and 55 arbitrary units, respectively. Data were collected in selected reaction monitoring (SRM) mode, with one m/z transition per analyte. SRM transitions was listed in the Table 1. Chromatographic data acquisition was performed using the

Xcalibur software (v2.0.7 SP1, Thermofisher). The percentage of each modified ribonucleoside was normalized to the total amount of quantified ribonucleosides with the same nucleobase, which decreases/eliminates errors caused by sample loading variation. For example, the percentage of m^5C = mole concentration (m^5C)/mole concentration ($m^5C + m^3C + Cm + C$). SRM transition were list in **Table 2.2**.

Table2.2: SRM transitions

	Identification	Q1/Q3	Retention time (min)	CE	S-Lens	Retention window (min) Start	Retention window (min) Stop
1	C	244.1/112.1	2.38	12	70	0	7.50
2	m^1A	282.1/150.1	4.70	18	85	0	17.00
3	Ψ(pseudouridine)	245.1/125.0	2.1	17	65	0	20.00
4	m^3U	259.1/127.1	15.11	12	77	0	28.80
5	m^7G	298.1/166.1	7.9	10	70	0	28.90
6	$m^{227}G$	326.1/194.1	22.75	17	93	0	29.07
7	$m^{2,2}G$	312.1/180.1	24.78	16	84	0	29.30
8	U	245.1/113.0	4.67	6	56	0	30
9	Cm	258.1/112.1	8.14	12	63	0	30
10	m^3C	258.1/126.1	3.7	14	70	0	30
11	m^5C	258.1/126.1	4.45	14	69	0	30
12	Um	259.1/113.1	14.6	6	55	0	30
13	m^5U	259.1/127.1	11.27	12	77	0	30
14	A	268.1/136.1	9.07	17	80	0	30
15	I	269.1/137.1	10.33	13	65	0	30
16	Am	282.1/136.2	15.95	9	86	0	30
17	m^6A	282.1/150.1	18.04	9	90	0	30
18	G	284.0/152.0	10.50	14	72	0	30
19	m^1G	298.1/166.1	16.80	10	70	0	30
20	m^2G	298.1/166.1	18.37	16	71	0	30
21	Gm	298.1/152.1	17.2	14	71	0	28.88
22	cAMP(IS)	335.0/136.0	11.42	26	109	0	30

tsRNA transfection and RG4 detection by QUMA-1

Hela cell was cultured in the DMEM with 10% Fetal Bovine Serum (FBS) and 1× antibiotics at 37 °C and 5% CO_2. The day before transfection, Hela cell was seeded in glass bottom chamber system (Lab-Tek). The transfection complex was prepared as follows: 1μl respective tsRNAs (10 μM) with 1μl Lipofectamine3000 and 20 μl Opti-MEM was mixed by vortexing and incubated at room temperature for 15 min. The media was discarded and 180 μl new Hela media (excluding antibiotics) was added to the wells. Then lipofectamine–RNA transfection complex was added to the wells and incubated for 24 h at 37 °C under 5% CO_2. For each transfection, three independent replicates were used. Vehicle-only transfection was used as a control.

Hela cells were fixed with 4% paraformaldehyde at room temperature for 15 min. After rinsing with nuclease-free PBS (pH 7.4), cells were permeabilized in 0.5% TritonX-100 at room temperature for 15 min. Cells were stained with 1μM QUMA-1 and 1ug/ml Hochest33342 for 30 min at 37 °C in dark. Wash the cells with nuclease-free PBS for 5 min at room temperature in dark. The emission of QUMA-1 was collected under excitation at 561nm, and signal of the Hochest33342 was collected under excitation at 361nm. Digital images were captured using a LSM700 laser scanning confocal microscope (Zeiss) with a 40 X objective lens. The images were analyzed with Imaris software (V9.0.1, Bitplane Corp) that RG4 foci were detected by the build-in vesicle-detection function in Imaris. After that, the number of G4 foci in each cell will be exported as Excel file. Then the statistics of RG4 foci number were performed with GraphPad Prism (v 8.1.0, GraphPad software LLC).

Polysome profiling and RNA extraction from different fractions

The polysome profiling steps were followed as previously described(Simsek et al. 2017). Briefly, Hela cells transfected with synthesized tsRNAs or endo-tsRNAs was treated by cycloheximide (100 μg/ml) for 5 min before harvest and followed by cold PBS (50 μg/ml cycloheximide) washing. Lysis the cell in the culture dish by adding 400μl lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 15 mM $MgCl_2$, 100 μg/ml cycloheximide, 1 mM DTT, 1% Triton X-100, 0.1 mg/ml heparin, 20 U/ml TURBO DNase, 200 U/mL SUPERase In RNase Inhibitor, 1x Combined Protease and Phosphatase Inhibitor). Incubate the cell on ice for 5 min, and then scrape the cell down the slope into the lysis buffer pooled in the lower corner. Collect the lysis buffer into 1.5 ml centrifuge tube and incubate on ice for 30 min. Triturate cells 10 times through a 26G gauge needle during the incubation. The lysate was clarified by centrifugation for 10 min at 10,000rpm and 4°C to remove nuclei and mitochondria. Transfer the supernatant into a new tube and store on ice for immediately sample running or store at -80 ℃ for future sample running. Cleared cell lysate was then loaded onto a 15%–50% sucrose gradient and centrifuged at 40,000 rpm and 4°C for 2.5 hours. Gradients were fractioned using a Brandel gradient fractionator system with continuous A260 measurement. Collected the fractions corresponding to the free RNA part (fraction 1 and 2), monosome-binding (fraction 3-6) and polysome-binding (fraction 7-13). Total RNAs are extracted from the fractions by TRIzol™ LS Reagent as per manufacturer instructions. The mRNAs for full cellular transcriptome were collected directly by adding TRIzol^{TM} reagent to Hela cells at

the 24-hour timepoint after tsRNA transfection. The extracted RNA was sent to Novogene to perform RNA-seq.

Transcriptome sequencing and data analysis

Transcriptome libraries were constructed using the TruSeq Stranded mRNA Library Prep Kit (Illumina). For each RNA library, 4 Gigabytes base pairs (raw data) were generated by Illumina Hi‐Seq 4000. After base composition and quality tests were passed, we removed the adapter sequence. The clean reads were used for downstream bioinformatics analysis. All mRNA library preparation, quality examination and RNA sequencing were performed by Novogene. For the data analysis, the kallisto program (Bray et al. 2016) were used to quantify mRNA abundance from sequencing data. For the dysregulated genes, groupwise comparison was performed using the edgeR package (Robinson et al. 2010) and the genes with false discovery rate (FDR) <1% and FC >1.5 will were deemed as differentially expressed. For the gene ontology analysis, the GSEA algorithm(Subramanian et al. 2005) will be applied to the GO databases (Harris et al. 2004). GSEA determines whether an a priori defined set of genes (e.g. translation initiation) show statistically significant (adjusted P <0.05 after Benjamini-Hochberg procedure), concordant expression differences between two biological states (e.g. control vs. endo-tsRNA^{Cys} transfection).

2.6 Reference

Akiyama, Y., P. Kharel, T. Abe, P. Anderson and P. Ivanov (2020). "Isolation and initial structure-functional characterization of endogenous tRNA-derived stress-induced RNAs." RNA Biol **17**(8): 1116-1124.

Benhalevy, D., S. K. Gupta, C. H. Danan, S. Ghosal, H. W. Sun, H. G. Kazemier, K. Paeschke, M. Hafner and S. A. Juranek (2017). "The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation." Cell Rep **18**(12): 2979-2990.

Biffi, G., M. Di Antonio, D. Tannahill and S. Balasubramanian (2014). "Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells." Nat Chem **6**(1): 75-80.

Bray, N. L., H. Pimentel, P. Melsted and L. Pachter (2016). "Near-optimal probabilistic RNA-seq quantification." Nat Biotechnol **34**(5): 525-527.

Chen, Q., W. Yan and E. Duan (2016). "Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications." Nat Rev Genet **17**(12): 733-743.

Chen, Q., X. Zhang, J. Shi, M. Yan and T. Zhou (2021). "Origins and evolving functionalities of tRNA-derived small RNAs." Trends Biochem Sci **46**(10): 790-804.

Chen, X. C., S. B. Chen, J. Dai, J. H. Yuan, T. M. Ou, Z. S. Huang and J. H. Tan (2018). "Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells." Angew Chem Int Ed Engl **57**(17): 4702-4706.

Drino, A., V. Oberbauer, C. Troger, E. Janisiw, D. Anrather, M. Hartl, S. Kaiser, S. Kellner and M. R. Schaefer (2020). "Production and purification of endogenously modified tRNA-derived small RNAs." RNA Biol **17**(8): 1104-1115.

Dumas, L., P. Herviou, E. Dassi, A. Cammas and S. Millevoi (2021). "G-Quadruplexes in RNA Biology: Recent Advances and Future Directions." Trends Biochem Sci **46**(4): 270-283.

Gay, D. M., A. H. Lund and M. D. Jansson (2021). "Translational control through ribosome heterogeneity and functional specialization." Trends Biochem Sci.

Gebetsberger, J., L. Wyss, A. M. Mleczko, J. Reuther and N. Polacek (2017). "A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress." RNA Biol **14**(10): 1364-1373.

Genuth, N. R. and M. Barna (2018). "The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life." Mol Cell **71**(3): 364-374.

Guo, J. U. and D. P. Bartel (2016). "RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria." Science **353**(6306).

Guzzi, N., M. Ciesla, P. C. T. Ngoc, S. Lang, S. Arora, M. Dimitriou, K. Pimkova, M. N. E. Sommarin, R. Munita, M. Lubas, Y. Lim, K. Okuyama, S. Soneji, G. Karlsson, J. Hansson, G. Jonsson, A. H. Lund, M. Sigvardsson, E. Hellstrom-Lindberg, A. C. Hsieh and C. Bellodi (2018). "Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells." Cell **173**(5): 1204-1216 e1226.

Harris, M. A., J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. White and C. Gene Ontology (2004). "The Gene Ontology (GO) database and informatics resource." Nucleic Acids Res **32**(Database issue): D258-261.

Herviou, P., M. Le Bras, L. Dumas, C. Hieblot, J. Gilhodes, G. Cioci, J. P. Hugnot, A. Ameadan, F. Guillonneau, E. Dassi, A. Cammas and S. Millevoi (2020). "hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma." Nat Commun **11**(1): 2661.

Hinnebusch, A. G. (2017). "Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation." Trends Biochem Sci **42**(8): 589-611.

Huppert, J. L. and S. Balasubramanian (2005). "Prevalence of quadruplexes in the human genome." Nucleic Acids Res **33**(9): 2908-2916.

Huppert, J. L., A. Bugaut, S. Kumari and S. Balasubramanian (2008). "G-quadruplexes: the beginning and end of UTRs." Nucleic Acids Res **36**(19): 6260-6268.

Ivanov, P., M. M. Emara, J. Villen, S. P. Gygi and P. Anderson (2011). "Angiogenin-induced tRNA fragments inhibit translation initiation." Mol Cell **43**(4): 613-623.

Kim, H. K., G. Fuchs, S. Wang, W. Wei, Y. Zhang, H. Park, B. Roy-Chaudhuri, P. Li, J. Xu, K. Chu, F. Zhang, M. S. Chua, S. So, Q. C. Zhang, P. Sarnow and M. A. Kay (2017).

"A transfer-RNA-derived small RNA regulates ribosome biogenesis." Nature **552**(7683): 57-62.

Kim, H. K., J. Xu, K. Chu, H. Park, H. Jang, P. Li, P. N. Valdmanis, Q. C. Zhang and M. A. Kay (2019). "A tRNA-Derived Small RNA Regulates Ribosomal Protein S28 Protein Levels after Translation Initiation in Humans and Mice." Cell Rep **29**(12): 3816-3824 e3814.

Kwok, C. K., G. Marsico and S. Balasubramanian (2018). "Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome." Cold Spring Harb Perspect Biol **10**(7).

Kwok, C. K., G. Marsico, A. B. Sahakyan, V. S. Chambers and S. Balasubramanian (2016). "rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome." Nat Methods **13**(10): 841-844.

Lee, D. S. M., L. R. Ghanem and Y. Barash (2020). "Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations." Nat Commun **11**(1): 527.

Luo, S. Q., F. He, J. J. Luo, S. Q. Dou, Y. R. Wang, A. N. Guo and J. Lu (2018). "Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response." Nucleic Acids Research **46**(10): 5250-5268.

Lyons, S. M., C. Achorn, N. L. Kedersha, P. J. Anderson and P. Ivanov (2016). "YB-1 regulates tiRNA-induced Stress Granule formation but not translational repression." Nucleic Acids Res **44**(14): 6949-6960.

Lyons, S. M., D. Gudanis, S. M. Coyne, Z. Gdaniec and P. Ivanov (2017). "Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs." Nat Commun **8**(1): 1127.

Malgowska, M., K. Czajczynska, D. Gudanis, A. Tworak and Z. Gdaniec (2016). "Overview of the RNA G-quadruplex structures." Acta Biochim Pol **63**(4): 609-621.

Mestre-Fos, S., C. Ito, C. M. Moore, A. R. Reddi and L. D. Williams (2020). "Human ribosomal G-quadruplexes regulate heme bioavailability." J Biol Chem **295**(44): 14855-14865.

Mestre-Fos, S., P. I. Penev, S. Suttapitugsakul, M. Hu, C. Ito, A. S. Petrov, R. M. Wartell, R. Wu and L. D. Williams (2019). "G-Quadruplexes in Human Ribosomal RNA." J Mol Biol **431**(10): 1940-1955.

Modelska, A., E. Turro, R. Russell, J. Beaton, T. Sbarrato, K. Spriggs, J. Miller, S. Graf, E. Provenzano, F. Blows, P. Pharoah, C. Caldas and J. Le Quesne (2015). "The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape." Cell Death Dis **6**: e1603.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.

Roth, S. H., E. Y. Levanon and E. Eisenberg (2019). "Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity." Nat Methods **16**(11): 1131-1138.

Schaeffer, C., B. Bardoni, J. L. Mandel, B. Ehresmann, C. Ehresmann and H. Moine (2001). "The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif." EMBO J **20**(17): 4803-4813.

Shi, J., Y. Zhang, D. Tan, X. Zhang, M. Yan, Y. Zhang, R. Franklin, M. Shahbazi, K. Mackinlay, S. Liu, B. Kuhle, E. R. James, L. Zhang, Y. Qu, Q. Zhai, W. Zhao, L. Zhao, C. Zhou, W. Gu, J. Murn, J. Guo, D. T. Carrell, Y. Wang, X. Chen, B. R. Cairns, X. L. Yang, P. Schimmel, M. Zernicka-Goetz, S. Cheloufi, Y. Zhang, T. Zhou and Q. Chen (2021). "PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications." Nat Cell Biol **23**(4): 424-436.

Shi, J., Y. Zhang, T. Zhou and Q. Chen (2019). "tsRNAs: The Swiss Army Knife for Translational Regulation." Trends Biochem Sci **44**(3): 185-189.

Simsek, D., G. C. Tiu, R. A. Flynn, G. W. Byeon, K. Leppek, A. F. Xu, H. Y. Chang and M. Barna (2017). "The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity." Cell **169**(6): 1051-1065 e1018.

Su, D., C. T. Chan, C. Gu, K. S. Lim, Y. H. Chionh, M. E. McBee, B. S. Russell, I. R. Babu, T. J. Begley and P. C. Dedon (2014). "Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry." Nat Protoc **9**(4): 828-841.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A **102**(43): 15545-15550.

Todd, A. K., M. Johnston and S. Neidle (2005). "Highly prevalent putative quadruplex sequence motifs in human DNA." Nucleic Acids Res **33**(9): 2901-2907.

Tosar, J. P., F. Gambaro, L. Darre, S. Pantano, E. Westhof and A. Cayota (2018). "Dimerization confers increased stability to nucleases in 5' halves from glycine and glutamic acid tRNAs." Nucleic Acids Res **46**(17): 9081-9093.

Varshney, D., J. Spiegel, K. Zyner, D. Tannahill and S. Balasubramanian (2020). "The regulation and functions of DNA and RNA G-quadruplexes." Nat Rev Mol Cell Biol **21**(8): 459-474.

Wolfe, A. L., K. Singh, Y. Zhong, P. Drewe, V. K. Rajasekhar, V. R. Sanghvi, K. J. Mavrakis, M. Jiang, J. E. Roderick, J. Van der Meulen, J. H. Schatz, C. M. Rodrigo, C. Zhao, P. Rondou, E. de Stanchina, J. Teruya-Feldstein, M. A. Kelliher, F. Speleman, J. A. Porco, Jr., J. Pelletier, G. Ratsch and H. G. Wendel (2014). "RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer." Nature **513**(7516): 65-70.

Zhang, X., A. E. Cozen, Y. Liu, Q. Chen and T. M. Lowe (2016). "Small RNA Modifications: Integral to Function and Disease." Trends Mol Med **22**(12): 1025-1034.

Zhang, Y., X. Zhang, J. Shi, F. Tuorto, X. Li, Y. Liu, R. Liebers, L. Zhang, Y. Qu, J. Qian, M. Pahima, Y. Liu, M. Yan, Z. Cao, X. Lei, Y. Cao, H. Peng, S. Liu, Y. Wang, H. Zheng, R. Woolsey, D. Quilici, Q. Zhai, L. Li, T. Zhou, W. Yan, F. Lyko, Y. Zhang, Q. Zhou, E. Duan and Q. Chen (2018). "Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs." Nat Cell Biol **20**(5): 535-540.

Chapter 3

Engineered 5′ tsRNA^{Cys} mimic endogenous 5′ tsRNA^{Cys} both in structure and biological function

3.1 Abstract

Endogenous 5′ tsRNA^{Cys} and synthetic 5′ tsRNA^{Cys} showed significant difference in their structure and biological function. Since the key difference between endogenous 5′ tsRNA^{Cys} and synthetic 5′ tsRNA^{Cys} is whether they form RG4 structure or not, we make two single-stand mutants of synthetic 5′ tsRNA^{Cys}, which lose the RG4 forming ability, by adding G to U point mutation or incorporating 4 m^1G modifications to the oligoguanine region of synthetic 5′ tsRNA^{Cys}. The engineered 5′ tsRNA^{Cys} mutants form much less RG4 foci in the Hela cell and induced similar effects of endogenous 5′ tsRNA^{Cys} in translational regulation after transfection into Hela cells.

3.2 Results

RNA modifications on guanine impeded the RG4 formation of synthetic 5′-TOG tsRNAs

The modified endogenous 5′ tsRNA^{Cys} form much less RG4 than the non-modified synthetic 5′ tsRNA^{Cys} (Chapter 2, Figure 2.4b), so we speculated that by adding site specific RNA modifications can disrupt the RNA G-quadruplex structure formed by oligoguanine in the RNA sequence. There are five RNA modifications commonly detected on the guanosine residue of RNA including m^1G, m^2G, $m^2{}_2G$, m^7G and Gm (Boccaletto et al. 2018). In the planar G-quartet, the atoms at the position 1,2,7 of guanosine are directly involved in the Hoogsteen hydrogen bond formation which hold the G-quartet (Figure 3.1a). We tested the impacts of different RNA modifications, such as m^1G, m^2G and Gm, on disrupting the RG4 formation using synthetic 5′-TOG tsRNA-Ala in our initial screening assay. We didn't test the m^7G and $m^2{}_2G$ modifications because these two modifications are not commercially available.

We tested different combinations of modification number and position on the terminal oligoguanine of synthetic 5′tsRNA-Ala, including one modification to the first or third guanine residues, two modifications to the first and third guanine residues, as well as four modifications to the first four guanine residues (Figure 3.1b, c, d). By *in vitro* G4 forming assay, we found that m^1G is the most effective modification to unfold the RNA G4-quadruplxe that four m^1G modifications on the terminal oligoguanine can totally disrupt the RG4 formation of 5′-TOG tsRNAs, and only very weak RG4 signal

were remained in the mutants with one m^1G modification or two m^1G modifications (Figure 3.1b). The effects of m^2G on disrupting RG4 forming are more complicated that it seems depend on the numbers and the position of the m^2G modifications. We observed more RG4 retained when two or four m^2G modifications were added to the oligoguanine of 5′tsRNA-Ala compared to the variants with one m^2G added (Figure 3.1c). It has been identified that m^2G happened at position 10 of the mammalian $tRNA^{Ala}$ (Boccaletto et al. 2018), so we also synthesized a variant with the m^2G at the position 10 of 5′tsRNA-Ala and found that this m^2G modification at position 10 which beyond the oligoguanine motif also can diminish RNA G-quadruplex formation, but this RG4-destroying ability is weaker than the m^2G modification at position 1 in oligoguanine motif (Figure 3.1c). The Gm modification happened on the atom located on the ring of ribose which didn't involve in the formation of Hoogsteen hydrogen directly, so it should be less potentially to block the RNA G-quadruplex formation theoretically. Our results supported this idea that only incorporating four Gm modifications can dramatically reduce the RG4 structure of synthetic 5′ tsRNA-Ala. The other modified variants still formed substantial amount of RG4s (Figure 3.1d).

Since the RNA modifications can influence the RNA stability, and the endogenous tsRNAs are more stable than the synthetic tsRNAs after incubation with zygote lysis or mouse serum (Chen et al. 2016, Zhang et al. 2014). So, we also performed the RNA degradation assay in our screening assay to determine the stability of the modified synthetic 5′tsRNA-Ala under RNase treatment with different modifications. We first compared the stability between endo-5′tsRNA-Cys, G4 forming synthetic 5′tsRNA-

Ala and non-G4-froming synthetic 5′tsRNA-Ala (UU) mutant under RNase A/T1 treatment. Short bands still can be observed on the gel after endo-5′tsRNA-Cys were treated with RNase A/T1, however there were no band detected in the synthetic 5′tsRNA-Ala (UU) mutants (Figure 3.1e). These results suggested that endo-5′tsRNA-Cys is more resistant to the cleavage of RNase A/T1 in some extent compared to the synthetic tsRNAs. For the G4-forming 5′tsRNA-Ala, there are dispersive bands on the gel which may be the vestige of RNA G-quadruplex (Figure 3.1c). Next, we treated the m^1G or Gm modified 5′tsRNA-Ala mutants with RNase A/T1 and found neither m^1G modification nor Gm modification can protect the synthetic 5′tsRNA-Ala from the cleavage of RNase A/T1(Figure 3.1 f, g). This may be because the RNase A/T1 cleave the nucleotide strand from 3′ end, and our modification are added at the 5′ end, so it cannot protect the RNA strand from the cleavage of RNase A/T1.

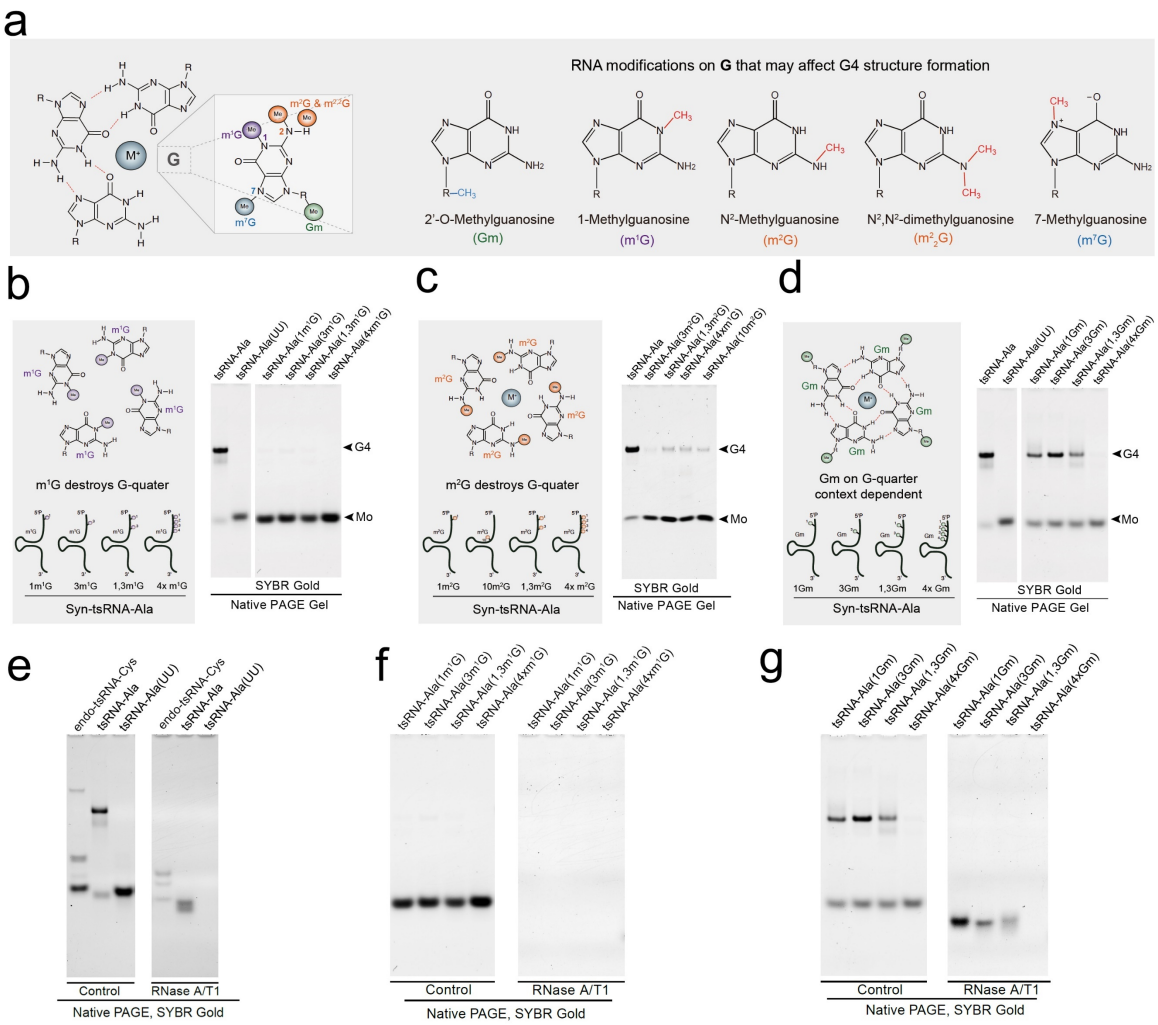
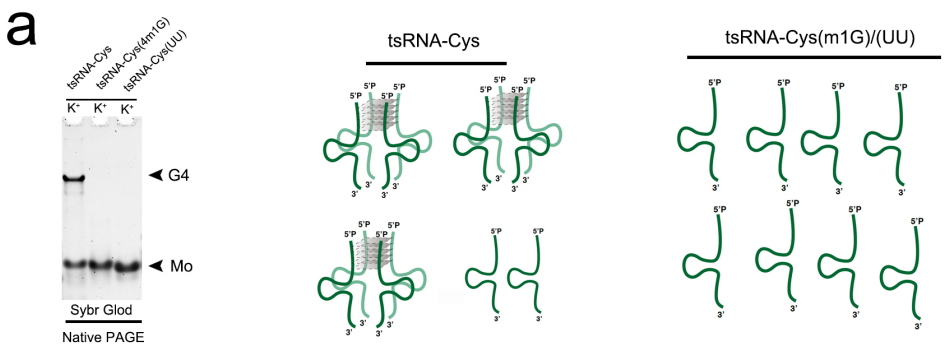


Figure 3.1: Impacts of RNA modifications on the RG4 formation of synthetic TOG tsRNAs. a: The Hoogsteen hydrogen bonds in the G-quarter planer and five common modifications on the guanine residue. b~c: the effects of m^1G, m^2G and Gm on the G4 formation of synthetic tsRNA-Ala; e: degradation assay of endo-tsRNA-Cys, synthetic tsRNA-Ala and tsRNA-Ala (UU) mutants; f: degradation assay of tsRNA-Ala with m^1G modification; g: degradation assay of tsRNA-Ala with Gm modification.

Engineered 5′ tsRNA^{Cys} are similar to endogenous 5′ tsRNA^{Cys} in structure

Since the four m^1G modifications and two G to U mutations are the most effective methods to disrupt the RG4 structure of 5′tsRNA-Ala in our screening assay, we decided to engineer the synthetic 5′tsRNA-Cys by these two methods, the resulting engineered tsRNA-Cys named 5′tsRNA-Cys (4m1G) and 5′tsRNA-Cys (UU). As shown in the Figure 3.2a, there were no RG4 band in native PAGE gel of the engineered tsRNA-Cys mutants when incubated with G4-supporting buffer which contains 100 mM potassium chloride (K^+), the synthetic 5′tsRNA-Cys form a substantial amount of RG4 structures. Moreover, the single-strand form is predominated in the engineered tsRNA-Cys mutants which are similar to the endo-5′tsRNA-Cys as we tested before (Chapter 2, Figure 2.4).

Next, we also transfected the endo-5′tsRNA-Cys, 5′tsRNA-Cys (4m1G), 5′tsRNA-Cys (UU) as well as two synthetic 5′tsRNA-Ala variants into Hela cells and comparatively compared their RG4-forming ability *in vivo*. The results showed that, the 5′tsRNA-Cys (4m1G) and 5′tsRNA-Cys (UU) didn't formed RG4 foci in the cytoplasm of Hela cells, the endo-5′tsRNA-Cys form less RG4 foci than the synthetic 5′tsRNA-Cys. Hela cells transfecting with synthetic tsRNA-Ala have the most G4 foci among all tested the groups (Figure 3.2b). Interestingly, we observed sporadic RG4 foci signal in the 5′tsRNA-Ala (UU) transfection group, but didn't observe the RG4 structure in the PAGE gel of tsRNA-Ala (UU) (Figure 3.1b, e). We also calculated of the average RG4 foci number in each Hela cells, the average RG4 foci number is no difference between control group and two engineered 5′tsRNA-Cys transfection groups (Figure 3.2c)


a
tsRNA-Cys
tsRNA-Cys(4m1G)
tsRNA-Cys(UU)
K+
K+
K+
G4
Mo
Sybr Glod
Native PAGE
tsRNA-Cys
tsRNA-Cys(m1G)/(UU)
5'P
3'

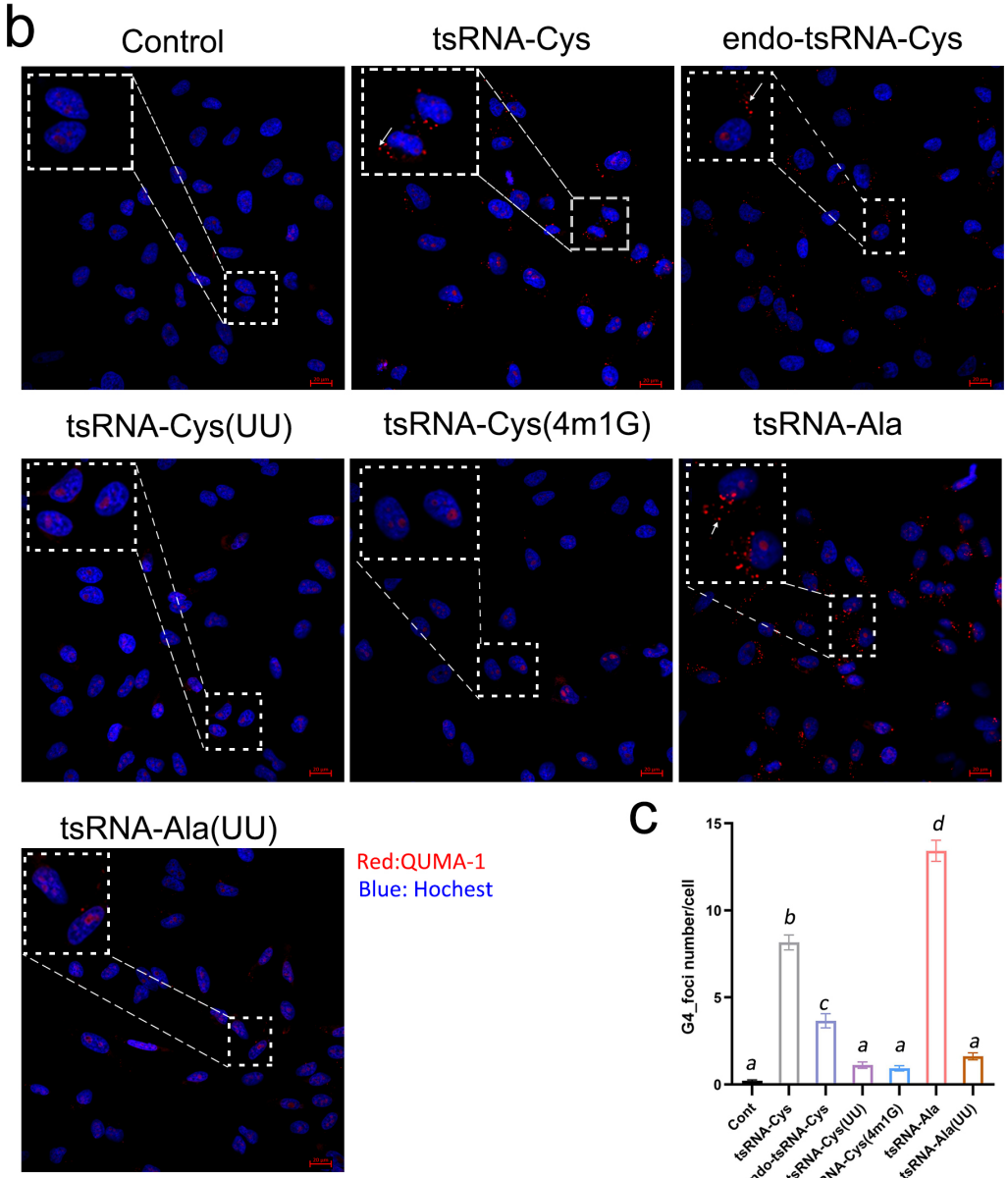

b
Control
tsRNA-Cys
endo-tsRNA-Cys
tsRNA-Cys(UU)
tsRNA-Cys(4m1G)
tsRNA-Ala
tsRNA-Ala(UU)
Red:QUMA-1
Blue: Hochest
c
G4_foci number/cell
15
10
5
0
a
b
c
a
a
d
a
Cont
tsRNA-Cys
endo-tsRNA-Cys
tsRNA-Cys(UU)
tsRNA-Cys(4m1G)
tsRNA-Ala
tsRNA-Ala(UU)

Figure 3.2 RG4 formation of engineered 5′ tsRNA^{Cys} *in vitro* and *in vivo*. a: RG4 detection by native PAGE gel of synthetic tsRNA-Cys, tsRNA-Cys (4m1G) and tsRNA-Cys (UU) in G4 supporting buffer (100 mM KCl); b: G4 foci detected in the cytoplasm of Hela cell by RG4 specific probe QUMA-1 at 24h timepoint after the transfection of different synthetic tsRNAs (tsRNA-Cys and tsRNA-Ala) and endo-tsRNA-Cys. Control group is empty transfection that no tsRNAs was added to the transfection reagents; the other group were transfected at the concentration of 50 nM tsRNAs. Red color represents the signal of QUMA-1; Blue color is the signal of Hochest33342. Arrowhead indicates the cytoplasmic RG4 foci in the Hela cell. Images are at 400 X magnification, the magnifying view indicated by rectangle is 200% of original view; c: The average G4 foci number in each cell. Image analyzed by Imaris software and data are plotted as mean ± SEM (n = 3 independent experiments). All statistical analysis was performed by one-way ANOVA, *p* value was corrected by Tukey's multiple comparison test. If the letter assigned to the column on the top is different to another column, it represents that the mean of these two groups is significantly different in statistic. The detailed *p* values are listed in the **Table 3.1**.

Table 3.1: Adjust *p* value of G4 foci number comparison

Tukey's multiple comparisons test	Summary	Adjusted P Value
Cont vs. tsRNA-Cys	****	<0.0001
Cont vs. tsRNA-Cys (m1G)	ns	0.7323
Cont vs. tsRNA-Cys (UU)	ns	0.5007
Cont vs. endo-tsRNA-Cys	****	<0.0001
Cont vs. tsRNA-Ala	****	<0.0001
Cont vs. tsRNA-Ala (UU)	ns	0.1699
tsRNA-Cys vs. tsRNA-Cys (m1G)	****	<0.0001
tsRNA-Cys vs. tsRNA-Cys (UU)	****	<0.0001
tsRNA-Cys vs. endo-tsRNA-Cys	****	<0.0001
tsRNA-Cys vs. tsRNA-Ala	****	<0.0001
tsRNA-Cys vs. tsRNA-Ala (UU)	****	<0.0001
tsRNACys (4m1G) vs. tsRNA-Cys (UU)	ns	0.9998
tsRNA-Cys (4m1G) vs. endo-tsRNA-Cys	****	<0.0001
tsRNA-Cys (4m1G) vs. tsRNA-Ala	****	<0.0001
tsRNA-Cys (4m1G) vs. tsRNA-Ala (UU)	ns	0.9017
tsRNA-Cys (UU) vs. endo-tsRNA-Cys	****	<0.0001
tsRNA-Cys (UU) vs. tsRNA-Ala	****	<0.0001
tsRNA-Cys (UU) vs. tsRNA-Ala (UU)	ns	0.9787
endo-tsRNA-Cys vs. tsRNA-Ala	****	<0.0001
endo-tsRNA-Cys vs. tsRNA-Ala (UU)	*	0.011

Transfection of engineered tsRNA^{Cys} induced similar biological effect of endo-tsRNA^{Cys}

We transfected the two engineered 5′tsRNA-Cys mutants, tsRNA-Cys (4m1G) and tsRNA-Cys (UU), into Hela cells and then analyzed the mRNAs in different fractions as we performed in Chapter 2. Combined with the transcriptome data from the synthetic tsRNA-Cys and endo-tsRNA-Cys transfection, we performed a comparative bioinformatic analysis (Figure 3.3). First, we performed the PCA analysis again with gene expression data from all samples. The results were consistent with our previous analysis that mRNAs in the same cellular category trend to cluster together regardless of the type of tsRNAs they were transfected with. From the comprehensive PCA analysis, we found that some samples in monosome binding category are overlapped with the samples from ribosome free fraction (Figure 3.3a).

The engineered tsRNA-Cys mutants show similar effect of endo-tsRNA-Cys in dysregulating the mRNAs in the polysome fraction which are under active translation. tsRNA-Cys (4m1G) and tsRNA-Cys (UU) both induced a huge number of differentially expressed genes in the polysome binding fraction (Figure 3.3b). Moreover, only few genes showed changes in the full cellular transcriptome after engineered tsRNA-Cys mutants were transfected, similar to the results we observed in Chapter 2, suggesting that engineered tsRNA-Cys mutants also regulate gene expression at post-transcriptional level too. However, the engineered tsRNA-Cys mutants were also different from endo-tsRNA-Cys in some extent. tsRNA-Cys (UU) has more dysregulated genes in the polysome binding fraction, but it only caused a few differentially expressed genes in the ribosome-

free fraction which is much less than those induced by endo-tsRNA-Cys and tsRNA-Cys (4m1G) (Figure 3.3b). The tsRNA-Cys (4m1G) induced more dysregulated genes in the monosome binding fraction than the endo-tsRNA-Cys, but it induced fewer dysregulated genes than endo-tsRNA-Cys in polysome binding fraction (Figure 3.3b).

With the focus on the mRNAs in polysome binding fraction, we clustered the dysregulated genes into a comprehensive heatmap which including the control, tsRNA-Cys, endo-tsRNA-Cys and two engineered tsRNA-Cys mutants. The heatmap showed that mRNAs with low expression in the control and synthetic tsRNA-Cys group were upregulated after the transfection of engineered tsRNA-Cys mutants which is similar to the effect of endo-tsRNA-Cys transfection (Figure 3.3c). The gene overlap analysis showed that there are 1476 upregulated genes overlapped between the endo-tsRNA-Cys and tow engineered tsRNA-Cys mutants, but only 28 upregulated genes were overlapped with synthetic tsRNA-Cys (Figure 3.3c, up). Similar situations also happened in the downregulated genes that the number of overlapped downregulated genes is 1009 vs. 4 (Figure 3.3c, bottom). We also analyzed the overlapped genes in ribosome free and monosome binding fractions, the trend was comparable to what we found in polysome binding fraction (Figure 3.4 a, b), but the number of overlapped gene was much smaller than that in polysome binding fraction. In addition, there were very few genes overlapped in the full cellular transcriptomes which due to the limited dysregulated gene number in this category (Figure 3.4c).

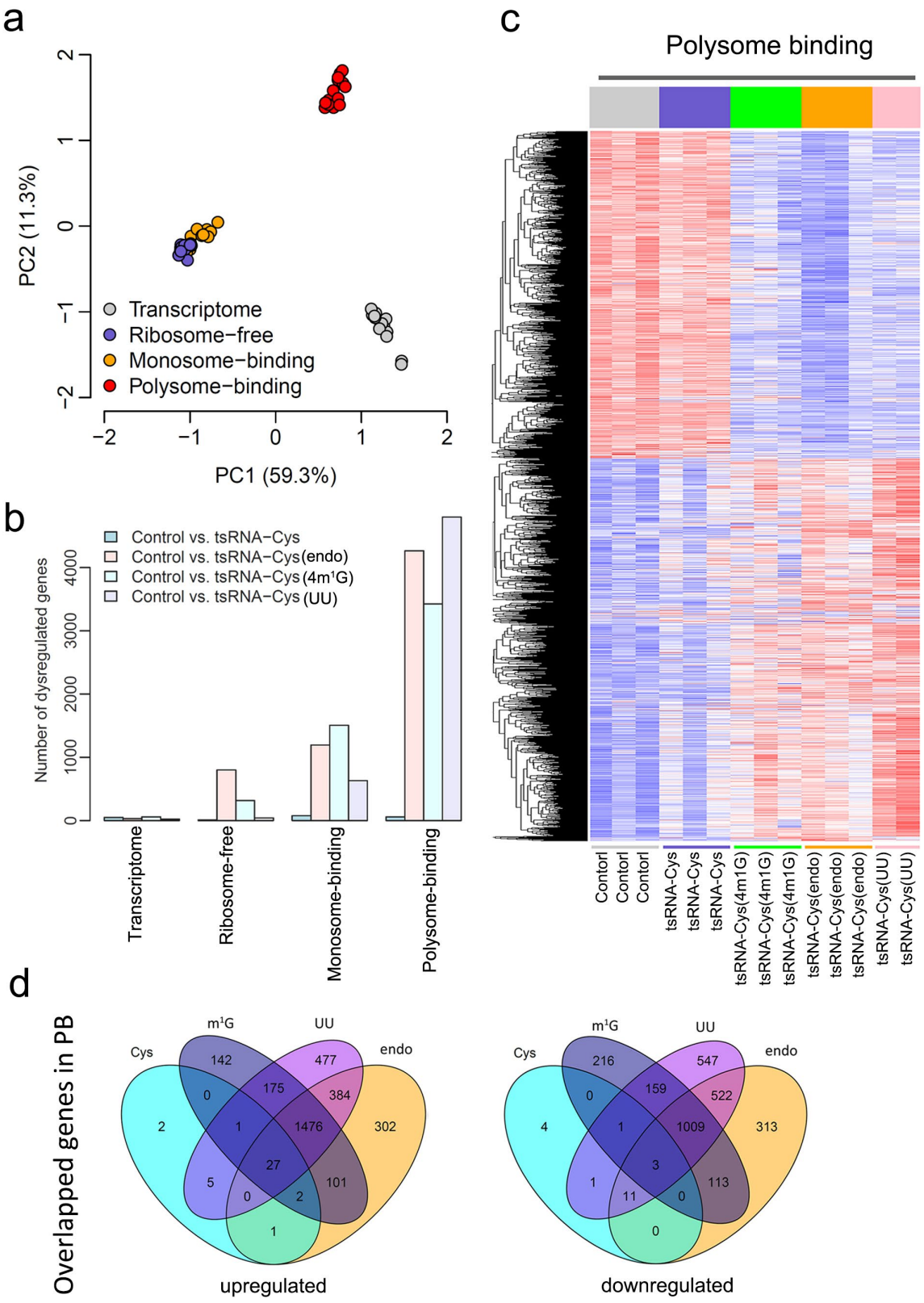

a
PC2 (11.3%)
PC1 (59.3%)
Transcriptome
Ribosome−free
Monosome−binding
Polysome−binding
b
Number of dysregulated genes
Control vs. tsRNA−Cys
Control vs. tsRNA−Cys(endo)
Control vs. tsRNA−Cys(4m¹G)
Control vs. tsRNA−Cys(UU)
Transcriptome
Ribosome-free
Monosome-binding
Polysome-binding
c
Polysome binding
Contorl
tsRNA-Cys
tsRNA-Cys(4m1G)
tsRNA-Cys(endo)
tsRNA-Cys(UU)
d
Overlapped genes in PB
Cys
m¹G
UU
endo
upregulated
downregulated

Figure 3.3: Comparative bioinformatic analysis of transcriptomic data from both engineered and endogenous 5′ tsRNA^{Cys} transfection. a: Principal component analysis on gene expression of induvial samples; PC1: the first principal component; PC2: the second principal component; b: Dysregulated genes induced by synthetics tsRNA-Cys, endo-tsRNA-Cys and tsRNA-Cys (4m1G), and tsRNA-Cys (UU) transfection in each fractions from plosyome profling and full transcriptome. c: Heatmap of dysregulated genes in the ploysome binding fraction of different tsRNA-Cys variants transfection; d: The number of overlopped gene in the polysome binding fraction between different tsRNA-Cys transfection groups.

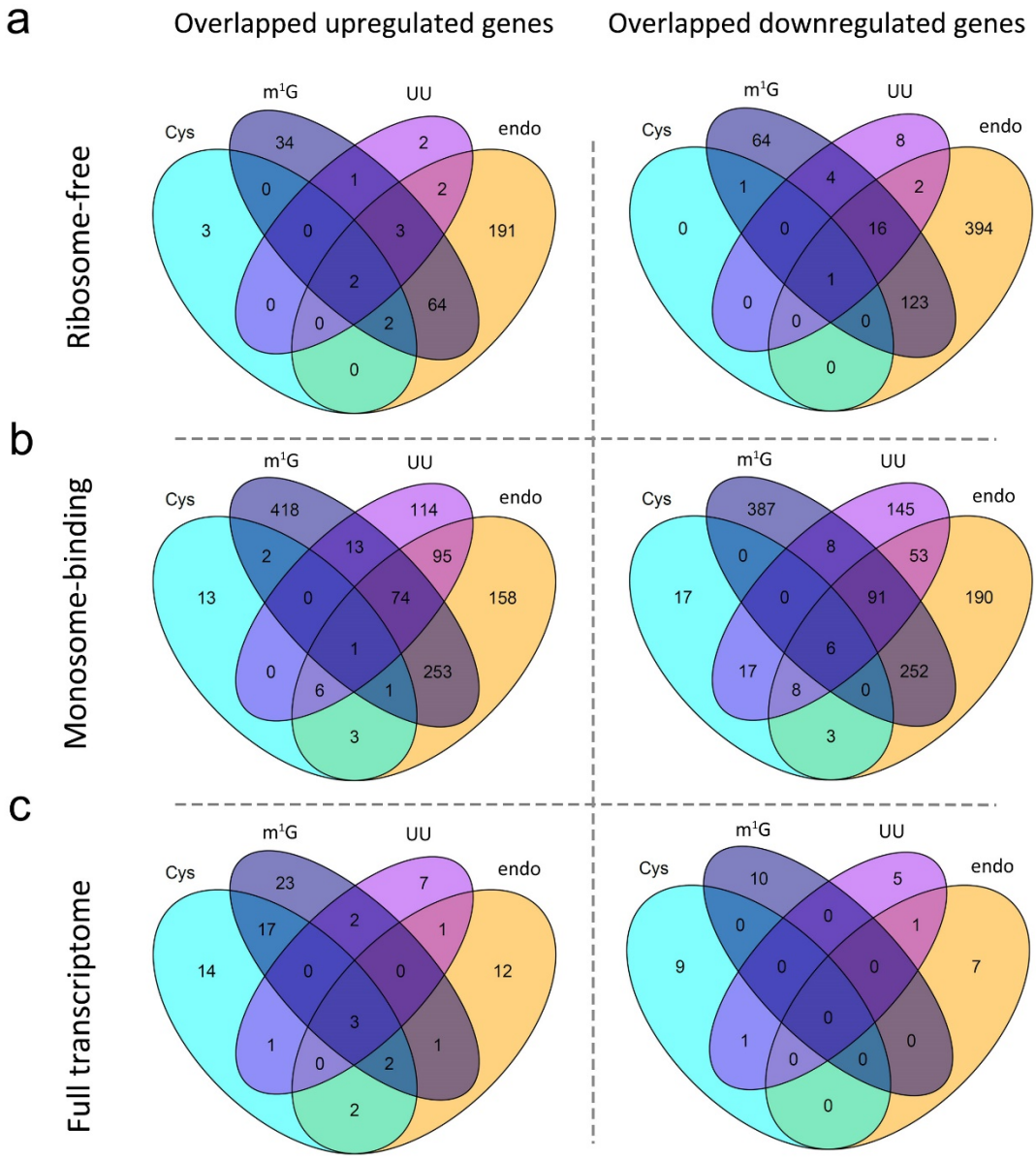


Figure 3.4: The overlapped dysregulated genes in different categories induced by different tsRNA^{Cys} variants transfection. The overlapped upregulated genes (left) and downregulated genes(down) in the ribosome free part (a), monosome binding fraction (b), and full cellular transcriptome (c). Cys: synthetic tsRNA-Cys, m^{1}G: tsRNA-Cys (4m1G), UU: tsRNA-Cys (UU); endo: endo-tsRNA-Cys.

To get further insight about the relation of tsRNA transfection and gene dysregulation in polysome binding fraction, we calculated the minimum free energy (MFE) of hybridization between the sequence of tsRNA-Cys and mRNAs in the polysome binding fractions *in silico*. mRNAs in the polysome binding fraction were grouped into three categories (upregulated, downregulated and not significant). From the cumulative distribution plots of MFE, we found that the MFE was lower when the single-strand tsRNA-Cys hybridized with the downregulated mRNAs, compared to hybridized with upregulated mRNAs (Figure 3.5). The MFE of hybridization between tsRNA-Cys sequence and unchanged mRNAs was lower than that of upregulated mRNAs, but higher than downregulated mRNAs, which meets our expectation. The lower MEF, the more stable of the RNA duplex (Rehmsmeier et al. 2004), so this result strongly suggested that single-strand tsRNA-Cys were preferred to interact with downregulated genes (with lower MFE) in the polysome binding fraction. In another words, single-stranded tsRNA-Cys preferred to target the downregulated genes and thus impact their loading into ribosomes.

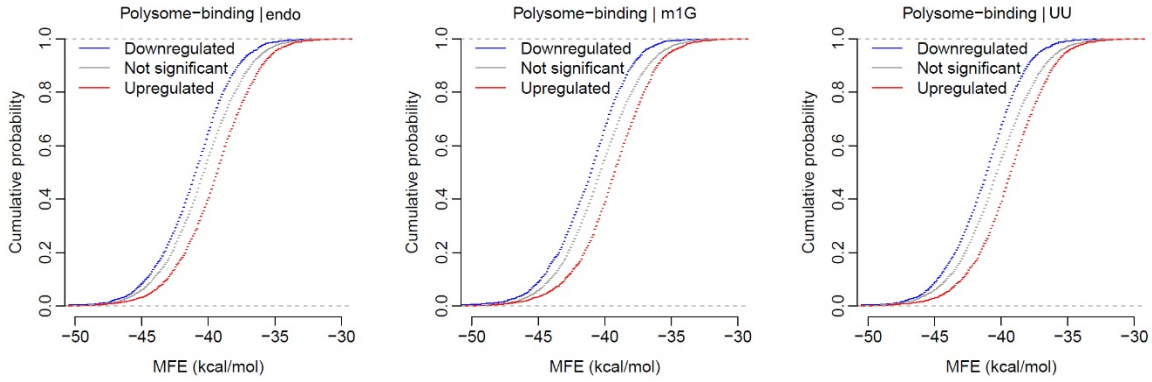


Figure 3.5: Cumulative probability plot of minimum free energy MEF. Cumulative distribution plot of minimum free energy (MFE) between tsRNA^{Cys} and mRNAs in polysome binding fractions from different transfection experiments; endo: endo-tsRNA-Cys, m1G: tsRNA-Cys (4m1G), UU: tsRNA-Cys (UU).

3.3 Discussion

We first identified the most effective modification (m^{1}G) to disrupt the RG4 formation of TOG-tsRNAs on guanine residues (Figure 3.1) and applied this RNA modification to engineer synthetic 5′ tsRNA^{Cys}. The engineered tsRNA-Cys mutants does not form RG4 structure both *in vitro* and *in vivo*, and induced strong gene dysregulation in the polysome binding fractions (Figure 3.2, 3.3). Moreover, much more dysregulated genes are overlapped between engineered tsRNA-Cys and endo-tsRNA-Cys than the synthetic tsRNA-Cys (Figure 3.3, 3.4). Our results of structural and functional analysis both manifested that the engineered tsRNA-Cys (4m1G) and tsRNA-Cys (UU) are more resemblant to the endo-tsRNA-Cys rather than the synthetic tsRNA-Cys.

Previous reports showed that the main functional mechanism of 5′ tsRNA^{Cys} or other 5′-TOG tsRNAs are forming RNA G-quadruplex to displace the translation

initiation factor from the m^7G capped mRNAs thus leading to the global translational inhibition (Ivanov et al. 2011, Lyons et al. 2017). In our study, we revealed a new regulation pattern induced by endogenous 5′ $tsRNA^{Cys}$ that single-strand 5′ $tsRNA^{Cys}$ interfere with the loading of mRNAs into ribosomes to regulate the cellular state by shaping a specific translatome, and this effect can be recapitulated by the engineered tsRNA-Cys mutants.

Besides that, the engineered tsRNA-Cys mutants are also different from endo-tsRNA-Cys in some extent. For example, the endo-tsRNA-Cys also can form a dimmer with unknown function which were not observed in the synthetic and engineered tsRNA-Cys. This may explain the difference we observed regarding the number of dysregulated genes in different fractions. Another factor may contribute to this difference is the RNA modification. Endogenous 5′ $tsRNA^{Cys}$ are highly modified and harbors multiple types of RNA modifications which are important for the small RNA's function (Chen et al. 2021, Zhang et al. 2016). The modified tsRNA vs. unmodified tsRNAs may have different affinity to RNA binding proteins thus led to different functional results. For instance, the 5′-TOG tsRNA with Ψ8 modification are preferring to bind to the polyadenylate-binding protein 1 (PABPC1), but TOG tsRNA without Ψ8 modification prefer to bind YBX1 and DEAH-box helicase 36 (DHX36) (Guzzi et al. 2018). One more specific case, the cytochrome c has a significantly higher affinity for the endogenous tsRNAs rather than the synthetic single-strand tsRNA mimics, which suggested the RNA modifications play a role in the binding selectivity of the cytochrome c (Saikia et al. 2014).

Even through, tsRNA can silence gene by targeting gene transcripts in the 3′ UTR in miRNA-like fashion which is AGO dependent (Luo et al. 2018, Xiao et al. 2021, Zong et al. 2021). But we only find very few dysregulated genes in the full transcriptome data which indicating that the endo-tsRNA-Cys and engineered tsRNA-Cys may not function through this gene silencing pathway in our experimental system. The single-strand endogenous 5′ $tsRNA^{Cys}$ and engineered 5′ $tsRNA^{Cys}$ mutants are more likely to bind with other RBP to interact with mRNAs wherever there is complementary sequence of tsRNAs, thus impeding the loading of mRNAs into ribosome rather than silencing them in a miRNA-like fashion. Recently, the 5′ terminal oligopyrimidine (5′-TOP) mRNAs were reported to be preferentially translated under the mTORC1-mediated regulation of mRNA translation (Thoreen et al. 2012), and the 5′TOP gene expression level were negatively correlated with the serum starvation condition in which 5′-TOG tsRNA were upregulated (Luo et al. 2018). These observations may indicate that TOG tsRNA could interact with 5′-TOP mRNA in the 5′ UTR to regulate their loading into ribosomes in addition to the miRNA-like fashion.

We also observed some genes were dysregulated in the monosome binding fraction. There are studies showing that some transcripts can be translated by monosome rather than polysome, these transcripts were enriched in the low abundance regulator proteins or contained short open reading frames (Biever et al. 2020, Heyer et al. 2016). The monosome fraction can be further analyzed by ribosome profiling sequencing to identify whether tsRNAs interact with the monosome binding mRNA to impede the elongation process rather than the mRNA loading.

3.4 Materials and Methods

G4 detection by Native PAGE gel

See the section of Materials and Method in Chapter 2.

tsRNA transfection and G4 detection by QUMA-1

See the section of Materials and Method in Chapter 2.

Polysome profiling and RNA extraction from different fractions

See the section of Materials and Method in Chapter 2.

mRNA sequencing and Transcriptome analysis

See the section of Materials and Method in Chapter 2

MEF calculation

The RNAhybrid program (Rehmsmeier et al. 2004) was used to compute the free energy of duplex structure in kcal/mol. In brief, the mRNAs in the polysome binding fraction from are uploaded to the RNA hybrid serve(https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid), and then the sequence of $tsRNA^{Cys}$ (GGGGGUAUAGCUCAGUGGUAGAGCAUUUGA) are also uploaded to the server. The default program parameters are applied, after the web server finish the MEF calculation, the results are downloaded to local to perform the cumulative distribution analysis by in-build statistic function of R.

3.5 Reference

Biever, A., C. Glock, G. Tushev, E. Ciirdaeva, T. Dalmay, J. D. Langer and E. M. Schuman (2020). "Monosomes actively translate synaptic mRNAs in neuronal processes." Science **367**(6477).

Boccaletto, P., M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm and J. M. Bujnicki (2018). "MODOMICS: a database of RNA modification pathways. 2017 update." Nucleic Acids Res **46**(D1): D303-D307.

Chen, Q., M. Yan, Z. Cao, X. Li, Y. Zhang, J. Shi, G. H. Feng, H. Peng, X. Zhang, Y. Zhang, J. Qian, E. Duan, Q. Zhai and Q. Zhou (2016). "Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder." Science **351**(6271): 397-400.

Chen, Q., X. Zhang, J. Shi, M. Yan and T. Zhou (2021). "Origins and evolving functionalities of tRNA-derived small RNAs." Trends Biochem Sci **46**(10): 790-804.

Guzzi, N., M. Ciesla, P. C. T. Ngoc, S. Lang, S. Arora, M. Dimitriou, K. Pimkova, M. N. E. Sommarin, R. Munita, M. Lubas, Y. Lim, K. Okuyama, S. Soneji, G. Karlsson, J. Hansson, G. Jonsson, A. H. Lund, M. Sigvardsson, E. Hellstrom-Lindberg, A. C. Hsieh and C. Bellodi (2018). "Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells." Cell **173**(5): 1204-1216 e1226.

Heyer, E. E. and M. J. Moore (2016). "Redefining the Translational Status of 80S Monosomes." Cell **164**(4): 757-769.

Ivanov, P., M. M. Emara, J. Villen, S. P. Gygi and P. Anderson (2011). "Angiogenin-induced tRNA fragments inhibit translation initiation." Mol Cell **43**(4): 613-623.

Luo, S., F. He, J. Luo, S. Dou, Y. Wang, A. Guo and J. Lu (2018). "Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response." Nucleic Acids Res **46**(10): 5250-5268.

Lyons, S. M., D. Gudanis, S. M. Coyne, Z. Gdaniec and P. Ivanov (2017). "Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs." Nat Commun **8**(1): 1127.

Rehmsmeier, M., P. Steffen, M. Hochsmann and R. Giegerich (2004). "Fast and effective prediction of microRNA/target duplexes." RNA **10**(10): 1507-1517.

Saikia, M., R. Jobava, M. Parisien, A. Putnam, D. Krokowski, X. H. Gao, B. J. Guan, Y. Yuan, E. Jankowsky, Z. Feng, G. F. Hu, M. Pusztai-Carey, M. Gorla, N. B. Sepuri, T. Pan and M. Hatzoglou (2014). "Angiogenin-cleaved tRNA halves interact with cytochrome c, protecting cells from apoptosis during osmotic stress." Mol Cell Biol **34**(13): 2450-2463.

Thoreen, C. C., L. Chantranupong, H. R. Keys, T. Wang, N. S. Gray and D. M. Sabatini (2012). "A unifying model for mTORC1-mediated regulation of mRNA translation." Nature **485**(7396): 109-113.

Xiao, Q., P. Gao, X. Huang, X. Chen, Q. Chen, X. Lv, Y. Fu, Y. Song and Z. Wang (2021). "tRFTars: predicting the targets of tRNA-derived fragments." J Transl Med **19**(1): 88.

Zhang, X., A. E. Cozen, Y. Liu, Q. Chen and T. M. Lowe (2016). "Small RNA Modifications: Integral to Function and Disease." Trends Mol Med **22**(12): 1025-1034.

Zhang, Y., Y. Zhang, J. Shi, H. Zhang, Z. Cao, X. Gao, W. Ren, Y. Ning, L. Ning, Y. Cao, Y. Chen, W. Ji, Z. J. Chen, Q. Chen and E. Duan (2014). "Identification and characterization of an ancient class of small RNAs enriched in serum associating with active infection." J Mol Cell Biol **6**(2): 172-174.

Zong, T., Y. Yang, X. Lin, S. Jiang, H. Zhao, M. Liu, Y. Meng, Y. Li, L. Zhao, G. Tang, K. Gong, Z. Wang and T. Yu (2021). "5'-tiRNA-Cys-GCA regulates VSMC proliferation and phenotypic transition by targeting STAT4 in aortic dissection." Mol Ther Nucleic Acids **26**: 295-306.

Chapter 4

RNA modification map of endogenous tRNAs revealed by MLC-seq

4.1 Abstract

RNA modifications are important to the structure and function of small RNAs. LC-MS/MS is the most widely used method to quantify the RNA modifications, but it loses the positional information of the RNA modifications. To further dissect the RNA modification on the endogenous small RNAs, we recently developed a LC-MS-based small RNA *de novo* sequencing method, called MLC-seq, which simultaneously get the sequence and the RNA modifications map of endogenous small RNAs. By this method we successfully sequenced the non-TOG containing tRNA ($tRNA^{Gln}$) and TOG containing tRNA ($tRNA^{Ala}$), and found new RNA modifications and point mutations on these two endogenous tRNAs. This method will be very promising to add the site-specific information about RNA modification of both tRNAs and tsRNAs.

4.2 Introduction

Beyond the primary RNA sequences, the various RNA modifications on small noncoding RNAs (sncRNAs) represent an additional layer of regulation by contributing to differential RNA stability, structure, and protein binding potentials (Guzzi et al. 2018, Zhang et al. 2016). The importance of RNA modifications has become more remarkable for sncRNAs with its impacts on the biogenesis and function of sncRNAs, such as tRNA, tsRNA and ribosomal RNA-derived small RNA (rsRNA). However, many modifications on sncRNAs remain undetectable due the current mainstream 'RNA-seq' methods need to convert the RNA to cDNA before sequencing which results in the loss of most of RNA modification information (Guzzi et al. 2018). Some methods are applied to specific high-throughput mapping of RNA modifications based on antibody or specific chemical conversion, but these methods are developed for long RNAs and only limited to a few well-known modifications, such as N6-methyladenosine (m^6A) (Dominissini et al. 2012), pseudouridine (Ψ) (Carlile et al. 2014, Li et al. 2015), inosine (I) (Roth et al. 2019) and m^5C (Pollex et al. 2010). Now the newly developed nanopore-based direct RNA sequencing can directly read the RNA sequence and their modification, but only one modification at a time (Anreiter et al. 2021) and it suffers from a high error rate for the highly modified RNA sequence like tRNAs (Thomas et al. 2021). In short, there are currently no efficient methods for high-throughput and comprehensive, quantitative mapping of multiple types of modifications in sncRNAs, or RNAs in general.

Liquid chromatography and tandem mass spectrometry (LC-MS/MS) is recognized as the 'gold standard' for RNA modification analysis. But this method

analyzed the RNA modification at nucleosides level that it requires to digest the RNA strand into single nucleoside, which lose the positional information of RNA modification on RNA sequence. Recently, the MS-based *de novo* sequencing methods have been developed to simultaneously get the sequence of a modified RNA strand and quantify different RNA modifications within their full-length RNA sequence context at single-nucleotide precision (Bjorkbom et al. 2015, Zhang et al. 2019, Zhang et al. 2020). MS-based *de novo* sequencing utilizes uniform RNA mass ladders to read the RNA sequence and the modification information. In the uniform RNA mass ladders, the larger mass ladder has one more nucleoside than the previous smaller one sequentially, so the RNA sequence can be directly read according to the mass difference along with the ladders. More important, the RNA modifications can also be deduced by the mass shift between contiguous RNA ladders, for instance, a methyl group on the nucleoside will cause an additional 14 Dalton (Da) mass shift compared to the unmodified nucleoside. This is conceptually similar to the sanger sequencing strategy regarding the formation of DNA ladder.

However, the uniform RNA ladder is difficult to get by the enzymatic or chemical digestion in practice (Bjorkbom et al. 2015, Yoluc et al. 2021, Zhang et al. 2019), and the RNAs extracted from the biological sample usually have different variants which are in low abundance and cannot meet the requirements to generate uniform RNA ladders. To overcome these limitations, we developed a *de novo* mass ladder complementation sequencing method (MLC-Seq) and successfully sequenced the endogenous $tRNA^{Gln}$ and $tRNA^{Ala}$ extracted from mouse liver.

4.3 Results

General workflow of MLC-seq

The general workflow of MLC-Seq to sequence cellular tRNA samples include three parts: sample processing, LC-MS measurement, and data processing. A given tRNA sample is firstly divided into two groups, one is intact group without formic acid (FA) treatment; the other one is acid-hydrolyzed group with FA treatment. The intact tRNA served as a control is injected directly into LC-MS to provide mass and relative abundance information of all the tRNAs which can be used to identify the isoforms in the data processing (Figure 4.1a, (I)). Acid-hydrolyzed tRNA is injected into LC-MS to provide information including mass and related retention time at tRNA ladder fragment level (Figure 4.1a, (II)). By plotting a 2D mass-retention time (tR) curve of the tRNA ladders, we can observe RNA modification events by the mass shift on the curve. For example, a methylation event can be distinguished by the 14 Da mass shift on the 2D mass-retention time curve of tRNA ladders which is shown as a branch of the curve (Figure 4.1a, (II)). After more data processing by multiple algorithms, which include the isoform identification, ladder gap filling, 5′/3′ ladder separation and complementation (Figure 4.1b), we can finally get computationally uniform RNA ladders corresponding to the nucleotides from the first to the last on an RNA molecule (Figure 4.1b). We can directly read the tRNA sequence and its RNA modification through this uniform RNA ladders. Also, stoichiometric quantification of RNA modification can be calculated with relative intensity of the corresponding ladder fragments in their branches. (See more details in the Material and Methods)

Notably, additional methods are needed to distinguish modified nucleotides with the same mass shift. For example, the sensitivity to AlkB enzyme treatment can be used for distinguishing between m^1A and m^6A, or between m^1G and m^2G, where m^1A, m^1G and m^3C can be demethylated by AlkB enzyme. Nucleotides with 2′-O-methylation (Am, Um, Cm, and Gm) can prevent the acid hydrolyzation and thus generate a mass gap in the RNA mass ladder (Bjorkbom et al. 2015, Zhang et al. 2019). For the pseudouridine, it can be converted to a CMC adduct by N-cyclohexyl-N'-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulfonate (CMC) which generates a mass shift different from uridine, so the pseudouridine can be called out from RNA sequence (Zhang et al. 2020).

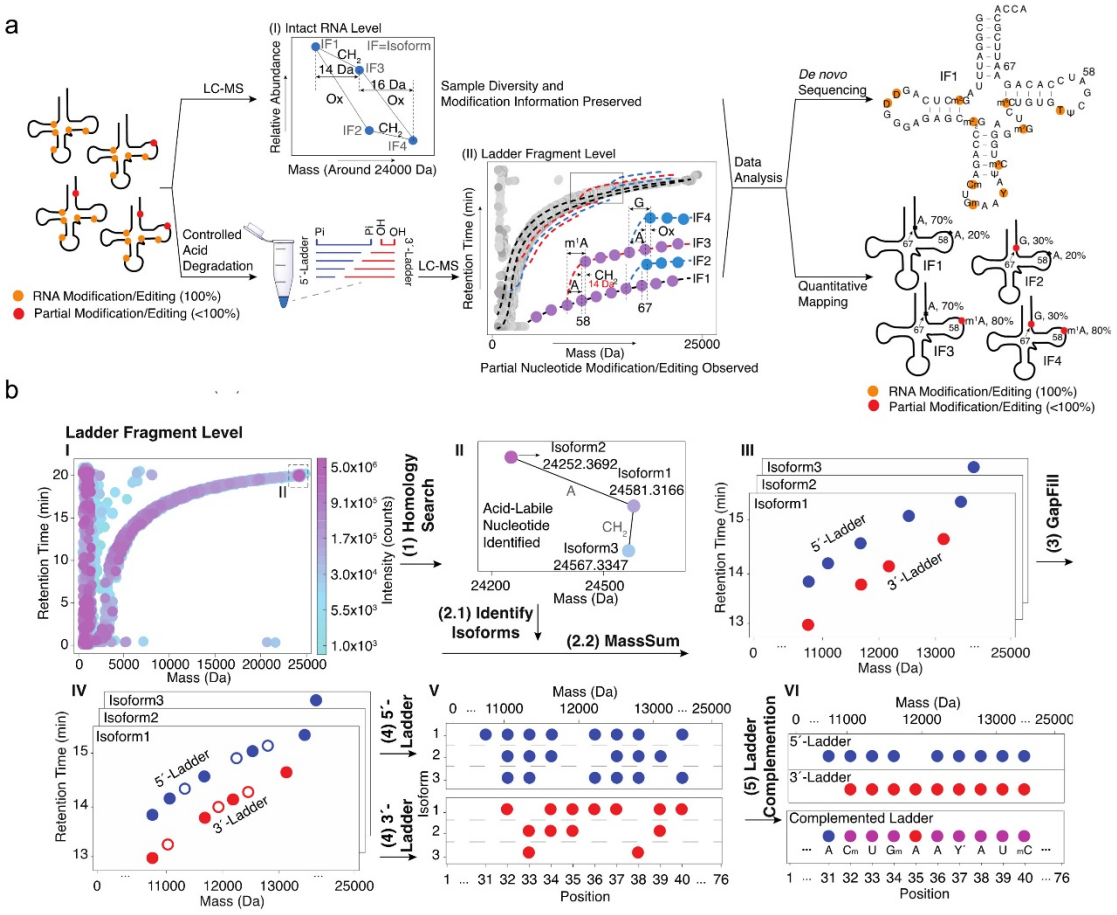


Figure 4.1: **General workflow of MLC-seq to sequence tRNA mixtures**. a: Schematic of general workflow of MLC-seq. It includes four major steps: RNA sample acid hydrolysis, LC-MS measurement; data processing and generation of sequences consisting of both canonical and modified nucleotides. b: Detailed illustration for "Data Processing" procedure; I-II, Homology Search allows related tRNAs to be identified; III, for each isoform identified, MassSum algorithm was applied to extract all its ladder fragments; IV, GapFill to find non-paired ladder fragments that are missed by MassSum. V: Ladder separation; VI, Ladder complementation to perfect MS ladders, allowing direct sequencing of full-length tRNA.

MLC-Seq profiled and pinpoint multiple RNA modifications on $tRNA^{Gln}$ and $tRNA^{Ala}$

To get more insights about the modifications of endogenous noncoding small RNAs, we purified non-5′TOG-containing $tRNA^{Gln}$ and 5′TOG-containing $tRNA^{Ala}$ from the mouse liver via the affinity pull-down assay for the MLC-seq. We also performed the MLC-seq of $tRNA^{Gln}$ and $tRNA^{Ala}$ with AlkB treatment (Figure 4.2), which aims to distinguish the methylated nucleotides with same mass shift as mentioned above.

By MLC-seq, we found that both tRNAs are highly modified with diverse RNA modifications, the list of identity, position, and stoichiometry of each modification in endogenous $tRNA^{Gln}$ and $tRNA^{Ala}$ is shown in Figure 4.2a and Figure 4.2e. For the $tRNA^{Gln}$, we found that the G at position 9 of $tRNA^{Gln}$ is 100% methylated, but after AlkB treatment, the methylated G at position 9 is only 55% left which suggests that this methylated G is m^1G because it is sensitive to AlkB treatment (Figure 4.2a). In wild type $tRNA^{Gln}$ which are not treated by AlkB, m^1A at position 57 of $tRNA^{Gln}$ was 100%, as shown by the 3′ RNA ladder of the wild-type $tRNA^{Gln}$ in the 2D mass-tR plot (Figure 4.2b, left). However, only 1% m^1A at position 57 was observed in the 2D mass-tR after AlkB-treatment which was shown by a branch with 14 Da mass shift with 1% intensity (Figure 4.2b, right).

More importantly, MLC-seq also leads to the discovery of new RNA modifications which have not been reported before, the 3,4-dihydrocytidine (C') at position 16 of $tRNA^{Gln}$. MLC-Seq showed that when reading the 5′ ladder of the $tRNA^{Gln}$ in the 2D mass-tR plot, a new ladder started to branch out at position 16, and the mass

differences of ladder fragments between position 16 and 15 are 307.0276 Da and 308.0340 Da, indicating a C' co-existing with D (Dihydrouridine) at position 16 (D/C': 79% vs. 21%) (Figure 4.2c). We propose a structure for this newly discovered nucleotide C' (Figure 4.2d), which contains two more hydrogens in positions 3 and 4 of Cytidine (C), based on reported structural stability analysis of the isomeric C' (Snider et al. 2002).

In the $tRNA^{Ala}$, we found the methylated G at position 10 is less sensitive to AlkB treatment that only 11% methylated G were demethylated, so we just call mG at this position for the $tsRNA^{Ala}$ which we thought is more accurate. Previous reports showed that this site is m^2G in some tRNAs, but it is possible that both m^1G and m^2G are co-existed at this site. And the A at position 57 of $tRNA^{Ala}$ is unmethylated (Figure 4.2e) which different from $tRNA^{Gln}$. In addition to the sites with full modification, we also observed partial modified sites in both tRNAs. For instance, the position 53 of $tRNA^{Ala}$ was co-occupied by the canonical U and methylated U at a ratio of 64% to 36% (Figure 4.2e). The position 49 of $tRNA^{Gln}$ are partially modified by m^5C and m^5U (Figure 4.2a). Points mutations on the tRNA sequence is also existed, for example, there are two G to A mutations happened on the sequence of $tRNA^{Ala}$ at the position 5 and position 27 with 30% percentage (Figure 4.2e), and an adenosine(A) instead of guanine (G) presented at position 68 of $tRNA^{Gln}$ with 3% percentage (Figure 4.2a). These results suggest that there are sequence variants among the same endogenous tRNAs.

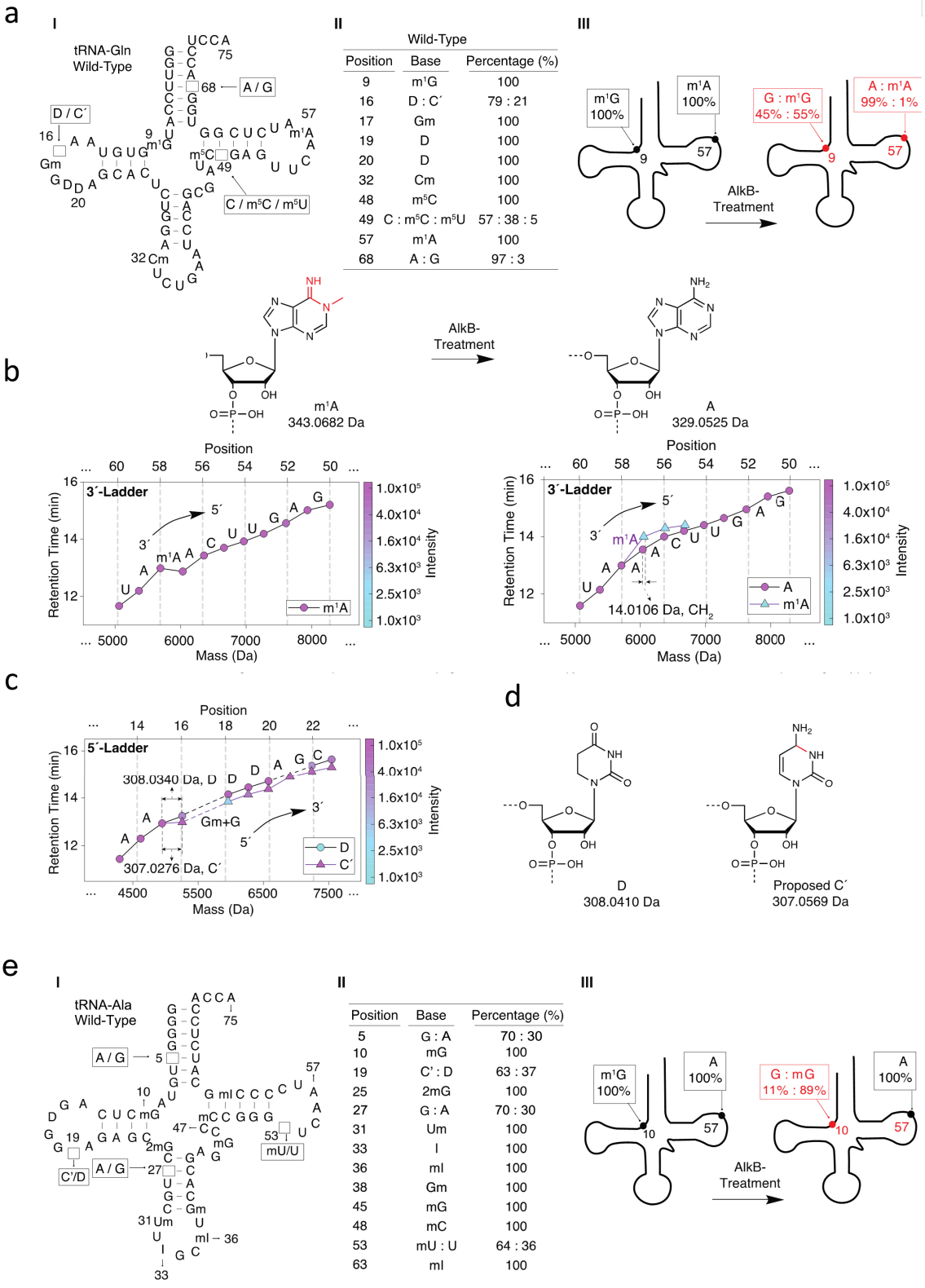

a
I
tRNA-Gln
Wild-Type
A / G
D / C´
C / m5C / m5U
II
Wild-Type
Position | Base | Percentage (%)
9 | m1G | 100
16 | D : C´ | 79 : 21
17 | Gm | 100
19 | D | 100
20 | D | 100
32 | Cm | 100
48 | m5C | 100
49 | C : m5C : m5U | 57 : 38 : 5
57 | m1A | 100
68 | A : G | 97 : 3
III
m1G 100%
m1A 100%
G : m1G 45% : 55%
A : m1A 99% : 1%
AlkB-Treatment
b
AlkB-Treatment
m1A
343.0682 Da
A
329.0525 Da
Position
3´-Ladder
Retention Time (min)
Mass (Da)
Intensity
m1A
A
14.0106 Da, CH2
c
5´-Ladder
308.0340 Da, D
307.0276 Da, C´
Gm+G
D
C´
d
D
308.0410 Da
Proposed C´
307.0569 Da
e
I
tRNA-Ala
Wild-Type
A / G
C´/D
mU/U
II
Position | Base | Percentage (%)
5 | G : A | 70 : 30
10 | mG | 100
19 | C´ : D | 63 : 37
25 | 2mG | 100
27 | G : A | 70 : 30
31 | Um | 100
33 | I | 100
36 | mI | 100
38 | Gm | 100
45 | mG | 100
48 | mC | 100
53 | mU : U | 64 : 36
63 | mI | 100
III
m1G 100%
A 100%
G : mG 11% : 89%
AlkB-Treatment

Figure 4.2: *De novo* sequencing result of endogenous tRNA^{Gln} and tRNA^{Ala} with RNA modification position. a: MLC-Seq results of tRNA^{Gln}, showing all the RNA modifications within their full-length tRNA sequence context as well as the stoichiometry of each modification site-specifically; b: MLC-Seq results of tRNA^{Ala}, showing all the RNA modifications within their full-length tRNA sequence context as well as the stoichiometry of each modification site-specifically; c: MLC-Seq results pinpoint stoichiometric changes of m^1A site-specifically at position 57 of tRNA^{Gln} after AlkB treatment; d: MLC-Seq leading to the discovery of a novel nucleotide of C′ (3,4-dihydrocytidine) at position 16 of tRNA^{Gln}; e: propose structure of C′ (3,4-dihydrocytidine)

4.4 Discussion

By the MLC-seq, we successfully revealed the RNA modification map of the endogenous $tRNA^{Gln}$ and $tRNA^{Ala}$ including the specific site and type of the RNA modifications as well as their stoichiometry (Figure 4.2). We found the modification profile of these two tRNAs from the liver tissue are totally different. For instance, the A residue at position 57 which located in the conserved T-loop region is methylated in the $tRNA^{Gln}$ but not in $tRNA^{Ala}$. The m^1A is reported to related the biogenesis of $tsRNA^{Gln}$ that the decrease of m^1A will result in the accumulation of 5′ $tsRNA^{Gln}$ (Chen et al. 2019) which is similar to the role of m^1G modification at position 9 for the 5′ $tsRNA^{Gln}$ generation (Cosentino et al. 2018). Moreover, this m^1A modification may also serves as a modular to regulate the biogenesis of 3′ tsRNAs from $tRNA^{Gln}$ which has been reported to exist abundantly in the human cardiac tissues (Torres et al. 2019). How the 3′ $tsRNA^{Gln}$ contribute to the normal cardiac function is an interesting question to be investigated in the future. For the $tRNA^{Ala}$, we found a G to A mutation happened in the terminal oligoguanine region (TOG) which may pass to the 5′ $tsRNA^{Ala}$ and impact the RG4 formation of 5′ $tsRNA^{Ala}$, thus influencing the effect of 5′ $tsRNA^{Ala}$ relating to translational inhibition (Ivanov et al. 2011, Lyons et al. 2017). Most recently, 5′ $tsRNA^{Ala}$, which only have 4 G residues in the TOG region, has been reported to regulate anti-fungal defense in plants via mRNA cleavage by binding to AGO1 (Gu et al. 2021). Whether the sequence variants of 5′ $tsRNA^{Ala}$ with the G to A mutation prefer to cooperate with AGO proteins rather than form RG4 to function in animal cells is an interesting question waiting to be explored. Interestingly, we also found a new RNA

modification, 3,4-dihydrocytidine (C′), in the tRNAs. C′ may be a general RNA modification on tRNAs or a tissue specific RNA modification in the liver tissue, which needs more investigations to confirm. MLC-seq is powerful tool to answer this question and to explore tissue specific RNA modification landscape by sequencing same tRNAs of different tissues in the further.

More important, MLC-seq can also be applied to tsRNAs and other modified small RNAs with easy optimization. If we can compare the RNA modification profile of tsRNAs and their corresponding tRNA, this will help to figure out whether tsRNAs can get new modifications by specific enzymes except those inherited from its precursor, which is a critical question regarding the biogenesis of tsRNAs. Moreover, profiling the RNA modification map of tsRNAs helps to get better understanding about the functional mechanism of tsRNAs which may help to solve the observed discrepancy relating to the effects of tsRNAs in same biological process such as cancer development and viral infection.

4.5 Materials and Methods

$tRNA^{Gln}$ and $tRNA^{Ala}$ pulldown

The tRNA was purified by affinity pulldown assay combined with gel recovery, with modified protocols from previous report (Drino et al. 2020). The total RNA of mouse liver was harvested by TRIzol™ reagent (Invitrogen™) as the manufacturer instructed. The concentration of total RNAs solution was adjusted to 2mg/ml by RNase-free water. Small RNAs fraction (<200 nt) were separated by the buffer containing 50%(w/v) PEG 8000 and 0.5M NaCl solution by centrifugation at 12000rpm and 4 °C for 20 min. The supernatant was collected followed by adding 1/10 volume NaAc solution (Invitrogen™). 1 ml supernatant was added with 3 ml Ethanol, and 5 µl Linear Acrylamide (Invitrogen™) to precipitate small RNAs (<200 nt) with -20 °C overnight incubation followed by centrifugation at 12000rpm at 4 °C for 20 min. Small RNA (<200 nt) solution was adjusted to 1mg/ml, 1 ml small RNA solution with 6 µl biotinylated probe(100 µM), 26 µl 20× SSC solution (Invitrogen™) and 15 µl RNase inhibitor (NEB) were incubated at 50°C overnight. 200 µl Streptavidin Sepharose (Cytiva) was added to the hybridization solution to enrich the biotin labeled probe that captured with the targeted tRNA. After incubation at room temperature for 30mins, the Streptavidin Sepharose was transferred to the 1.5 ml Ultrafree-MC tube (Millipore) and washed by 0.5× saline-sodium citrate (SSC) buffer, the washing step was repeated five times. 500 µl nuclease-free water was added to the MC tube and incubated at 70°C for 15 min followed by centrifugation at 2500g at room temperature for 1 min to elute the RNAs that are complementary to the biotinylated probe. The eluent was collected followed by adding

1/10 volume NaAc solution (Invitrogen™). 1 ml eluent was added with 3 ml Ethanol, and 5 µl Linear Acrylamide (Invitrogen™) to precipitate RNAs with -20 °C overnight incubation followed by centrifugation at 12000rpm at 4 °C for 20 min. Nuclease-free water was added to dissolve the RNAs pallets. RNAs were loaded into the 7M Urea-PAGE gel to run electrophoresis, the main tRNA band was recovered from the PAGE gel as previously described (Shi et al. 2021) to obtain enriched tRNAs for MLC-Seq. The DNA probes for pull down experiments were synthesized by IDT and the sequences were listed blew:

tRNA-Ala pulldown probe: 5′-Biotin-CGCTCTACCACTGAGCTACACCCCC

tRNA-Gln pulldown probes: 5′-Biotin-TGGAGGTTCCACCGAGATTTGA

Treatment of tRNA with AlkB

200ng tRNA was incubated in a 50 µl reaction mixture containing 50 mM Na-HEPES, pH 8.0 (Alfa Aesar), 75 µM ferrous ammonium sulfate (pH 5.0), 1 mM α-ketoglutaric acid (Sigma Aldrich), 2 mM sodium ascorbate, 50 µg/ml BSA (Sigma-Aldrich), 2.5 µl RNase inhibitor (NEB), and 200 ng AlkB at 37 °C for 30 min (the recommended mass ratio of AlkB to RNA is 1:1 or greater). 500 µl TRIzol™ reagent was added to the mixture and RNA was extracted as per manufacturer instructions.

Controlled Acid Degradation of tRNA Samples

Formic acid was applied to degrade tRNA samples, including tRNA-Gln and tRNA-Glu from mouse live. To produce mass ladders according to reported experimental

protocols (Zhang et al. 2020). In brief, we divided each RNA sample solution into three equal aliquots of 5 µl (each in a RNase-free, thin walled 0.2 ml PCR tube) for formic acid degradation, each using 50% (v/v) formic acid at 40 °C in a PCR machine, with one reaction running for 2 min, one for 5 min, and one for 15 min. The reaction mixtures were each immediately frozen on dry ice at the specified times, followed by centrifugal vacuum concentrator (Labconco Co., Kansas City, MO, USA) to dryness, which was typically completed within 30 min. The dried samples of three different time points for each specific tRNA sample were then combined and suspended in 20 µl nuclease-free, deionized water for LC-MS measurement.

LC-MS Measurement of Intact tRNA and Resulting Acid-Degraded tRNA Samples

Each combined acid-hydrolyzed tRNA sample was individually analyzed on an Orbitrap Exploris 240 mass spectrometer (ThermoFisher Scientific, Bremen, Germany) coupled to a Vanquish Horizon UHPLC using a DNAPac reversed-phase (RP) column (2.1 mm x 50 mm, ThermoFisher Scientific, Sunnyvale, California, USA) with 200 mM HFIP and 10 mM DIPEA as eluent A, and methanol, 7.5 mM HFIP, and 3.75 mM DIPEA as eluent B. A gradient of 2% to 38% B over 15 min was used for analysis of the samples. The flow rate was 0.4 ml/min, and all columnar separations were performed with the column temperature maintained at 40 °C. Injection volumes were 5–25 µl and sample amounts were 1-10 pmol (or ~20 ng) of tRNA-Ala and tRNA-Gln. tRNA samples were analyzed in negative ion full MS mode from 410 m/z to 3200 m/z with a scan rate of 2 spectra/s at 120k resolution at m/z 200. The data were processed using the Thermo

BioPharma Finder 4.0 (ThermoFisher Scientific), and a compound detection workflow with a deconvolution algorithm was used to extract relevant spectral and chromatographic information from the LC-MS experiments as described previously (Zhang et al. 2020).

Data processing steps for MLC-seq

Homology Search: Once LC-MS data is displayed as a two-dimensional (2D) mass-retention time (tR) plot, we start a homology search of intact tRNAs in the monoisotopic mass range of >~24k Dalton using an in-house developed algorithm in Python (see Github) to first identify related tRNA isoforms that may share the same ancestral precursor tRNA, but are different in absolute sequence, *e.g.*, in posttranscriptional profiles of nucleotide modifications, editing, and truncations. Mass differences between two intact tRNA isoforms are calculated and matched to the known mass of each nucleotide or nucleotide modification in the database (Bjorkbom et al. 2015). For example, known mass differences between intact tRNAs of 14.0157 Da and 329.0525 Da (with parts per million (ppm) difference <10 ppm) (Brenton et al. 2010) can be assigned to a methylation (Me/-CH_2-) event and an additional A nucleotide, respectively. Therefore, these intact tRNAs are assigned to the same tRNA group and considered as homologous isoforms of a specific tRNA for sequencing together. The homology search is a non-target pre-selection to group possible related tRNA isoforms together for sequencing. However, only one monoisotopic mass difference of two intact masses is used to identify the tRNA isoforms differing, *e.g.*, by RNA

editing/modifications and/or 3′-CCA truncations. Thus, there may be errors when grouping a tRNA isoform that does not belong to this specific tRNA group or *vice versa*, missing a tRNA isoform when cataloging a group. These errors can be fixed later when sequencing each group of tRNA isoforms, and sequencing results can further verify the inter-connection between isoforms.

Detection of all acid-labile nucleotide modifications: Acid-labile nucleotides are identified using another algorithm in Python (see Github). The algorithm analyzes the connections between the compounds (with a monoisotopic mass >24 K Dalton for tRNAs) measured by LC-MS before acid degradation and the LC-MS-detected compounds after acid degradation. For each such a compound pair, if the monoisotopic mass difference can be matched to a known mass difference (calculated from the possible structural change to a specific nucleotide modification during acid hydrolysis), or be matched to the mass difference sum of a subset of different acid-labile nucleotide modifications, the compound pair will be selected and further considered as potentially containing acid-labile nucleotide modifications. In general, if the intact mass of an RNA species does not change after acid degradation, this intact mass will be used for MassSum data separation (See below). Otherwise, an acid-labile nucleotide may be identified by matching to the observed mass difference with the theoretical mass difference caused by acid-mediated structural changes of the nucleotide.

5′- and 3′-ladder separation: The t_R differences can be used to separate these two ladders further computationally (5′- and 3′-ladders), breaking two adjacent sigmoidal curves into two isolated curves: one for the 3′- and the other for the 5′-ladder. Due to the

large number of RNA/fragment compounds, the dividing line between two subsets of 5′- and 3′-ladder fragments is not visually decisive in the 2D mass-tR plot. Thus, we developed a computational tool (see Github) to separate the 5′- and 3′-fragments. We roughly divide all compounds in each LC-MS data pool into two subgroup areas by circling compounds in the top collective curve of the 2D mass-t_R plot and marking the compounds as 5′- ladder fragment compounds, while the compounds in the bottom curve are marked as 3′- ladder fragment compounds. The purpose of selecting the top area is to include as many 5′- fragment compounds as possible, while minimizing the 3′- fragments in this area. Similarly, the purpose of selecting the bottom area is to include as many 3′- fragment compounds as possible, while minimizing the 5′- fragments in this area. Overlap between two selected ladder subgroups is inevitable, due to limited t_R differences between these two subgroups. The aim of the manual selection step is not to separate the 5′- and 3′-fragments from each other completely, but to serve as two input ladder fragments for MassSum algorithm (See below) to output 5′- and 3′- ladder fragments separately for each tRNA isoform/species.

MassSum data separation: MassSum is an algorithm in Python (see Github) developed based upon the acid degradation principle. Taking advantage of the fact that each fragmented pair from two ladder groups (5′- and 3′- ladders) sums to a constant mass value that is unique to each specific tRNA isoform/species, MassSum can isolate ladder compounds corresponding to a specific tRNA isoform. MassSum simplifies the dataset by grouping MS ladder components into subsets for each tRNA isoform/species based on its unique intact mass. Since the well-controlled acid degradation reaction

cleaves RNA oligonucleotides at one specific site of the phosphodiester bond (on average, one cut per RNA (Bjorkbom et al. 2015) the masses of two RNA fragments (Mass $_{3'\text{- portion}}$ and Mass $_{5'\text{- portion}}$) from the same strand sum to a constant value ($Mass_{sum}$).

$$Mass_{3'-portion} + Mass_{5'-portion} = Mass_{intact} + Mass_{H_2O} = Mass_{sum}$$

(1)

Taking advantage of this unique mass sum between the paired 3′- and 5′- ladder fragments (*Equation 1*), the algorithm chooses two random compounds from the acid-degraded LC-MS dataset and adds their mass values together, one pair at a time. If the sum of the two selected compounds is equal to a specific Masssum, these two compounds will be selected into the pool for each RNA accordingly. The process repeats until all compound pairs have been inspected. In the end, MassSum will cluster the dataset into different groups; each group is a subset that contains 3′- and 5′- ladders of one specific RNA sequence.

GapFill: GapFill is another Python-based algorithm (see Github) developed as a complement to MassSum. Since MassSum handles compounds in pairs, in the case that one ladder fragment is missing, e.g., in the 5′- ladder, the corresponding single-cut ladder fragment, even if it exists in the 3′- ladder, will not be separated/called by the MassSum algorithm. In order to extract all ladder fragments from the complex MS data, a GapFill algorithm was designed to “rescue” any ladder fragments missed by MassSum separation. GapFill identifies a gap that has ladder fragments missing between ladder fragment compounds, e.g., two adjacent compounds with Mass values Mass5′- i and Mass5′- j in the 5′- ladder which were found by MassSum algorithm. Within the gap, there exist many

ladder compounds in the degraded LC-MS dataset, but presumably none were selected by the MassSum algorithm during data separation. GapFill iterates over each potential compound in the gap in the original LC-MS dataset, and examines the mass differences of this compound and the two ending compounds, Mass5′-i and Mass5′-j. If the mass difference is equal to the sum of one or more nucleotides or modifications in the RNA modification database (Bjorkbom et al. 2015), we define it as a connection. If the compound in the gap has connections with both ending compounds, this compound would be selected into a candidate pool for the later sequencing process. After iteration, GapFill calculates connections of the compounds in the candidate pool and assigns weights to them based on the frequency of each connection. The compounds that contain the highest weights would be the ones chosen to fill in the gap.

Ladder complementation and generation of RNA sequences: After MassSum and GapFill, each tRNA isoform has its own set of separate 5′- and 3′- ladders (not combined). Each ladder (5′- or 3′-) consists of a ladder sequence, and we can determine if these ladders are perfect (without missing any ladder fragments), which would allow reading of the full RNA sequence (from the first to the last nucleotide in the sequence). If not, we can complement ladders from other related isoforms in order to obtain a more complete ladder for sequencing (ideally no missing ladder fragment or as complete as possible). A Python-based computational algorithm (see Github) was designed to align ladders from related isoforms based on the position of the ladder fragment in the 5′ →3′ direction. For example, we lay out each tRNA-Phe isoform's full 5′-ladder, *e.g.*, the 5′-ladder in Fig. 1b, on top of each other vertically; horizontally, we align the 5′- ladder of

each isoform according to the position of each corresponding ladder fragment, ranging from nucleotide position 1 to nucleotide position 76 for tRNA-Phe. Ladder complementation can be performed separately on 5′- or 3′-ladders separately (but not mixed ladders), resulting in one final 5′-ladder or one final 3′-ladder. Additionally, all 3′-ladder fragments can be converted to their corresponding 5′ -ladder fragments for each tRNA isoform based on the MassSum principle. As such, each tRNA isoform could have two 5′-ladder fragments in each position on the 5′-ladder: one original 5′-ladder fragment, and a second ladder fragment that was converted from its corresponding 3′- ladder fragment, for reaffirmation and/or complementation.

Stoichiometric quantification of partial nucleotide modifications/editing

Stoichiometries/percentages of partial nucleotide modifications/editing were quantified according to reported protocols (Zhang et al. 2019, Zhang et al. 2020, Zhang et al. 2020). In brief, to accurately determine the ratio of partial modification/editing, from datasets of multiple experimental trials ($n \geq 3$) of a given tRNA sample, 3 pairs of ladder fragments (one in the original ladder, and the other in the branched ladder) were taken among the partial modification position (first nucleotide after the branch point where the partial modification was observed) or otherwise its immediate next ladder fragment in the same ladder in case the ladder fragment at the partial modification position does not exist, *e.g.*, due to methylation on the 2′-hydroxyl group of Cm that blocks acid degradation. Mean ratio and standard deviation were calculated.

4.6 Reference

Anreiter, I., Q. Mir, J. T. Simpson, S. C. Janga and M. Soller (2021). "New Twists in Detecting mRNA Modification Dynamics." Trends Biotechnol **39**(1): 72-89.

Bjorkbom, A., V. S. Lelyveld, S. Zhang, W. Zhang, C. P. Tam, J. C. Blain and J. W. Szostak (2015). "Bidirectional Direct Sequencing of Noncanonical RNA by Two-Dimensional Analysis of Mass Chromatograms." J Am Chem Soc **137**(45): 14430-14438.

Bjorkbom, A., V. S. Lelyveld, S. Zhang, W. Zhang, C. P. Tam, J. C. Blain and J. W. Szostak (2015). "Bidirectional direct sequencing of noncanonical RNA by two-dimensional analysis of mass chromatograms." Journal of the American Chemical Society **137**(45): 14430-14438.

Brenton, A. G. and A. R. Godfrey (2010). "Accurate mass measurement: terminology and treatment of data." J Am Soc Mass Spectrom **21**(11): 1821-1835.

Carlile, T. M., M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli and W. V. Gilbert (2014). "Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells." Nature **515**(7525): 143-146.

Chen, Z., M. Qi, B. Shen, G. Luo, Y. Wu, J. Li, Z. Lu, Z. Zheng, Q. Dai and H. Wang (2019). "Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs." Nucleic Acids Res **47**(5): 2533-2545.

Cosentino, C., S. Toivonen, E. Diaz Villamil, M. Atta, J. L. Ravanat, S. Demine, A. A. Schiavo, N. Pachera, J. P. Deglasse, J. C. Jonas, D. Balboa, T. Otonkoski, E. R. Pearson, P. Marchetti, D. L. Eizirik, M. Cnop and M. Igoillo-Esteve (2018). "Pancreatic beta-cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes." Nucleic Acids Res **46**(19): 10302-10318.

Dominissini, D., S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek and G. Rechavi (2012). "Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq." Nature **485**(7397): 201-206.

Drino, A., V. Oberbauer, C. Troger, E. Janisiw, D. Anrather, M. Hartl, S. Kaiser, S. Kellner and M. R. Schaefer (2020). "Production and purification of endogenously modified tRNA-derived small RNAs." RNA Biol **17**(8): 1104-1115.

Gu, H., B. Lian, Y. Yuan, C. Kong, Y. Li, C. Liu and Y. Qi (2021). "A 5' tRNA-Ala-derived small RNA regulates anti-fungal defense in plants." Sci China Life Sci.

Guzzi, N., M. Ciesla, P. C. T. Ngoc, S. Lang, S. Arora, M. Dimitriou, K. Pimkova, M. N. E. Sommarin, R. Munita, M. Lubas, Y. Lim, K. Okuyama, S. Soneji, G. Karlsson, J. Hansson, G. Jonsson, A. H. Lund, M. Sigvardsson, E. Hellstrom-Lindberg, A. C. Hsieh and C. Bellodi (2018). "Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells." Cell **173**(5): 1204-1216 e1226.

Ivanov, P., M. M. Emara, J. Villen, S. P. Gygi and P. Anderson (2011). "Angiogenin-induced tRNA fragments inhibit translation initiation." Mol Cell **43**(4): 613-623.

Li, X., P. Zhu, S. Ma, J. Song, J. Bai, F. Sun and C. Yi (2015). "Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome." Nat Chem Biol **11**(8): 592-597.

Lyons, S. M., D. Gudanis, S. M. Coyne, Z. Gdaniec and P. Ivanov (2017). "Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs." Nat Commun **8**(1): 1127.

Pollex, T., K. Hanna and M. Schaefer (2010). "Detection of cytosine methylation in RNA using bisulfite sequencing." Cold Spring Harb Protoc **2010**(10): pdb prot5505.

Roth, S. H., E. Y. Levanon and E. Eisenberg (2019). "Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity." Nat Methods **16**(11): 1131-1138.

Shi, J., Y. Zhang, D. Tan, X. Zhang, M. Yan, Y. Zhang, R. Franklin, M. Shahbazi, K. Mackinlay, S. Liu, B. Kuhle, E. R. James, L. Zhang, Y. Qu, Q. Zhai, W. Zhao, L. Zhao, C. Zhou, W. Gu, J. Murn, J. Guo, D. T. Carrell, Y. Wang, X. Chen, B. R. Cairns, X. L. Yang, P. Schimmel, M. Zernicka-Goetz, S. Cheloufi, Y. Zhang, T. Zhou and Q. Chen (2021). "PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications." Nat Cell Biol **23**(4): 424-436.

Snider, M. J., D. Lazarevic and R. Wolfenden (2002). "Catalysis by entropic effects: the action of cytidine deaminase on 5,6-dihydrocytidine." Biochemistry **41**(12): 3925-3930.

Thomas, N. K., V. C. Poodari, M. Jain, H. E. Olsen, M. Akeson and R. L. Abu-Shumays (2021). "Direct Nanopore Sequencing of Individual Full Length tRNA Strands." ACS Nano **15**(10): 16642-16653.

Torres, A. G., O. Reina, C. Stephan-Otto Attolini and L. Ribas de Pouplana (2019). "Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments." Proc Natl Acad Sci U S A **116**(17): 8451-8456.

Yoluc, Y., G. Ammann, P. Barraud, M. Jora, P. A. Limbach, Y. Motorin, V. Marchand, C. Tisne, K. Borland and S. Kellner (2021). "Instrumental analysis of RNA modifications." Crit Rev Biochem Mol Biol **56**(2): 178-204.

Zhang, N., S. Shi, T. Z. Jia, A. Ziegler, B. Yoo, X. Yuan, W. Li and S. Zhang (2019). "A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures." Nucleic Acids Res **47**(20): e125.

Zhang, N., S. Shi, X. Wang, W. Ni, X. Yuan, J. Duan, T. Z. Jia, B. Yoo, A. Ziegler, J. J. Russo, W. Li and S. Zhang (2020). "Direct Sequencing of tRNA by 2D-HELS-AA MS Seq Reveals Its Different Isoforms and Dynamic Base Modifications." ACS Chem Biol **15**(6): 1464-1472.

Zhang, N., S. Shi, B. Yoo, X. Yuan, W. Li and S. Zhang (2020). "2D-HELS MS Seq: A General LC-MS-Based Method for Direct and de novo Sequencing of RNA Mixtures with Different Nucleotide Modifications." J Vis Exp(161).

Zhang, X., A. E. Cozen, Y. Liu, Q. Chen and T. M. Lowe (2016). "Small RNA Modifications: Integral to Function and Disease." Trends Mol Med **22**(12): 1025-1034.

Chapter 5: Conclusion and future direction

Conclusion and future direction

Here we showed an example that how the different functional effects can be induced by the different structure conformation between the endogenous small RNAs and its synthetic version without RNA modification. And by adding site specific RNA modification to the synthetic tsRNAs, the synthetic tsRNAs can recapitulate the cellular function of the endogenous tsRNAs. These results highlight the importance of RNA modifications for the proper function of tsRNAs, and the possibility to engineer synthetic tsRNAs to mimic endogenous tsRNAs.

By applying the LC-MC based *de novo* RNA sequencing MLC-seq, we successfully profiled the highly modified tRNAs, with both stoichiometry and positional information of RNA modifications. With the help of MLC-seq, it is possible to decipher the RNA modification map of each endogenous tsRNAs which can provide key information for engineering synthetic tsRNAs to mimic its endogenous sibling in the future, since our results showed that one proper RNA modification is sufficient for the synthetic tsRNAs to mimic the cellular function of endogenous tsRNAs. Moreover, if we can profile the same tsRNA species from different tissues to picture the tissues specificity of RNA modification on endogenous tsRNAs, it will help us to get better understanding about the RNA modification, structure and functionality of endogenous tsRNAs. With other technologies developed for RNA therapeutics, such as antibody-conjugation, LNA, aptamer conjugations (Dammes et al. 2020), our effort may guide future understanding and designing of modified tsRNAs with precise gene regulation toward our needs for precision medicine.

Reference

Dammes, N. and D. Peer (2020). "Paving the Road for RNA Therapeutics." Trends Pharmacol Sci **41**(10): 755-775.