THE UNIVERSITY OF CHICAGO

REVEALING RNA DYNAMICS AND PROCESSING USING ADVANCED MICROSCOPY TECHNIQUES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

AND

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOPHYSICAL SCIENCES

BY

JIACHENG ZHANG

CHICAGO, ILLINOIS

DECEMBER 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS BASED ON WORK PRESENTED IN THIS THESIS[†]

1. **J Zhang**[※], B Chetnani[※], ED Cormack, W Liu, A Mondragón, J Fei. (**2018**) Specific structural elements of the T-box riboswitch drive the two-step binding of the tRNA ligand. *eLife* **7**: e39518
2. S Park, **J Zhang**, MA Reyer, J Zareba, AA Troy, J Fei. (**2018**) Conducting Multiple Imaging Modes with One Fluorescence Microscope. *JoVE (Journal of Visualized Experiments)*, e58320
3. M Bujnowska, **J Zhang**, Q Dai, EM Heideman, J Fei, (**2020**) Deoxyribozyme-based method for absolute quantification of N^6-methyladenosine fraction at specific sites of RNA, *Journal of Biological Chemistry,* **295**:6992

[†]. The following chapters of this dissertation contain sections and figures adopted from the listed publications with modifications. Chapter 1: publication 2; Chapter 2: publication 1; Chapter 3: publication 3.

[※]. Denotes equal contribution

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Professor Jingyi Fei for her guidance and support during these past years. Jingyi has encouraged and motivated me by her passion for science and enthusiasm for challenging researches. She's always been around lab and ready to answer our questions and trouble shoot unexpected problems. I am especially grateful for her trust on me leading challenging projects and the opportunities for all of the collaborations outside of the lab. Her advice on presentations has dramatically improved my skills and it helped a lot on my fear of public speaking. I am really glad that after all these years of devotion in lab, she is able to spend more time with the new member of her family, Fermi, so that she cannot be bothered for a moment with all of the mess we made. This thesis would not have been possible without Jingyi's support and advice.

I would also like to thank members in Fei' lab. It is a great experience to know and work with different people from different part of the world. I want to take the opportunity to thank the individuals who helped me a lot during my PhD. I didn't know anything about microscopy when I joint the lab but Dr. Seongjin Park patiently got my hands on the complicated machinery and helped me to be independent and even creative on manipulating optics and software. I wish him a great career and wonderful life. Carter Stout, Eric Cormack and Magda Bujnowska are UChicago undergrads whom I am super lucky to work with. They are all smart and dedicated on research and I would say I actually

learned more from them. I have no doubt that they are going to have great success pursuing their PhD. Emily Heideman was a technician but she has done much more helping with everyone's experiment and organizing lab stuff. It would take much more time to finish my PhD without her help. I would like to thank Dr. Matt Reyer and Dr. Amine Driouchi for their accompany and humor. I would like to thank all of the collaborators that I have been worked with: Dr. Bhaskar Chetnani, Dr. Dulce Alonso and Prof. Alfonso Mondargón on Tbox project; Dr. Susan Liao, Dr. Mauricio and Prof. Oded Regev on nuclear speckle project. I also appreciate Dr. Zijie Zhang, Dr. Jiangbo Wei and Dr. Hailing Shi in Chuan He's lab for their kind help with assays and unlimited resource of reagents.

I would like to specially thank my Biophysics cohort. I was quite uncomfortable with the new surroundings after coming to U.S. They were the first people I met, yet delivered full security and relax in my toughest first year and got me familiar with everything necessary for graduate school. At the same time, I also appreciate help and advice from Biophysics Admin faculties, Prof. Tobin Sosnick, Prof. Greg Engel, Dr. Adam Hammond and especially Dr. Michele Wittels and Dr. Julie Feder. I couldn't survive without their help.

I would also like to sincerely thank my thesis committee, Prof. Yamuna Krishnan, Prof. Chuan He and Prof. Tobin Sosnick for their insightful input, rigorous requirement and kindly encouragement during my PhD. I always felt refreshed and excited after the committee meetings.

I would like to thank people who have the courage and kindness to befriend me, especially Ruyi Wang, Yining Han, Hailing Shi and Fangfang Yu. They always brought

new thoughts and ideas but we also communicate deep about work, life ect. You can never get tired talking with them and I have been so honored to have friends like them.

There are a special group people whom I never met but we talk basically every day. We knew each other from online gaming during COVID but our relationship has grown far beyond that. I appreciate their tolerance of my arrogance and negative emotions and their positive energy.

Finally, I would like to thank my parents and my family for their invaluable love and support. To be honest, I am not sure if it is worth sacrificing family accompanies for a PhD but I am sure they sacrificed a lot without gaining anything. In their eyes, I have always been a good kid but I hope I can be a good man to be responsible and support my family in the future.

ABSTRACT

RNA molecules are emerging as a more essential role in central dogma of molecular biology. They not only deliver genetic information from DNA to proteins but also participating in gene regulations at different levels. These regulations are mediate by dynamic interactions with other RNA and proteins in certain subcellular context. Fluorescent microscopy is most direct way to obtain the spatial and temporal details of RNA mediated gene regulation. In this thesis, I present my research of studying RNA molecules at three different aspects, RNA-RNA interaction, RNA processing and RNA modifications. In the first part, using single-molecule fluorescence resonance energy transfer (smFRET), we characterized different steps of the dynamic interaction between T-box riboswitch and its ligand tRNA and dissected the structural features that drive the interaction. In the second part, we used diffraction limited and super-resolution microscopy to study the sequence dependence of RNA localization and splicing outcome. In the third part, we developed biochemical assay to quantify the N^6-methyladenosine (m^6A) modification at a given site and further apply it for imaging m^6A modification in situ. Our studies of the RNA dynamics and processing may provide valuable insights into gene regulation as well as new targets for therapeutic purpose.

CHAPTER 1

Introduction

1.1 RNA molecules play diverse roles in biology

Genetic information flows from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) through transcription and to protein through translation, in all living organisms. In order to adapt to a variable environment or stimuli, gene expression can be regulated with different mechanisms at different aspects. RNA molecule, which sits in the middle of the central dogma, not only serves as the intermediate molecule to generate protein product, but also has emerged to be a critical role on regulating the gene expression at transcriptional and translational level.

1.1.1 RNA transcription and processing

In bacterial cells, RNA molecules are transcribed from the same type of RNA polymerase and the transcription is coupling with translation[1]. The RNA polymerase is composed by four subunits, two small α subunits, β and β' subunits and σ subunit which is the key component for transcription initiation. As for transcription termination, bacteria cells adapt two distinct strategies, Rho dependent and independent. Both strategies require palindrome sequences called terminator[1], which forms a stable hairpin structure. In Rho independent pathway, the RNA polymerase reaches the terminator and the hairpin

structure is thermodynamically more favorable than the DNA-RNA pairing within the transcription bubble[2]. Sequences following the terminators are an array of U. Because of the weak A-U base pair, it also promotes disassociation of the RNA polymerase from the DNA template. In Rho dependent pathway, a helicase called Rho attaches to the 5' end of the transcribed RNA and slides on the RNA towards the polymerase. The polymerase stalls at the terminator sequence and Rho is able to catch up and unwind DNA-RNA base pair, leading to the disassociation (Genome). Bacteria mRNA do not undergo much processing. Their degradation is controlled by endoribonucleases, e.g. RNase E[3], and 3'-exoribonuclease, e.g. RNase R[4]. *In B. subtilis,* mRNA decay can happen from the 5' end by 5'-exoribonuclease, e.g. RNase J1[5].

In eukaryotic cells, unlike bacteria cells, only pre-mRNAs are generated through transcription in nucleus. The pre-mRNAs have to undergo processing steps to become mature and be translated in the cytosol. These processing events include capping, splicing and polyadenylation (Figure 1.1). Capping of RNA is coupled with RNA transcription initiation by RNA polymerase II and is required for RNA export and translation and stability [6]. The transcription termination is followed by polyadenylation. In mammalian cells, the polyadenylation is guided by the sequence at the 3' end of the mRNA and performed by polyA polymerase[7]. Two protein factors, cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) that are crucial for polyadenylation can interact with the CTD of the RNA polymerase II. This

indicates that the polyadenylation is part of the termination process[8]. The polyA tail can stabilize the mRNA and promote re-initiation of translation [9].

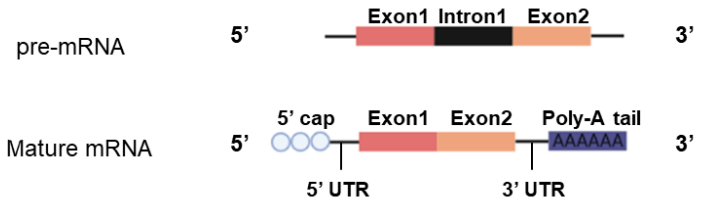


Figure 1.1 RNA processing in eukaryotic cells. Pre-mRNA undergoes 5' capping, 3' polyadenylation and RNA splicing to become mature.

RNA splicing is an essential RNA processing step in eukaryotic cells, during which introns in pre-mRNA are removed and exons are ligated to generate the mature mRNAs. RNA capping and polyadenylation affect RNA stability and translation efficiency. RNA splicing, however in a different perspective, expanded the variation of genetic products in a great scale. During the splicing, particular exons can be included in or excluded from the pre-mRNA to generate different mRNA isoforms encoding different protein, known as alternative splicing. It is estimated that 94% of human genes are alternatively spliced and 50% of the disease mutations are related with splicing[10–12]. The majority of human genes contain introns which can be recognized and process by splicing machinery. Most of these genes are capable of undergoing alternative splicing to generate multiple isoforms of mRNA and protein products. This is stringently controlled by splicing consensus sequences and regulated by cis-acting regulatory sequences and trans-acting proteins within a temporal and cell-specific manner. The consensus sequences contain 5' and 3' splice sites, branch site, polypyrimidine tract and the additional nucleotides of the 5' splice

site[13,14] (**Figure 1.2**). In addition to the consensus sequences, exons and introns have auxiliary elements that can either enhance or repress the corresponding exon inclusion or skipping. In many cases, splicing regulatory elements (SRE) are also involved with regulating the RNA splicing outcomes [15–18]. SREs are short mRNA motifs that can be recognized by their specific splicing factor proteins, including serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoprotein splicing factors (hnRNPs) [19,20]. The splicing factor proteins regulate the splicing outcomes by interacting with SRE and spliceosomes[21] (**Figure 1.2**). However, the exact mechanism underlying the SRE sequence dependent regulation is not fully understood.

Interestingly, the splicing factor proteins shows distinct localization inside the nucleus. SR proteins containing long serine and arginine repeats are more enriched inside nuclear speckles while hnRNP proteins are more likely to be found in the nucleoplasm[22–24]. Nuclear speckles are membrane-less granules inside the nucleus composed by multiple RNA and protein complexes. The RNA and proteins can on one hand form scaffold to stabilize the structure of the nuclear speckle, on the other hand can regulate mRNA processes such as RNA transcription, splicing and exportation[25,26]. RNA splicing is well believed to occur co-transcriptionally, associated with nuclear speckle and directly coupled with RNA exportation[27–29].

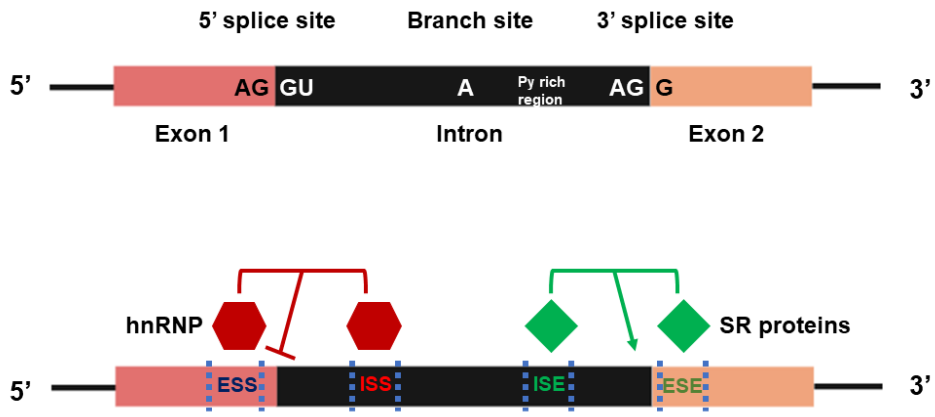


Figure 1.2 Cis-elements and trans-factors affect RNA alternative splicing. (Top) Canonical RNA splicing sequences. (Bottom) Trans-regulatory proteins interact with splicing regulatory elements (SRE) to either enhance or inhibit exon inclusion. ESS = exonic splicing silencer, ISS = intronic splicing silencer, ISE = intronic splicing enhancer, ESE = exonic splicing enhancer.

In addition to those regulation processes, individual nucleosides can be chemically modified by small chemical groups similar as DNA nucleotides. These chemical modifications have been found to exist on mRNA, rRNA lnRNA (long non-coding RNA) and mostly tRNAs and modulate biogenesis, metabolism and functions of these RNAs[30–32]. Using chemical or antibody based high-throughput sequencing methods, over a hundred RNA modifications have been found and they are involved in RNA structure, stability and translation efficiency[33,34] (**Figure 1.3**). Among these RNA modifications, the most prevalent one in mRNA and lnRNA is m^{6}A modification and it is the most well characterized RNA modification so far. It was first discovered early in 1974 but only till 2000s, study of the m^{6}A modification has surged[35,36]. Methyltransferases (m^{6}A writers), demethylases (m^{6}A erasers) and binding proteins (m^{6}A readers) were found, proving the dynamic regulation of m^{6}A modification[37–40]. Taking advantage of the sequencing technique, m^{6}A sites have been successfully mapped on the transcriptome and its unique

stop codon and 3' UTR enrichment in human and mice cells has been revealed[41,42]. People have demonstrated its effect on RNA translation, splicing, stability and export and found association of it with cancer and other diseases [43–50].

Transcriptomic sequencing data have illustrated that an mRNA can have multiple methylation sites, with 3-5 on average, and often clustered. The fraction of m6a modification varies within 10~80% at a given site, with an average of 25%[41,51,52]. The incomplete modification at each site within on RNA potentially leas to large heterogeneity of methylation states of individual RNAs of the same species. However, current sequencing technique is not capable of resolving this heterogeneity and so far it is still not clear what determines the installation and remove of m^6A modification at a particular site. Even though people have created computational models to predict modification sites on the transcriptome, the heterogeneity information is still missing[53,54]. This drawback limits engineering and therapeutic applications targeting the m^6A modification.

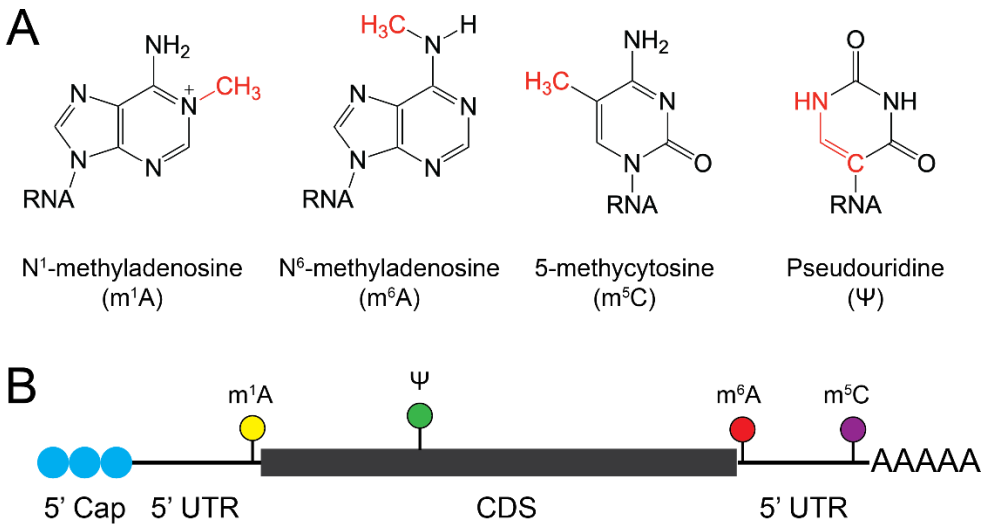


Figure 1.3 Structures and distributions of major mRNA modifications. (A) Chemical structures of known RNA modifications with the chemical modification highlighted in red. (B) Distributions of RNA modifications in (A) along a mRNA transcript.

1.1.2 Regulatory RNAs

Various regulatory RNA molecules in bacteria cells modulate a wide range of physiological responses with different mechanisms. One class of the regulatory RNA molecules are riboswitches that regulate the gene expression in cis. Riboswitches are leader sequences sitting at the 5' end of mRNAs in response to environmental signals, which can be temperature, uncharged tRNA molecules or metabolites in the cells[55,56]. The riboswitches are generally composed by an aptamer domain and an expression platform. The aptamer domain binds with the ligand and the expression platform regulates the downstream gene expression through conformational changes[57,58]. In general, the conformational changes lead to an alternative secondary hairpin structure that can disrupt transcriptional terminators or antiterminators or expose or hide ribosome-binding sites (**Figure 1.4**).

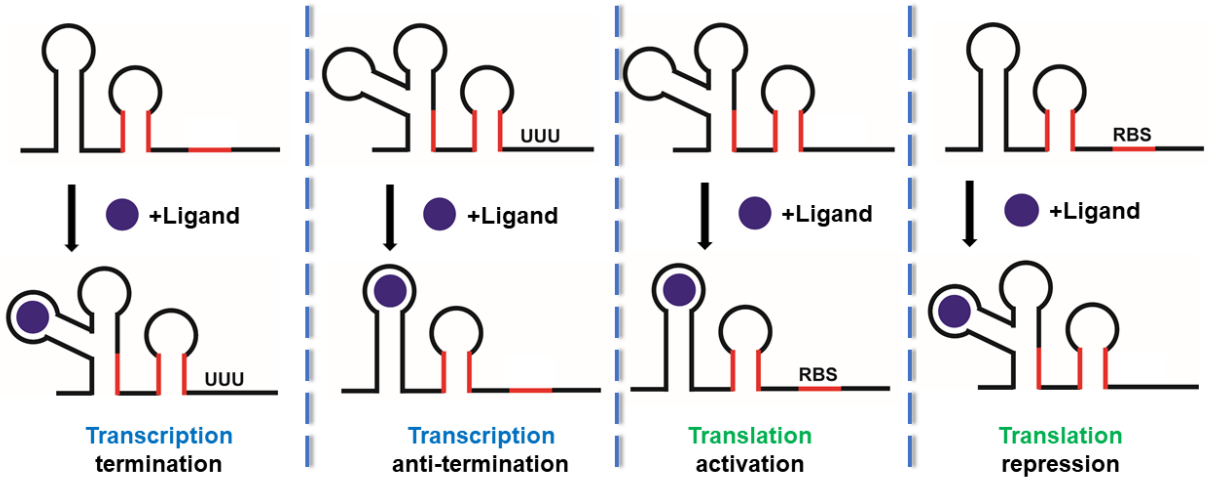


Figure 1.4 Mechanisms of riboswitch mediated gene regulation. (Left panel) Adding the ligand alters the expression platform and forms the terminator conformation to terminate the transcription. (Middle-left panel) Adding the ligand disrupts the terminator conformation allowing transcription to continue. (Middle-right panel) Ligand binding induces conformational changes of the expression platform and exposes the ribosome binding site (RBS) for translation initiation. (Right panel). The RBS is buried in the hairpin upon ligand binding, preventing translation.

Another class of regulatory RNA in bacteria cells are small RNAs (sRNA). Unlike riboswitches, biogenesis of sRNAs is in response to the changes of environmental conditions and gene regulation is through base pairing with target RNAs to modulate their stability and translation[56]. The majority of sRNA utilizes RNA as their target where as a small population of sRNA can bind with protein and modulate its activity[59]. sRNAs targeting RNAs are usually divided into two major classes: cis- and trans-coded sRNAs. The cis-targeting sRNAs are transcribed from one DNA strand that is complementary to the one where the target mRNA is transcribed so that they share an extended region of complementarity[60]. Whereas, the trans-targeting sRNAs are transcribed from an entirely different genomic location thus only share limited complementarity with their target mRNAs[61]. In many cases, the trans-targeting sRNAs require the assistance of chaperone

proteins Hfq or ProQ to stabilize the sRNA-mRNA base pairing[61,62]. Although most of the sRNAs downregulates gene expression, in some cases they can upregulate the expression of their target genes[63]. The regulatory mechanism varies from RNA to RNA but it basically includes modifying RNA transcription, translation and stability[64]. The most common ones are base pairing with ribosome binding site of the target mRNAs inhibiting translation, releasing the internal secondary hairpin structure at the Shine-Dalgarno sequence and activating translation, recruitment of RNase E for target RNA degradation, protecting target RNA from ribonucleases, base pairing with target RNA and promoting transcriptional terminator formation[63,65–69].

In eukaryotic cells, there are much more regulatory RNAs with a large variety of regulation mechanisms. Small nuclear RNAs (snRNAs), transcribed by either RNA polymerase II or III are pivotal cofactors during RNA splicing[70]. U1, U2, U4, U5 and U6 forms canonical spliceosome complex with other RNA and protein molecules. A functionally similar, minor spliceosome is composed by snRNAs, U11, U12, U5 and U4atac and U6atac. Multiple RNA modifications have been identified on the snRNAs and they are involved with snRNA-snRNA and snRNA-mRNA interaction in the process of splicing[71].

In the last two decades, short RNA regulators (20~30 nucleotides) have emerged as a critical role in gene regulation. MicroRNAs (miRNAs) and short interfering RNAs (siRNAs) are the two major categories and it has been shown they can function on

chromatin structure, transcription, RNA processing, RNA stability and translation through a mechanism called RNA interference (RNAi)[72]. The first miRNA, lin-4, was discovered in 1993 to regulate genes controlling development timing of C. elegans[73]. More miRNAs were documented in the 2000s through computational or experimental methods in different species[74]. About 50% of the miRNA genes reside in the introns of other genes and are co-expressed with the host genes[75]. Another number of miRNA genes are from distal regions of protein coding sequences and have their own promotor[72]. The precursors of miRNAs are single strand RNAs transcribed by RNA polymerase II that form a ~33 bp stem-loop. It is first processed by Drosha/Dcl1 proteins in the nucleus to create a pre-miRNA product. In animals, the pre-miRNA is exported to the cytosol and cleaved into a 21-22 bp double strand RNA (dsRNA) product by Dicer protein[76]. In plants, the pre-miRNA is trimmed in the nucleus by Dcl1 and then exported to cytosol[77]. The dsRNA is then loaded on to Ago protein and unwound[78,79]. One strand of it (miRNA*) is discarded and the left strand (miRNA strand) together with the associated protein forms an effector assembly know as RNA-induced silencing complexes (RISCs). The decision of retaining which strand is not fully understood yet, but major factors are the asymmetry of thermodynamic stability from the two ends of the complex and the identities of the 5' nucleotide of each strand. 5' Uracil is more preferred within the binding pocket of Ago and lower thermodynamic stability at the 5' end is also favored[80].

The miRNAs serve as an adaptor of RISCs to recognize and regulate the target RNAs through base pairing. The regulatory mechanism is dependent on the

complementarity of the miRNA with target mRNA[81]. Limited complementarity of base pairing leads to translational repression and fully complementary base paring can result in both translation repression and RNA cleavage by Ago[81,82]. As miRNA binding sites lies at the 3' UTR of animal mRNAs, the mechanism of translational repression is speculated as blocking circularization of mRNA for re-initiation or cap binding. One miRNA can potentially have multiple mRNA targets and one mRNA can be regulated by multiple miRNAs, the reason of which is not yet fully explained[83,84].

siRNAs were initially found in 1998 to inhibit exogenous dsRNA in plants and the sources of siRNAs were found to be virus, transposon or transgene[85,86]. siRNAs were thought to be a defense interference system against exogenous RNAs. However, endogenous siRNAs from genomic transcripts were identified shortly[87]. Similar as miRNA RISC assembly, the precursors were processed sequentially by Dicer and Ago[88]. TRBP protein was recruited to form the RISC complex. The guide strand is retained and the passenger strand is digested by Ago[89,90]. The mechanism of selecting guide strand is similar as miRNA strand selection[91]. However, since the sources of siRNAs are dsRNAs, they require a different type of Dicer protein from the miRNA[92]. The gene silencing mechanism of siRNAs is based on perfect complementary base pairing of the guide strand with the target RNAs followed by precise cleavage at nucleotide 10-11 phosphodiester bond by Ago protein [91]. Cellular exonucleases further digest the RNA fragments after the initial cut[93]. The RISC can also associate with nascent transcripts and recruit histone methyltransferase (HMT) or DNA methyltransferase (DMT) to promote

heterochromatin formation [94]. One extinct feature of the siRNA regulation is the primary siRNA can induce the secondary siRNA synthesis by RNA-dependent RNA polymerase (RdRP)[92]. The amount of siRNA is dramatically amplified to make sure an effective and systematic response to the exogenous RNA.

1.2 Cellular RNA localization

1.2.1 RNA localization in bacteria cells

Bacteria cells have no subcellular compartments yet certain molecules have distinct localization pattern inside the cell. Bacteria genomic DNA forms a compact organization, the nucleoid, facilitated by nucleoid-associated proteins (NAPs). The nucleoid, even though occupies the majority of the space in the cell, is separated from the poles. In addition, it also exhibits a three-dimensional helical ellipsoid structure[95]. The transcriptional machinery, RNA polymerase, was found to be colocalize with the nucleoid[96]. With both fluorescent and electronic microscopy, people have demonstrated that the actively translated ribosomes are close to the cell membrane[97,98]. These observations showed the transcription and translation machineries are well separated and indicated the mRNAs have been relocated from nucleoid to the ribosome, which is contradict with the old idea that transcription and translation are coupled [99]. The localization of the mRNA molecules inside the bacteria cells, therefore, is speculated with two different models. The first model suggests mRNAs stay near the transcription site in the nucleoid whereas the second model suggests the mRNAs are close to where their

protein products function[100]. These two hypotheses have been proven by fluorescent imaging and are not fully contradicting. Results have shown that genomic loci that coding membrane proteins can be repositioned close to the cell membrane upon their expression[101]. Regulatory sRNAs, in contrast to mRNAs, are less compartmentalized examined by FISH. This localization scenario is consistent with the fact that each sRNA species can regulate multiple RNA targets[102].

1.2.2 RNA localization in eukaryotic cells

The eukaryotic cells are much larger than bacteria cells and they form organelles with or without membrane to compartment a large variety of molecules for systematic and effective biofunctions. The molecules can dynamically shuffle in and out these organelles. Many methods have been developed to study RNA localization in eukaryotic cells including fluorescent imaging and fractionation followed by sequencing[103,104]. Recently, APEX-seq was developed and it is capable to screen the RNA localization at transcriptome-wide level [105].

The pre-RNAs are first transcribed and processed to be mature RNA in the nucleus and then retained in nucleus or transported into cytosol. The transcription and maturation of RNAs are highly regulated and involves different ribonucleoproteins that localize in many sub-nuclear membrane-less bodies[106]. These nuclear bodies are formed based on liquid-liquid phase separation (LLPS) and many noncoding RNAs (ncRNAs) provide multivalency for interacting with proteins and other RNAs simultaneously to strengthen

the droplet formation[107,108]. Nucleoli are one of the well characterized nuclear bodies. They are associated with actively transcribing ribosome DNA (rDNA) and packed with many proteins and RNAs. Ribosome RNAs (rRNAs) are transcribed from rDNA and processed by small nucleolar RNAs (snoRNAs) in the nucleoli. In addition, the nucleoli also contribute to biogenesis of multiple ribonucleoprotein particles [108] (**Figure 1.5**).

Nuclear speckles (NS) are another category of nuclear territory. They composed by proteins function on transcription amplification, splicing, RNA modification, 3' end processing and export [109] (**Figure 1.5**). Long noncoding RNAs (lncRNAs), small nuclear RNAs (snRNAs) and mRNAs are found to be existed in the NS. lncRNA *MALAT1,* also known as *NEAT2*, is associate with multiple serine arginine splicing factors (SRSFs), chromatin, and directly interacts with U1 snRNA and is found to be regulating alternative splicing through modulating the phosphorylation of SRSF proteins and regulating gene activation by recruiting corresponding machinery[110–112]. mRNAs with or without introns can localize in the NS and get exported by the transcription export (TREX) components[28,113]. Depletion of the TREX leads to the accumulation of mRNAs inside NS so that NS serves as a hub for coordinating RNA processing and export[109,113].

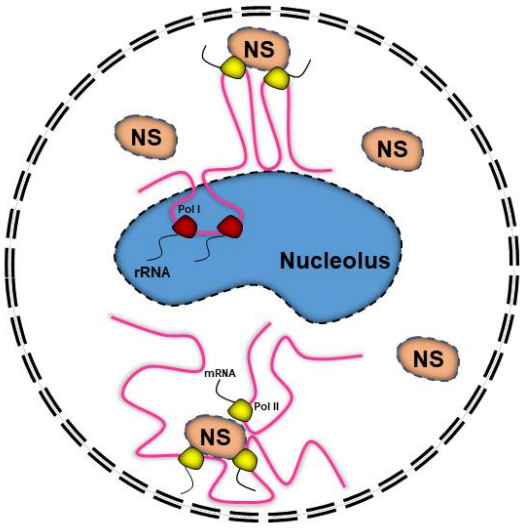


Figure 1.5 Cartoon of membrane-less territories inside nucleus. Ribosome RNAs are generated inside the nucleoli and nuclear speckle is associated with actively transcribing genes.

Both mRNAs and noncoding RNAs are found in cytosol. The localization of cytosolic mRNAs is largely associated with its coding protein localization. To be more specific, mRNAs that are close to endoplasmic reticulum (ER) membrane code protein that localize to the ER, Golgi and vesicles[105]. Interestingly, different transcript isoforms of the same gene exhibit different localizations[105]. A special cytosolic organelle, mitochondria, has its own DNA (mtDNA) and ribosome to express protein. However, only 13 mitochondria proteins are encoded from mtDNA whereas the majority of proteins are from the genomic DNA [114]. The mitochondria-proximal transcripts are identified to code mitochondria proteins. This raises the question whether the accumulation of these transcripts is driven by the local protein translation by mitochondria ribosome or they are transported to the mitochondria post-transcriptionally By treating cells with different translational inhibitors, Fazal et al. have found out that the mitochondria-proximal transcripts are clustered into

two groups, ribosome dependent and ribosome independent, the latter of which may bind directly to RNA binding protein on mitochondria membrane. Regulation of RNA localization and local translation may benefit for the cells in different perspective. First, transporting one mRNA that can be repeatedly used as a template for translation to a distant position require less energy and response time than trafficking multiple protein products, especially in axons[115,116]. Second, it prevents protein product from entering wrong cellular compartments and introducing toxic effects[117,118]. Third, in many cases, protein subunits are clustered together to form a large complex and the local translation ensures the effective complex assembly[119,120]. Therefore, RNA localization is essential for coordinating cellular organization and function.

1.3 Imaging tools to study RNA dynamics in vitro and in vivo

It is increasingly evident that RNA molecules have emerging as diverse roles on regulating gene expression and mediate cellular events. Much more information is required to reveal the biophysical properties of RNA molecules and the complexity of their interactions and functions. Fluorescent imaging has become a great tool to fulfill these requirements since it can provide spatial and temporal information both in vitro and in vivo. Here, I will introduce imaging tools that are commonly applied to label RNA molecules and the microscopy for imaging RNA molecules with different purpose.

1.3.1 RNA labeling in vitro

Even though RNA molecules, same as DNA molecules, are composed by nucleotide chemically, it is well appreciated that RNA is less stable and more difficult being synthesized and labeled. Since in vitro RNA were purified from cell or synthesized from scratch, there is no challenge of resolving spatial resolution or influence from other molecules. Imaging RNA is largely limited by the labeling methods. The most common chemical approach to introduce fluorophore or other modifications into RNA is solid phase synthesis [121]. The step-by-step incorporation of modified nucleotides allow us to insert the modification at different positions of the nucleotide and different sites of the RNA. However, the RNA product length is limited within 100 nucleobases[122]. In order to label a long RNA, mimicking the physiological properties in vivo, post-transcriptional labelling method is widespreadly applied. In vitro transcription is the fastest and cheapest method to obtain a long RNA fragment with desired sequence. The RNA product can be labeled at 5' end, 3' end and internally.

Labeling at 5' requires functionalization and this can be achieved by chemical modification with carbodiimides e.g. 1-Ethyl-3(3-dimethylaminopropyl)carbodiimide (EDC)[123]. The 5' monophosphate end of the RNA can be functionalized by a reaction with EDC, imidazole and ethylenediamine and the reaction results in an amine modified phosphate at the 5' end which can be used for labeling fluorophore by coupling with N-hydroxysulfosuccinimide (NHS ester). Another 5' labeling method utilizes methyltransferases[124,125]. Cap (guanine-N7-)-methyltransferases methylate the N7

position of the 5'-cap with the substrate S-Adenosyl-L-methionine (AdoMet). Using a chemically modified AdoMet, we can incorporate functional groups to the 5' end and labeling the RNA via click chemistry. 3' labeling can be achieved enzymatically as well. An in vitro transcribed RNA can be incorporated with another 2' or 3' modified nucleotide at its 3' end by deoxynucleotidyl transferase (TdT) or polyA polymerase (PAP)[126,127]. The modified nucleotide carries a functional group that can be utilized for click labeling to introduce the fluorophore [128]. And the first 3' modified NTP can block the incorporation of additional NTPs to the 3' end. Internal fluorophore labeling of RNA can be performed with deoxyribozymes[129,130]. A deoxyribozyme is a single strand DNA functions as catalyst. Deoxyribozyme 10DM24 is able to hybridize with both target RNA molecule and a mononucleotide to form a three-helix junction. The deoxyribozyme can further attach the mononucleotide to the 2' hydroxyl group of a specific internal adenosine of an RNA. Using a chemically modified GTP as the mononucleotide, this approach can incorporate functional groups to the internal adenosine and further be labeled with fluorophore through click chemistry. In order to label multiple fluorophores on one RNA molecule, a well-established method is to ligate multiple labeled RNA fragments into one with T4 RNA or DNA ligase and DNA splints that are complementary to the 5' and 3' end of RNA fragments[121]. It is of note that the labeling potentially interferes with the structural dynamics or the binding affinity. This has to be considered during construct design and control experiments are always necessary to rule out these possibilities.

The versatile RNA labeling methods enable studying RNA dynamics and interaction

in vitro by imaging[123]. A common tool is single-molecule fluorescence energy transfer (smFRET)[131]. Energy transfer efficiency can be calculated through the donor acceptor fluorophore intensity or the number of photons received from both. The energy transfer efficiency indicates the distance between the donor- acceptor fluorophore pair with a range normally between 1 and 10 nm, varying with different FRET pair and the surrounding electrochemical condition. Therefore, monitoring the donor and acceptor intensities in a period of time can provide us kinetic information of the distance changes of the donor and acceptor, indicating the structural dynamics of individual RNA molecule or interactions between two different RNA molecules. There are mainly two ways to perform smFRET: confocal and total internal reflection fluorescence (FRET) microscopy [131]. In confocal smFRET, molecules are free diffusing in the solution and only one molecule in a small volume of the solution is observed, and the FRET efficiency in a short time frame can be determined[132]. Time resolution below milliseconds is possible for confocal smFRET. However, structural transitions or molecule interactions which take from milliseconds to seconds cannot be monitored because of the fast-diffusing property of the molecules. In TIRF smFRET, the molecules are anchored on a glass imaging chamber and the camera acquires the intensity profile of the whole imaging area which contains hundreds or thousands of molecules. The time resolution, therefore, is limited to the order of 10 to 100 milliseconds rage. The molecule immobilization requires additional modifications on the molecules. For nucleic acids, attaching biotin to DNA or RNA enables forming a stable biotin-streptavidin complex on the imaging chamber[133].

Another way to investigate molecule-molecule interaction by imaging is through two-color colocalization analysis. In this case, both fluorophores are excited simultaneously and the interaction of the two molecules is detected by the coincident colocalization of both fluorophores[134]. Therefore, the labeling of the two fluorophores does not have to fall into the FRET range and this method is better for constructs lacking structural information but cannot fully reveal the dynamic transitions as smFRET.

1.3.2 RNA labeling in vivo

In order to study the spatial distribution and quantify the copy number of target RNA molecule in cells, researchers have developed single molecule fluorescent in situ hybridization (smFISH)[135,136]. In general, the cells are fixed and permeabilized, a set of DNA probes labeled with fluorophore are added to the cells to bind with the target RNA through base pairing. For low copy number of RNAs, diffraction limited images can be used to quantify the copy number. For super abundant of RNAs, super-resolution (SR) imaging methods is capable to resolve the sub-limited distribution pattern and quantify the copy number. This method can reveal the heterogeneity of RNA distribution among different cells. However, it is not able to provide temporal information of individual RNA.

RNA molecules, unlike protein, cannot be covalently fused with fluorescent molecules in live cells. In order to label and track RNA in live cells, the target RNA is usually fuse with RNA aptamers[137]. An example would be a well-characterized RNA

hairpin sequence (e.g. MS2, PP7, etc.) on the 3' UTR[138,139]. The hairpin can be recognized by its protein binding partner which can be fused with a fluorescent protein. To increase the signal to noise ratio, multiple copies of the RNA hairpin sequence are fused to the target RNA and multiple fusion proteins can bind to the RNA. Using this tagging method, people are able to monitor RNA processing such as transcription[139,140], splicing[141] and even translation[142–144] dynamics. Except for using RNA hairpin sequence, researchers have also leveraged CRISPR targeting technology and successfully labeled target RNA in vivo. Currently, the fluorescent signal is either tagged on CRISPR-Cas13 protein[145] or on the gRNA directly[146]. In addition to the protein tagging method, a different type of RNA aptamer was developed to avoid large repeats of the RNA hairpin sequences[137]. These RNA aptamers themselves and their specific ligands alone do not contain fluorescent signal. So that a major advantage of these aptamers is the low fluorescent background in the absence of the aptamers[147]. It is also a protein free tagging system in which the sequence of the aptamers can be optimized by cycles of transcription, selection and reverse transcription[148]. The two benefits also allow quantification of the RNA signal more directly and easier compared to quantifying with GFP. So far, several RNA aptamers for imaging are generated based on 4-hydroxybenzylidene imidazolinone (HBI), a key component for GFP fluorescence, and its derivatives [147,149,150]. With many modifications to improve the brightness though, these RNA aptamers are questionable to image low copy number mRNAs so multiple arrays of them are still required.

1.3.3 Microscopy techniques of imaging RNA molecules

Fluorescence microscopes are important tools for the modern biological science research and fluorescent imaging is routinely performed in many biology laboratories. By tagging biomolecules of interest with fluorophores, we can directly visualize them under the microscope and record the time-dependent changes in the localization, conformation, interaction, and assembly state in vitro and in vivo. Conventional fluorescence microscopes have a diffraction-limited spatial resolution, which is ~200–300 nm in the lateral direction and ~ 500-700 nm in the axial direction, and are, therefore, limited to imaging at the 100s of nanometers-to-micron scale[151,152]. In order to reveal finer details in the molecular assembly or organization, various super-resolution (SR) microscopies that can break the diffraction limit have been developed. Strategies used to achieve SR include non-linear optical effect, such as stimulated emission depletion (STED) microscopy[153,154] and structured illumination microscopy (SIM)[155–157], stochastic detection stochastic detection of single molecules, such as stochastic optical reconstruction microscopy (STORM)[158] and photoactivated localization microscopy (PALM)[159], and a combination of both, such as MINFLUX[160]. Among these SR microscopies, single-molecule detection-base SR microscopes can be relatively easily modified from a single-molecule microscope set-up. With repetitive activation and imaging of photoactivatable fluorescent proteins (FPs) or photo-switchable dyes tagged on biomolecules of interest, spatial resolution can reach 10-20 nm[161]. To gain information on molecular interactions and conformational dynamics in real-time, angstrom-to-nanometer resolutions is required.

smFRET is one approach to achieve this resolution[133,162]. Generally, depending on the biological questions of interest, imaging methods with different spatial resolutions are needed.

Typically, for each type of imaging, specific excitation and/or emission optical configuration is needed. For instance, one of the most commonly used illumination methods for single-molecule detection is through total internal reflection (TIR), in which a specific excitation angle needs to be achieved either through a prism or through objective lens. For smFRET detection, emissions from both donor and acceptor dyes need to be spatially separated and directed to different parts of the electron-multiplying, charge-coupled device (EMCCD), which can be achieved with a set mirrors and dichroic beam splitters placed in the emission path. We built a cost-effective, hybrid system that provides adjustable and reproducible switches among three different imaging method: conventional epi-fluorescent imaging with diffraction-limited resolution, single-molecule detection-based SR imaging and multi-color single-molecule detection, including smFRET imaging (Fig). Specifically, the set-up contains fiber-coupled input lasers for multi-color excitation and a commercial illumination arm in the excitation path, which allows programmed control of the excitation angle, to switch between epi-mode and TIR mode. In the emission path, commercial beam splitter is placed before an EMCCD camera that can be selectively enabled to detect multiple emission channels simultaneously.

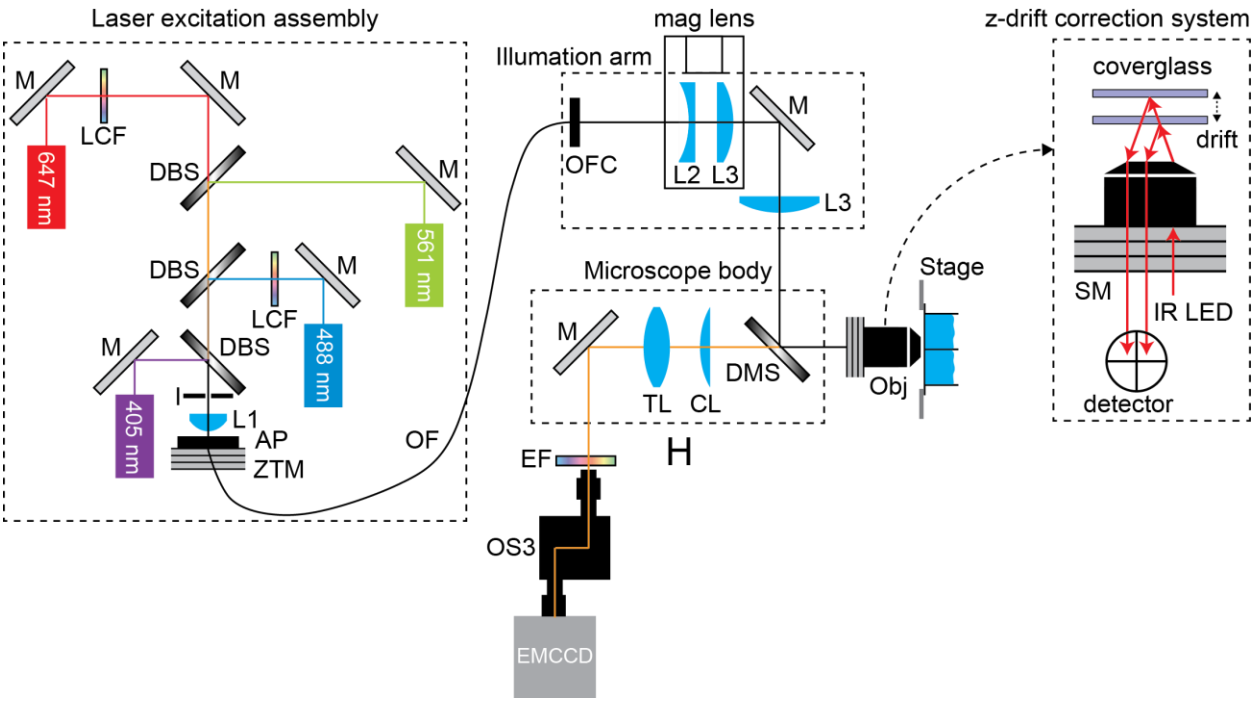


Figure 1.6 The design of the microscope set-up. M=mirror, DBS = dichronic beam splitter, LCF = laser clean-up filter, I = Iris, L = lens, AP = adapter plate, ZTM = z-axis translational mount, OF = optical fiber, OFC = optical fiber coupling, CL = cylindrical lens, TL = tube lens, EF = emission filter(s), Obj = objective lens, and SM = stem motor. The z-drift correction system moves the step motor on the nosepiece of the objective lens to the opposite direction of the z-drift, which is calculated by the infrared (IR) signal generated by its own LED and detected by its own detector.

1.4 Scope of this thesis

Gene expression is regulated at RNA level from different aspects. In this thesis, I will be focusing on the study of RNA interaction, modification and processing in different biological system.

Chapter 2 presents the hierarchy and dynamics of T-box riboswitch and tRNA interaction followed by co-transcriptional model for gene expression.

Chapter 3 presents the method development for quantification and imaging m^6A modification.

Chapter 4 presents the early study on the association of RNA motif with RNA localization and splicing outcome.

CHAPTER 2

Hierarchy and dynamics of Tbox riboswitch and tRNA ligand interaction

2.1 Introduction

As introduced in section 1.2, riboswtiches are cis-regulatory RNA elements that recognize and respond to defined external signals to affect transcription or translation of downstream messenger RNAs (mRNAs). Riboswitches generally consist of two domains: a sensory or aptamer domain and a regulatory domain or expression platform[163]. T The aptamer of each riboswitch class contains conserved sequence motifs and unique secondary or tertiary structural elements that help distinguish and bind specific ligands. Bacterial T-box riboswitches represent a unique class of riboswitches that do not bind small molecule ligands, instead they recognize and bind tRNA molecules and sense directly their aminoacylation state[164]. T-box riboswitches serve as excellent paradigms to understand RNA-RNA interactions and RNA-based regulation.

T-box riboswitches are found in Gram-positive bacteria and are usually located in the region upstream of mRNA sequences encoding aminoacyl tRNA synthetases and proteins involved in amino acid biosynthesis and transport and hence participate directly in amino acid homeostasis[164]. In general, the aptamer domain of all T-box riboswitches contains a long stem, Stem I, responsible for specific tRNA binding[165]. The expression platform can adopt either a terminator or anti-terminator conformation, depending on whether the bound tRNA is charged or uncharged[164,166]. In most T-box riboswitches, binding of a charged tRNA to the T-box leads to rho independent transcription termination whereas an uncharged tRNA stabilizes the anti-terminator conformation and leads to transcription read-through[164,166]. Whereas Stem I and the anti-terminator domain are highly conserved among T-box riboswitches, the region connecting them can vary. The

Bacillus subtilis glyQS T-box riboswitch, involved in glycine regulation, represents one of the simplest T-box riboswitches[167]. Only a short linker and a small stem, Stem III, connect Stem I and the anti-terminator domain (**Figure 2.1**).

Recognition of tRNA by a T-box riboswitch involves three main structural elements of the tRNA: the anticodon region, the "elbow" region formed by the conserved T- and D-loops, and the 3' NCCA sequence (**Figure 2.1**). The anticodon and elbow regions of the tRNA interact with Stem I directly. Stem I contains several phylogenetically conserved structural motifs[165], including a K-turn motif, a specifier loop, a distal bulge, and an apical loop[165] (**Figure 2.1**). Bioinformatics and structural analyses have collectively revealed the interactions between Stem I and the tRNA[168–170]. Specifically, the co-crystal structures of Stem I/tRNA complexes show that Stem I flexes to follow closely the tRNA anticodon stem and interacts directly with the anticodon loop and the elbow through its proximal and distal ends, respectively[170]. The distal bulge and the apical loop fold into a compact structural module of interdigitated T-loops[171,172], which interact directly with conserved unstacked nucleobases at the tRNA elbow[168,170]. In addition, the structures revealed that Stem I turns sharply around two hinge regions using a conserved dinucleotide bulge and the K-turn motif[170,173]. Sensing of the aminoacylation state involves direct binding of the tRNA 3' end to a highly conserved bulge in the T-box, the t-box sequence[174] (**Figure 2.1**). A free NCCA end can base pair with the t-box sequence, enabling the anti-terminator conformation, whereas a charged NCCA end prevents the formation of the t-box/NCCA interactions, leading to the more stable terminator conformation[164,166]. Importantly, discrimination between the charged and uncharged tRNA does not require any additional proteins, such as EF-Tu[175], and is driven solely by RNA/RNA interactions.

Although there are no atomic-level structural details on the interactions between tRNA and the anti-terminator region, Small Angle X-ray Scattering (SAXS)-derived models of the entire *B. subtilis glyQS* T-box riboswitch in complex with tRNA are available[176,177]. The two models are distinct, one presenting a more compact structure where all the previously

observed interactions between Stem I and tRNA are preserved and the 3' NCCA sequence of the tRNA helps stabilize a coaxial stem formed by Stem III and the anti-terminator region[176], while the second model shows a more extended and relaxed structure where the interactions with the anticodon are preserved but the contacts with the tRNA elbow are not present[177]. In addition, there is a dearth of information on the kinetics of the binding process. Whereas it is clear that tRNA recognition involves several specific interactions, their binding temporal sequence remains elusive. In addition, it is unclear whether sensing of the 3' end of the tRNA involves any additional conformational changes in the T-box. Here, by introducing donor-acceptor fluorophore pairs at several locations in the tRNA and the T-box riboswitch, and using single-molecule fluorescence resonance energy transfer (smFRET), we demonstrate the temporal order of events in the trajectory of tRNA binding. Our results demonstrate that tRNA binds to the riboswitch in two steps, with its anticodon being recognized first, followed by NCCA binding accompanied with an inward motion of the 3' region of the T-box riboswitch, including Stem III and the anti-terminator stem, relative to Stem I. In addition, by introducing mutations at different locations of the T-box, we further show that the two-step binding kinetics are regulated by the modular structural elements in the T-box riboswitch.

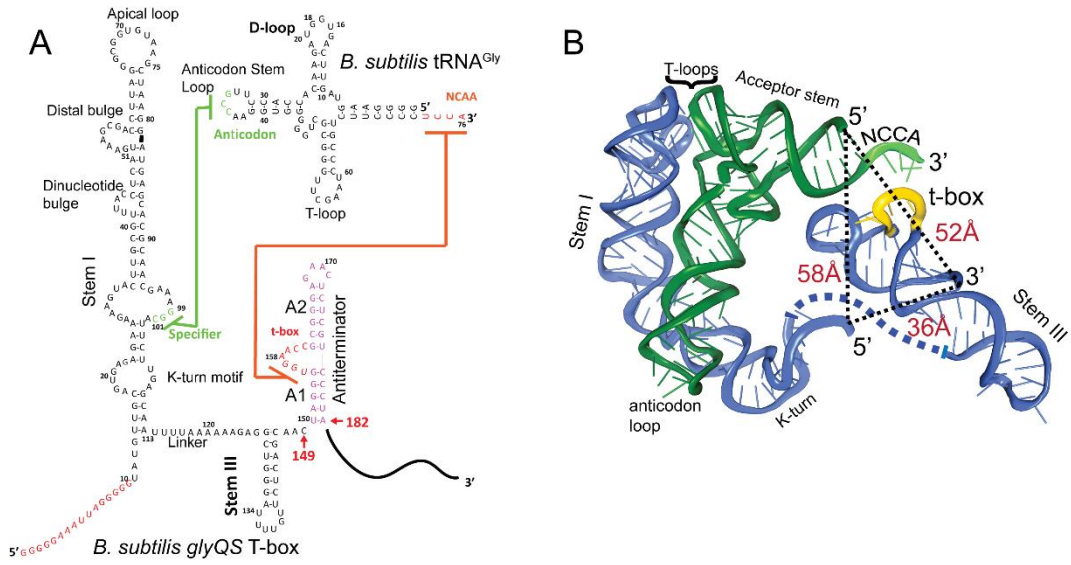


Figure 2.1. Secondary and tertiary structures of *B. subtilis glyQS* T-box riboswitch and tRNA^{Gly}. **(A)** Secondary structure diagrams of the *B. subtilis glyQS* T-box riboswitch and *B. subtilis* tRNA^{Gly} used in this study. Green and orange lines indicate interactions between the T-box specifier loop and the tRNA anticodon and between the T-box t-box sequence and the tRNA 3' NCCA, respectively. The sequence of *glyQS* T-BOX in red is added for surface immobilization. **(B)** Ribbon diagram of a model of a complex between the *B. subtilis glyQS* T-box riboswitch (blue) and *B. subtilis* tRNA^{Gly} (green) based on SAXS data. Distances between the 5' and 3' ends of the T-box and the 5' end of the tRNA^{Gly} are shown (black dash lines). The NCCA sequence at the 3' end of the tRNA is shown in light green and the t-box sequence in the T-box is shown in yellow.

2.2 Results

2.2.1 Binding of cognate tRNA by the glyQS T-box results in two distinct FRET states

To observe directly the binding of tRNA to the T-box, we placed the donor dye (Cy3) on the 3' end of a T-box fragment (T-box$_{182}$), and the acceptor dye (Cy5) on the 5' end of the tRNA^{Gly}, where the subscript "182" denotes the length of the T-box construct (**Figure 2.1A**). T-box$_{182}$ spans Stem I, the linker sequence, Stem III and the anti-terminator, but

does not contain the terminator sequence, thereby preventing the transition to the terminator conformation. A short RNA extension sequence was added to the 5' end of the T-box for surface immobilization (**Figure 2.2A, Figure S2.1.1**). Single-molecule fluorescence images were recorded under equilibrium condition in the presence of 30nM $tRNA^{Gly}$-Cy5. Binding of $tRNA^{Gly}$-Cy5 results in a major distribution of FRET values around 0.7, with 79±4% of the traces showing a stable signal at 0.7 and (9±5)% traces sampling from 0.7 to 0.4 (**Figure 2.2B, C**). The SAXS model[176] predicts a distance between the labeling positions at the 3' end of the $T\text{-}box_{182}$ and the 5' end of the tRNA to be around 52 Å. (**Figure 2.1B**). Based on a Förster distance of 54-60 Å[178,179], our measured FRET value is within the range of estimated FRET values (0.56-0.70). Therefore, we assign the 0.7 FRET state to be the fully bound state of the $tRNA^{Gly}$ by the T-box.

In order to assign the 0.4 FRET value to specific tRNA binding states, $tRNA^{Tyr}$-Cy5 and $tRNA^{\Delta NCCA}$-Cy5 ("ΔNCCA" denotes a $tRNA^{Gly}$ with deleted 3' NCCA sequence) were flowed in the flow-chamber with pre-immobilized $T\text{-}box_{182}$-Cy3(3') (3' denotes that the label was added at the 3' end). We did not observe any binding of $tRNA^{Tyr}$-Cy5 (**Figure S2.2.1**), confirming that recognition of the anticodon by the specifier region is required for tRNA binding. In the presence of $tRNA^{\Delta NCCA}$-Cy5, we observed a fluctuating signal between 0.4 and 0 FRET (**Figure 2.2B, D**), with an average lifetime of the 0.4 FRET state of 3.6±0.6 s and an average waiting time before binding of 31.3±5.3 s (**Figure S2.2.2B**). Taken together with the results from the $tRNA^{Gly}$, $tRNA^{\Delta NCCA}$ and $tRNA^{Tyr}$ binding

experiments, we assign the 0.4 FRET state to a partially bound state where only the anticodon interactions have been established.

To further confirm the assignment of the FRET states, we generated T-box$_{149}$, where the anti-terminator sequence is truncated (**Figure 2.1A, Figure S2.1.1**). Based on the structure model from the SAXS data[176] we predicted that a Cy3 dye placed either at the end of Stem III (T-box$_{149}$) or at the end of the anti-terminator stem (T-box$_{182}$) are localized in close proximity in three dimensions, further confirmed by the distance measurement using smFRET (**Figure S2.2.3**). Therefore, we expect that if tRNA^{Gly}-Cy5 can reach the same fully bound state in T-box$_{149}$ as in T-box$_{182}$, a high FRET state centered at 0.7 would be observed. However, using T-box$_{149}$-Cy3(3') in combination with tRNA^{Gly}-Cy5, we again observed transient binding of tRNA^{Gly} with a FRET value centered at ~0.4 with the same average lifetime as observed with the T-box$_{182}$-Cy3(3') and tRNA$^{\Delta NCCA}$-Cy5 combination (**Figure 2.2B-D, S2.2.3C**). Therefore, these two complexes (T-box$_{182}$ + tRNA$^{\Delta NCCA}$ and T-box$_{149}$ + tRNA^{Gly}) represent the same binding state of the tRNA, i.e. the state where binding of the anticodon to the specifier region has been established, but is unstable without the further interactions between the NCCA and the t-box region.

Collectively, our results suggest a two-step binding model where the establishment of the interaction with the anticodon precedes the interactions with the NCCA. Without the interaction between the NCCA and the t-box sequence the binding of tRNA^{Gly} is not stable. From the binding kinetics of tRNA$^{\Delta NCCA}$, we estimated the association rate

constant (k_1) and the disassociation rate constant (k_{-1}) for the first binding step to be (5.0±1.7) x 10^5 $s^{-1}\cdot M^{-1}$ and 0.28±0.04 s^{-1}, respectively (**Figure 2.5; Figure 2.6E**).

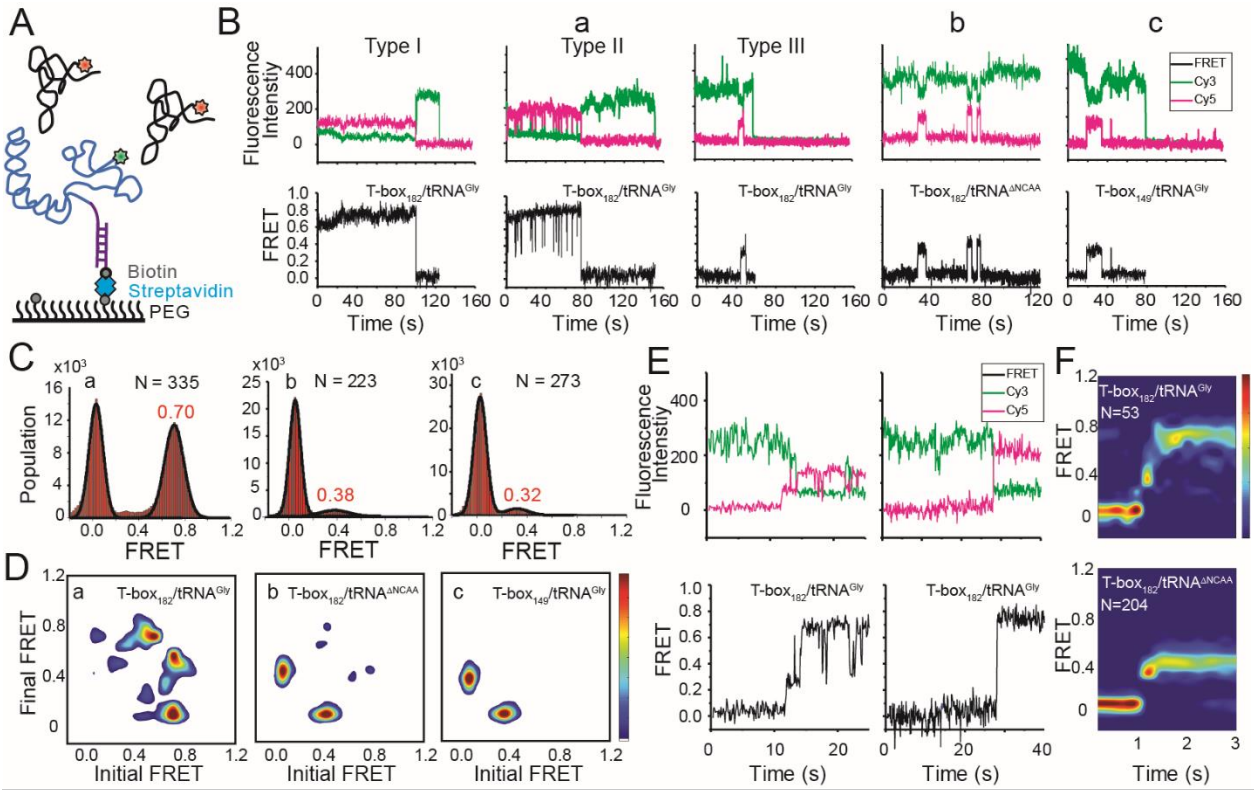


Figure 2.2. Two-step binding of uncharged tRNA to the *glyQS* T-box riboswitch. (A) FRET labeling scheme for the T-box and tRNA. Cy3 (green star) and Cy5 (red star) fluorophores are attached at the 3' of the T-box (blue) and the 5' of the tRNA (black), respectively. *glyQS* T-box riboswitch molecules are anchored on slides through a biotinylated DNA probe (purple) hybridized to a 5' extension sequence on the T-box. **(B)** smFRET *vs.* time trajectories of T-box$_{182}$-Cy3(3') with (a) tRNA^{Gly}-Cy5 or (b) tRNA$^{\Delta NCCA}$-Cy5 and (c) T-box$_{149}$-Cy3(3') with tRNA^{Gly}-Cy5. Cy3 and Cy5 fluorescence intensity traces (upper panel), and their corresponding smFRET traces calculated as I_{Cy5} / (I_{Cy3}+I_{Cy5}) (lower panel). **(C)** One-dimensional FRET histograms of combinations (a), (b) and (c) as described above. FRET peaks are fit with a Gaussian distribution (black curve) and the peak centers are shown in red. "N" denotes the total number of traces in each histogram from 3 independent experiments. **(D)** Transition density plot (TDP) of combinations (a), (b) and (c). Contours are plotted from blue (lowest population) to red (highest population). TDPs are generated from all smFRET traces from 3 independent experiments. **(E)** Representative smFRET trajectories showing real-time binding of tRNA^{Gly}-Cy5 to T-box$_{182}$-Cy3(3') in a steady-state measurement. Traces showing transitions from unbound state (0 FRET) to partially bound state (0.4 FRET) to fully bound state (0.7 FRET) (left) and unbound state directly to fully bound state (right). **(F)** Surface contour plot of time-evolved FRET histogram of T-box$_{182}$-Cy3(3') with tRNA^{Gly}-Cy5 (top) and tRNA$^{\Delta NCCA}$-Cy5 (bottom). "N" denotes the total number of traces in each histogram from 3 independent experiments, which are a subset of traces showing real-time binding event in the steady-state measurements. Total numbers of traces in each data sets are indicated in **(B)**. Traces that reach the 0.7 FRET state (cutoff >0.55) are included in the plot for tRNA^{Gly}-Cy5 to reveal better the transition from the 0.4 to the 0.7 FRET state. Time-evolved FRET histograms of all traces are shown in **Figure S2.2.4D** for comparison.

2.2.2 The transition from anticodon recognition to NCCA binding is rapid for uncharged tRNA

We classified smFRET traces for T-box$_{182}$-Cy3(3') in complex with tRNA^{Gly}-Cy5 into three types (**Figure 2.2B**): (I) traces stably sampling the 0.7 state (79±4%), (II) traces transiently transitioning from the 0.7 state to the 0.4 state (9±5%), and (III) traces only sampling the 0.4 state without reaching 0.7 state (12±5%). The low percentage of Type III traces indicates that once the anticodon is recognized, the commitment to the next binding step, NCCA interactions, is high. The majority of the traces showed that the tRNA^{Gly} remained mostly in the fully bound state (Type I) until the fluorophore photobleached, with the actual lifetime limited by photobleaching ($T_{0.7}$ > 24 s, where $T_{0.7}$ denotes the lifetime of the 0.7 FRET state) (**Figure S2.2.2A**). The observation that tRNA^{Gly} is able to transit from the fully bound state back to the partially bound state (Type II) suggests that the NCCA/t-box interaction can break occasionally (**Figure 2.2B**). We estimated the lifetime of the transiently sampled partially bound state to be 0.35±0.09 s (**Figure S2.2.2A**), ~10-fold shorter than the partially bound state without the NCCA end.

While the majority of the T-box molecules were already bound to tRNA^{Gly} before starting data acquisition, we could detect that some molecules show real-time binding during imaging acquisition. We observed only a few traces briefly sampling the 0.4 FRET state from the zero FRET (unbound) state before reaching the 0.7 FRET state, while most traces directly sampled the 0.7 FRET state without a detectable 0.4 FRET, likely due to our imaging time resolution (100 ms per frame). We post-synchronized the FRET traces

at the transition point from the zero FRET state to the first sampled 0.4 FRET state, and plotted in a time-evolved FRET histogram. From the time-evolved FRET histogram (**Figure 2.2F**), we estimated roughly that the upper limit of the lifetime spent at the 0.4 FRET state is ~100 ms, very rapidly followed by establishment of NCCA/t-box interactions. In contrast, $tRNA^{\Delta NCCA}$ could not pass the 0.4 FRET state. To capture better real-time binding, we performed a flow experiment, where $tRNA^{Gly}$-Cy5 was flowed in to a chamber with immobilized $T\text{-}box_{182}$-Cy3(3') during imaging acquisition. The corresponding post-synchronized time-evolved FRET histogram again shows a fast transition into the fully bound state (**Figure S2.2.4**). In addition, the association rate constant of $tRNA^{Gly}$ in the real-time flow experiment is $(7.5\pm0.7) \times 10^5\ s^{-1}\cdot M^{-1}$, consistent with the k_1 of $tRNA^{\Delta NCCA}$ and confirming that the NCCA end of the tRNA does not participate in the first binding step.

From the real-time binding kinetics of $tRNA^{Gly}$ to $T\text{-}box_{182}$, we estimated a transition rate constant from the partially bound state to the fully bound state (k_2) of ~10 s^{-1} (**Figure 2.2F, Figure S2.2.4**). On the other hand, as transitions back to the partially bound state from the fully bound state were only observed in ~10% traces, we interpreted this to mean that the reverse transition rate constant (k_{-2}) is very small, and the second binding step in the wild-type (WT) T-box with uncharged $tRNA^{Gly}$ is close to irreversible (**Figure 2.5; Figure 2.6E)**.

2.2.3 Establishment of the NCCA/t-box interaction is accompanied by conformational changes in the T-box riboswitch

We next investigated whether tRNA binding requires any conformational changes in the T-box itself. Using doubly labeled T-box$_{182}$, with Cy3 at the 3' end and Cy5 at the 5' hybridization extension, we observed a high FRET state (centered at ~0.75) in the absence of tRNA (**Figure S2.3.1**). Based on the structural model (17), we estimated the distance between the 5' and 3' ends of the T-box$_{182}$ to be ~36 Å (**Figure 2.1B**). Our measured FRET value is slightly less than the predicted FRET value (~0.9), likely due to the engineered 5' extension sequence used to immobilize the T-box. No noticeable change was detected upon incubation with unlabeled tRNA^{Gly} (**Figure S2.3.1**), indicating that the 3' portion (Stem III plus the anti-terminator stem) does not move away from the 5' portion (Stem I). Given that the measured FRET efficiency of 0.75 is already located beyond the FRET sensitive region, it is unlikely that any inward motion of the 3' portion relative to the 5' could be detected. To overcome this limitation, we added extensions at both the 3' and 5' ends (**Figure 2.3A, Figure S2.1.1**). ITC experiments suggest that addition of a 5' and/or a 3' extension sequences to the T-BOX does not affect tRNA binding (**Figure S2.3.2**). With this intra-T-box FRET scheme, we observed a FRET shift from ~0.5 to ~0.65 when tRNA^{Gly} was added (**Figure 2.3B**), indicating that the 3' half of the T-box moved closer to the 5' half, potentially with the T-box becoming more compact due to the presence of the cognate tRNA^{Gly}. Adding non-cognate tRNA^{Phe} or tRNA$^{\Delta NCCA}$ gave similar FRET values as the T-box alone (**Figure 2.3B**), suggesting that the conformational

change is associated with binding of the NCCA, not with anticodon recognition.

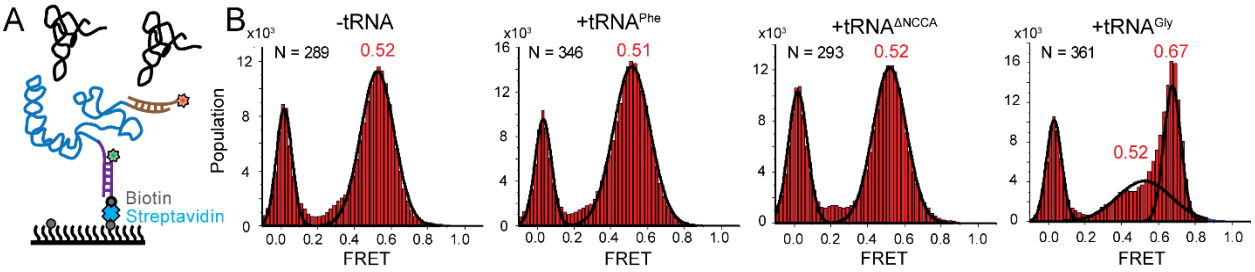


Figure 2.3. Conformational changes following tRNA binding in the *glyQS* T-box riboswitch revealed by an intra-T-box FRET pair. (A) Intra-T-box FRET scheme. Cy3 (green star) and Cy5 (red star) are attached at the 5' and 3' extensions of T-box (blue), respectively. **(B)** One-dimensional FRET histograms of T-box$_{182}$ alone, with tRNA^{Phe}, with tRNA$^{\Delta NCCA}$, and with tRNA^{Gly}. "N" denotes the total number of traces in each histogram from 3 independent experiments.

2.2.4 The NCCA end of the uncharged tRNA maintains its relative position to the K-turn region during the second binding step

Using the above two FRET pairs, we observed that the 3' portion of the T-box moves towards the base of Stem I as well as the NCCA end of the tRNA during the second binding step. To ascertain whether the NCCA end of the tRNA also moves relative to the base of Stem I, we measured FRET between a Cy3 placed at the 5' end of the T-box (T-box182-Cy3(5')) and tRNA^{Gly}-Cy5 (**Figure 2.4A**). Using this FRET pair, binding of both tRNA^{Gly} and tRNA$^{\Delta NCCA}$ generated a similar FRET value centered at ~0.35 (**Figure 2.4B, C**). However, the FRET traces behaved differently for these two tRNA molecules. For tRNA$^{\Delta NCCA}$, the signal fluctuated between zero and 0.35 (**Figure 2.4D**), with a lifetime of the 0.35 FRET state of 4.5 ± 1.0 s, reminiscent of the 0.4 FRET state using the tRNA/T-box182-Cy3(3') FRET pair **(Figure S2.4.1**). For tRNA^{Gly}, the signal was more stably centered at 0.35 (**Figure 2.4B**). Since the tRNA-Cy5/T-box182–Cy3(5') FRET pair cannot

distinguish the partially bound from the fully bound state, we fit the lifetime with a double-exponential decay. The fast dissociation fraction has a lifetime of 3.9 ± 0.7 s (46 ± 21% of population), consistent with the lifetime for the partially bound state, and the low dissociation fraction has a lifetime of 15.7 ± 0.8 s (54 ± 21%), representing the stable fully bound state (**Figure S2.4.1**). Overall, the measurements with the tRNA-Cy5/T-box–Cy3(5') FRET pair further validate the two-step binding model and reveal that the NCCA end of the uncharged tRNA maintains its relative position to the base of Stem I during the second binding step.

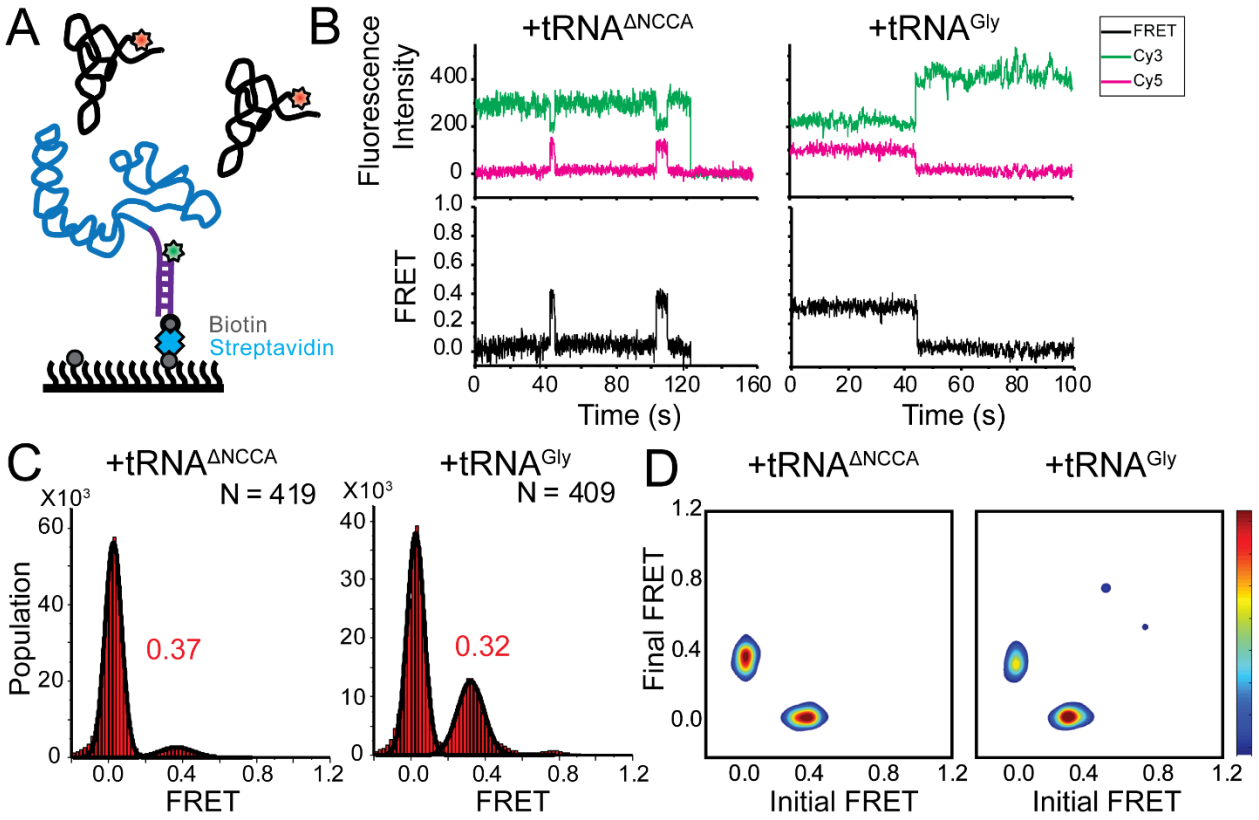


Figure 2.4. FRET between fluorophores at the 5' end of the *glyQS* T-box riboswitch and 5' end of tRNA^{Gly} is insensitive to the two binding states. (A) Cy3 (green star) and Cy5 (red star) are attached at the 5' extension of the T-box (blue) and the 5' of the tRNA (black), respectively. **(B)** smFRET trajectories of T-box-Cy3(5') with tRNA$^{\Delta NCCA}$-Cy5 (left) and tRNA^{Gly}-Cy5 (right). "N" denotes the total number of traces in each histogram from 3 independent experiments. **(C)** One-dimensional FRET histograms of T-box-Cy3(5') with tRNA$^{\Delta NCCA}$-Cy5 (left) and tRNA^{Gly}-Cy5 (right). **(D)** TDP of T-box-Cy3(5') with tRNA$^{\Delta NCCA}$-Cy5 (left) and tRNA^{Gly}-Cy5 (right). TDPs are generated from all smFRET traces from 3 independent experiments.

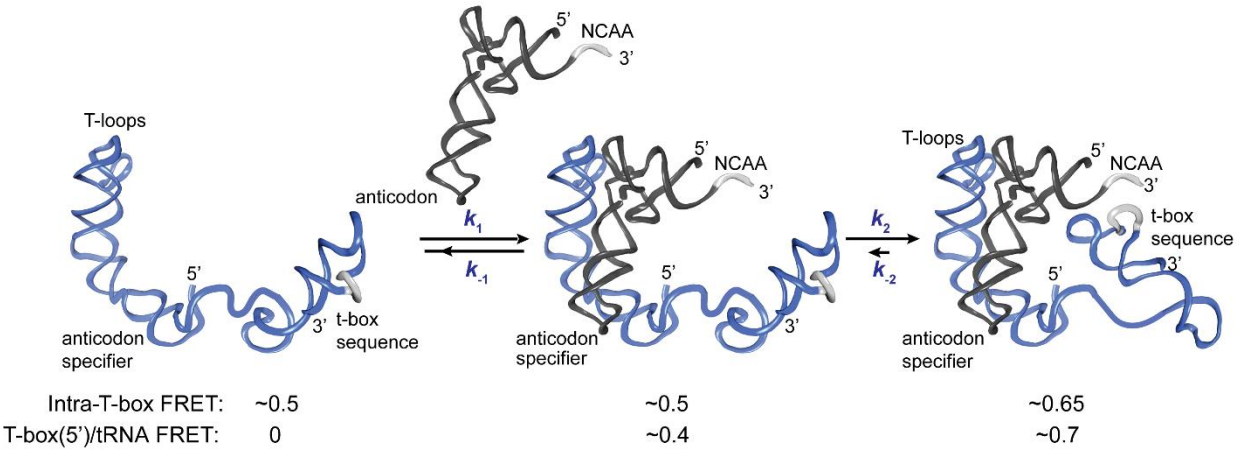


Figure 2.5. Kinetic model for the two-step binding of *glyQS* T-box riboswitch and uncharged tRNA^{Gly}. Details of the model are described in the text. Rate constants are summarized in **Figure 2.6E**.

2.2.5 Structural elements of glyQS Tbox affects the binding dynamics

2.2.5.1 A mutation in the T-loop region affects the first binding step but has minimal effect on the second binding step.

The interdigitated T-loops structure formed by the interactions between the distal bulge and the apical loop at the distal end of Stem I has been shown to be important for tRNA binding[168–170]. Specifically, C56 of T-box stacks on a nucleobase in the D-loop of tRNA, and a point mutation of C56 to U has been shown reduce the tRNA binding affinity by ~40 fold[170]. We introduced the same mutation in the T-box$_{182}$ backbone (T-box$_{C56U}$) (**Figure 2.6A, Figure S2.1.1**). The smFRET trajectories for tRNA^{Gly} binding to T-box$_{C56U}$ are overall similar to the trajectories for WT T-box$_{182}$, with a majority of traces (73±6%) showing stable binding at 0.7 FRET state, and 13±4% of the traces showing transitions back to the 0.4 FRET state (**Figure 2.6C, D**). $T_{0.7}$ was estimated to be at least ~23 s

(limited by the photobleaching of the fluorophore) (**Figure 2.6E**). Post-synchronized time-evolved histogram on the subset of traces that demonstrated real-time binding shows fast transition to the fully bound state (**Figure S2.6.1**). Comparison of $tRNA^{Gly}$ binding to T-box$_{C56U}$ and T-box$_{182}$ suggest that the C56U mutation does not affect the second binding step. To investigate whether the mutation at the T-loop region affects the first binding step, we analyzed the binding and dissociation of $tRNA^{\Delta NCCA}$-Cy5 to T-box$_{C56U}$-Cy3(3'). We found that the k_1 of tRNA binding to T-box$_{C56U}$ was roughly 16-fold slower compared to tRNA binding to T-box$_{182}$, and the dissociation was roughly 2.5-fold faster compared to T-box$_{182}$ (**Figure 2.6E**), leading to a ~40 fold higher dissociation constant for the first binding step. Our results suggest that the T-loop region of the T-box is critical during the first binding step, potentially aiding in anticodon recognition, but does not contribute significantly to the second binding step.

2.2.5.2 A truncation of Stem III has a minor effect on tRNA binding

The functional role of Stem III is unclear. It has been speculated that Stem III might serve as a transcription stalling site to allow co-transcriptional folding and regulation of the T-box riboswitch. In addition, a SAXS data-derived model suggested coaxial stacking of Stem III and the anti-terminator stem, leading to a plausible role of Stem III in stabilizing the anti-terminator conformation in the presence of uncharged $tRNA^{Gly176}$. To investigate the latter hypothesis, we generated a T-box mutant (T-box$_{SIII-\Delta 4bp}$), in which four base pairs in Stem III are deleted to significantly shorten its length (**Figure 2.6A, Figure S2.1.1**).

smFRET studies using T-box$_{SIII-\Delta 4bp}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5 and tRNA^{Gly}-Cy5 revealed insignificant difference in overall kinetics in the first and second step bindings (**Figure 2.6C-E, Figure S2.6.2**). Noticeably, the $T_{0.7}$ was around 50% shorter than that for the T-box$_{182}$ (**Figure 2.6E**), indicating that Stem III may contribute to the stabilization of the fully bound state, potentially through coaxial stacking with the anti-terminator stem, but the effect is minor.

2.2.5.3 A K-turn mutation affects both binding steps

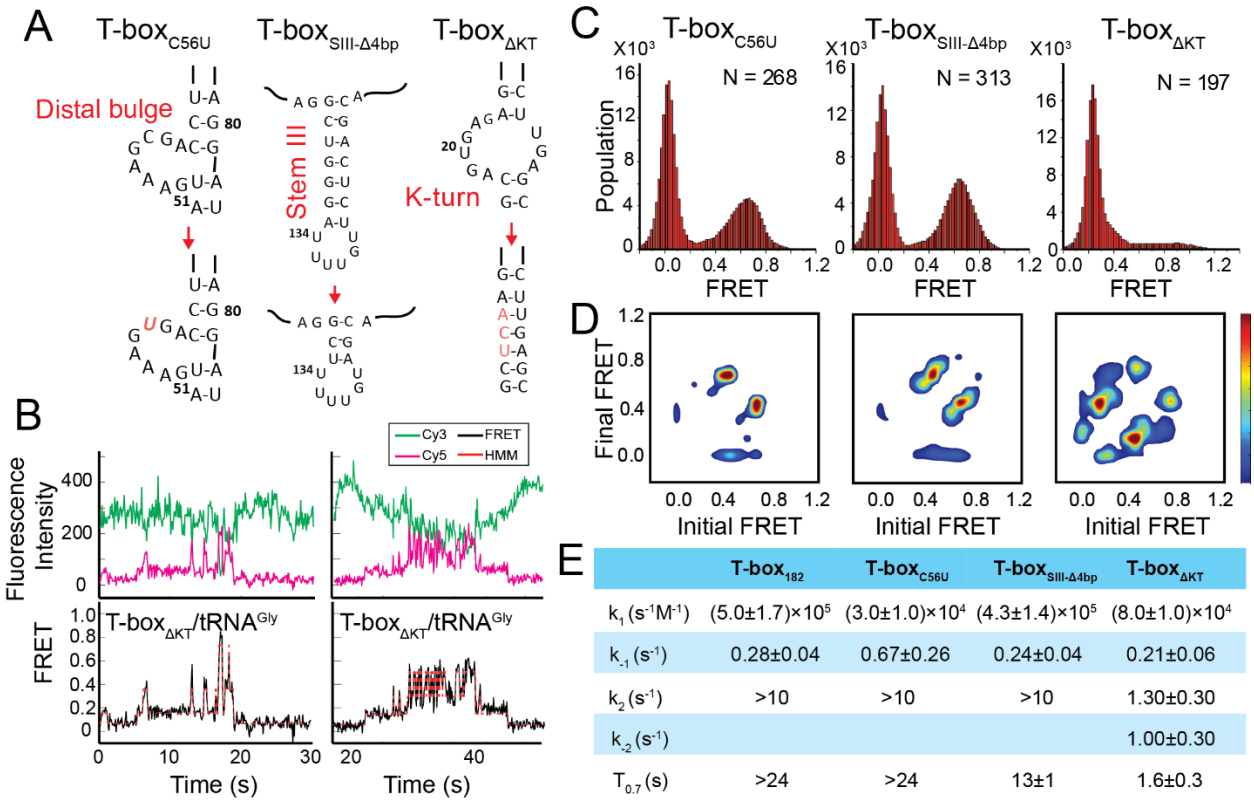
As our smFRET data highlight the role of a region near the K-turn as the hinge of the tRNA binding-dependent conformational change, we investigated the role of the K-turn in regulating tRNA binding kinetics. We disrupted the K-turn (T-box$_{\Delta KT}$) by changing the 6 bulged nucleotides to 3 nucleotides (UCA) to replace the K-turn with a 3 base pair stem (**Figure 2.6A, Figure S2.1.1**). In contrast to binding of tRNA^{Gly} to T-box$_{182}$, binding to T-box$_{\Delta KT}$ binding results in three FRET states centered on 0.2, 0.4, and 0.7. (**Figure 2.6B**). While the exact boundary of each FRET state is difficult to determine accurately from the FRET histogram (**Figure 2.6C**), a transition density plot (TDP) clearly revealed interconversion between the 0.2, 0.4, and 0.7 states (**Figure 2.6D**), with transitions between the 0.2 and 0.4 FRET states, and between the 0.4 and 0.7 FRET states more populated. Binding of tRNA$^{\Delta NCCA}$-Cy5 to T-box$_{\Delta KT}$, on the other hand, leads to the loss of population of the 0.7 state; however, both the 0.2 and 0.4 FRET states and fluctuations between these two states are frequently sampled (**Figure S2.6.3A, B**).

Comparing the $tRNA^{Gly}$ and $tRNA^{\Delta NCCA}$ binding, we speculate that both the 0.2 and 0.4 FRET states observed in the case of the T-box$_{\Delta KT}$ represent the partially bound state in which only the anticodon and elbow are recognized. In contrast to T-box$_{182}$, T-box$_{\Delta KT}$, with the K-turn replaced by an extension of Stem I, could potentially favor a relaxed conformation of Stem I, as observed in the NMR structure of an isolated K-turn and specifier loop domain[180], generating a lower FRET value centered at 0.2. However, the sampling of the 0.4 FRET in the T-box$_{\Delta KT}$ construct suggests that the interactions between the specifier and anticodon of the tRNA may transiently force open the extended base pair region and bend the T-box to adopt a similar conformation to the one observed in T-box$_{182}$. The lifetimes of the 0.4 state transition back to the 0.2 state and the transition forward to the 0.7 state are 0.30±0.03 s and 0.13±0.05 s, respectively, indicating that this forced bent state is energetically unfavorable. However, this 0.4 FRET state is very likely to be required for the NCCA/t-box interaction to occur, as in the presence of $tRNA^{Gly}$ the transitions from the 0.2 FRET to the 0.7 FRET state often pass through the 0.4 FRET state (**Figure 2.6B**). Furthermore, we observed a small region in the TDP corresponding to direct transitions between the 0.2 and 0.7 states. Given the very short lifetime of the 0.4 state, which is close to the time resolution of our experiments, the 0.2 to 0.7 state transition is likely to represent populations whose 0.4 FRET state lifetime is even shorter than the time resolution of the experiment.

The 3-step kinetic scheme for tRNA binding to T-box$_{\Delta KT}$ is presented in **Figure**

S2.6.3C. The association rate constant k_1 ((8.0±1.0)10^4 $s^{-1}\cdot M^{-1}$), estimated from binding of $tRNA^{\Delta NCCA}$ to the T-$box_{\Delta KT}$, is ~ 6-fold smaller than binding to T-box_{182}, suggesting that disruption of the K-turn affects anticodon recognition. Considering both the 0.2 and 0.4 FRET state as the partially bound state in T-$box_{\Delta KT}$, the apparent dissociation rate from the partially bound state k_{-1_app} of $tRNA^{\Delta NCCA}$ to T-$box_{\Delta KT}$ is 0.21±0.06 s^{-1}, similar to that for T-box_{182}, suggesting that disruption of the K-turn does not affect the stability of the partially bound state. The 0.7 FRET state observed for T-$box_{\Delta KT}$ in the presence of $tRNA^{Gly}$ is consistent with the FRET value for the fully bound state in the WT T-box, indicating that the NCCA/t-box interactions in T-$box_{\Delta KT}$ can still be formed. However, $tRNA^{Gly}$ bound to T-$box_{\Delta KT}$ ($T_{0.7}$ = 1.6±0.3 s) is at least 10-fold less stable compared to $tRNA^{Gly}$ bound to T-box_{182} ($T_{0.7}$ >24 s). Furthermore, in contrast to the signal observed for $tRNA^{Gly}$ binding to T-box_{182}, in which fewer than 10% of the FRET traces show transitions back to 0.4 FRET, in T-$box_{\Delta KT}$ the vast majority of the traces show backward transitions to the 0.4 and 0.2 FRET states, contributing largely to the instability of the fully bound state. Based on the transition rates between 0.2, 0.4, and 0.7, we estimated the apparent forward (k_{2_app}) and (k_{-2_app}) reverse transition rates between the partially bound and the fully bound state of the T-$box_{\Delta KT}$ to be 1.3±0.3 s^{-1} and 1.0±0.3 s^{-1}, respectively (**Figure 2.6E, Figure S2.6.3C**). The dramatically reduced k_{2_app} and increased k_{-2_app} in T-$box_{\Delta KT}$ leads to a ~100-fold change in the equilibrium constant in the second binding step compared to T-box_{182}, implying that inflexibility of the K-turn region largely inhibits the conformational change in the T-box required to form the NCCA/t-box interaction, and

strongly destabilizes the fully bound state.



	T-box$_{182}$	T-box$_{C56U}$	T-box$_{SIII-\Delta 4bp}$	T-box$_{\Delta KT}$
k_1 ($s^{-1}M^{-1}$)	(5.0±1.7)×10^5	(3.0±1.0)×10^4	(4.3±1.4)×10^5	(8.0±1.0)×10^4
k_{-1} (s^{-1})	0.28±0.04	0.67±0.26	0.24±0.04	0.21±0.06
k_2 (s^{-1})	>10	>10	>10	1.30±0.30
k_{-2} (s^{-1})				1.00±0.30
$T_{0.7}$ (s)	>24	>24	13±1	1.6±0.3

Figure 2.6. Regulation of the tRNA^{Gly} binding kinetics by structural elements in the *glyQS* T-box riboswitch. **(A)** Schematic diagram of three different mutations introduced to the T-box$_{182}$ backbone (T-box$_{C56U}$, T-box$_{SIII-\square 4bp}$ and T-box$_{\Delta KT}$). **(B)** Representative smFRET traces of T-box$_{\Delta KT}$-Cy3(3') and tRNA^{Gly}-Cy5. **(C)** FRET histograms of the T-box mutants with tRNA^{Gly}-Cy5. "N" denotes the total number of traces in each histogram from 3 independent experiments. **(D)** TDP of the T-box mutants with tRNA^{Gly}-Cy5. TDPs are generated from all smFRET traces from 3 independent experiments. **(E)** Table of kinetic parameters for tRNA^{Gly}-Cy5 binding to different T-box constructs. k_{-1}, k_2, and k_{-2} of T-box$_{\Delta KT}$-Cy3(3') are apparent rate constants estimated to allow comparison as described in **Materials and Methods**. All rate constants are reported as mean ±standard deviation (S.D.) from 3 or 4 independent experiments.

2.3 Conclusion and discussion

T-box riboswitches represent a unique class of riboswitches as they recognize a

macromolecule and require interactions at multiple spatially separated sites on the ligand, unlike other riboswitches that respond to the binding of small ligands. Previous structural studies suggested that tRNA recognition by a T-box riboswitch is a bipartite process [170,173] with Stem I largely responsible for discriminating non-cognate tRNAs while the t-box sequence in the expression platform senses the charged state of the tRNA. We used smFRET to elucidate the binding kinetics of $tRNA^{Gly}$ by the *glyQS* T-box riboswitch. With three FRET pairs between different T-box riboswitch and tRNA ligand constructs, our data collectively reveals a two-step model of uncharged $tRNA^{Gly}$ binding to the *glyQS* T-box riboswitch (**Figure 2.5**). The first binding step involves recognition of the anticodon of the tRNA by the specifier sequence located in Stem I of the T-box riboswitch, leading to a partially bound state. In the second step, the 3' end of the T-box docks into the NCCA end of the tRNA through interactions with the t-box sequence, which leads to a fully bound state. Without the NCCA interaction, the binding of tRNA is unstable, with an average lifetime of ~4 s, whereas with interactions both with the anticodon and the NCCA end, the binding of tRNA is very stable, with an average lifetime > 24 s. The later lifetime measurement is limited by fluorophore photobleaching and is likely to be much longer. In addition, an intra-T-box FRET pair at the 5' and 3' ends of the T-box demonstrates that, while the T-box is largely pre-organized in a folded state before tRNA binding, it still exhibits conformational rearrangement in a tRNA-dependent manner. Specifically, the 3' half of the T-box (including Stem III and the anti-terminator) moves inward relative to the 5' half (Stem I) of the T-box to accommodate the interaction with the NCCA end in the

second binding step. From the FRET pair placed on the tRNA and at the 5' end of Stem I, we further confirm that the hinge is likely to be near the K-turn region.

While our manuscript was in preparation, Suddala *et al.*[181] reported a single-molecule study on tRNA binding to the *glyQS* T-box riboswitch and proposed a similar two-step binding model. Although both studies propose consistent kinetic models, in the study by Suddala *et al.*[181], the FRET pair attached at the variable loop of the tRNA and the 3' or 5' ends of the *glyQS* T-box cannot distinguish between the partially bound state from the fully bound state; therefore the two binding states are distinguished by different dissociation rates of the tRNA from these states, aided by using a Stem I-only mutant that cannot interact with the NCCA end of the tRNA. Neither transitions between the partially bound and the fully bound states, nor the order of events during tRNA binding can be resolved in the study[181]. In contrast, by employing a FRET pair located at the 5' end of the tRNA and the 3' end of the *glyQS* T-box, we observed directly two FRET states corresponding to the recognition of the anticodon (0.4 FRET) and the binding of the NCCA (0.7 FRET), therefore our study allows the discrimination between different states and generates a more complete kinetic framework describing the full trajectory of the tRNA binding process.

Our data reveal that anticodon recognition precedes the NCCA end interactions, and that after anticodon recognition the commitment to further establishment of the NCCA/t-box interaction is high. The T-box/tRNA^{Gly} complex transits rapidly from the partially bound

state to the fully bound state, with a rate constant (k_2) on the order of 10 s^{-1}. Considering that the intracellular concentration of tRNA is on the order of μM in bacteria[182–184], the binding of the tRNA ligand to the T-box riboswitch is rate limited by the first step. Interestingly, our data also reveal that in the fully bound state the NCCA/t-box interaction is not highly-stable or ultra-stable. Brief disruption of the NCCA/t-box interactions can occur, but transition back to the fully bound state is rapid, ~10 fold faster than dissociation of the tRNA from the partially bound state. Therefore, tRNA can remain bound during the breaking and reforming of the NCCA/t-box interaction. However, such transient breaking of NCCA/t-box interaction was only observed in ~10% of the total population, suggesting that the reverse transition rate constant (k_{-2}) is very small, and the second binding step in the WT T-box with uncharged $tRNA^{Gly}$ is close to be irreversible. The relatively slow dissociation rate (k_{-1}) from the partially bound state and the very rapid and almost irreversible transition from the partially bound state to the fully bound state is kinetically beneficial during co-transcriptional folding of the T-box and sensing of the tRNA ligand. On one hand, slow dissociation from the partially bound state can ensure the tRNA ligand stay bound until the completion of the anti-terminator sequence. On the other hand, very rapid and almost irreversible transition into the fully bound state helps secure the interaction between the NCCA end of the tRNA and the t-box sequence and trap the T-box in the anti-terminator conformation before completion of transcription of the terminator sequence, which is immediate downstream.

Importantly, our results hint at the critical role of the K-turn region in promoting the fast coordination between anticodon sensing and locking of the NCCA end. Crystal structures of the Stem I/tRNA complex[170,173] show that Stem I flexes around the K-turn region and that this flexing seems to be important to establish the interactions between the anticodon and specifier sequence. Using a mutant where the K-turn region is removed, our data show that in the absence of the K-turn motif both the first and second binding steps are affected, but with a much more dramatic effect on the second step. In this specific K-turn mutant, where the K-turn is replaced by an extended Stem I helix, the conformational change in the T-box that brings closer the 3' and 5' portions becomes highly energetically unfavorable, leading to a ~10-fold reduction in k_2, and more than 10-fold destabilization of the fully bound state. Our kinetic measurement of the K-turn mutant explains the *in vivo* loss-of-function K-turn mutants[185], and emphasizes the importance of the flexing around the K-turn region and the associated conformational changes in the T-box itself in the overall recognition and binding process.

Our results with various T-box mutants demonstrate that the T-box structural elements involved in tRNA recognition can drive the two-step binding process in a modular fashion. For example, in the first binding step, interactions with Stem I involve both anticodon recognition and the contacts between the interdigitated T-loops in Stem I and the elbow region (D- and T-loops) of the tRNA[170,173]. When introducing a point mutation that impairs the T-loops/elbow region interaction[170], we observed a dramatic

decrease in the association rate constant and a moderate increase in the dissociation rate constant, leading to an overall ~40 fold reduction on the binding affinity for the first step, however the second step is unaffected. This observation suggests that establishment of the T-loops/elbow interactions is an important part of the Stem I/tRNA recognition process in the first binding step, but does not play any role in NCCA recognition in the second binding step. Similarly, truncation of Stem III has a minor effect on the stability of the fully bound state, but with no influence on the first binding step. Finally, the K-turn region (discussed above), which links the 5' and 3' portions of the T-box, plays a key role in coordinating the two binding steps by providing structural flexibility. It is worth mentioning that while Stem III contributes to the stabilization of the anti-terminator conformation *in vitro*, deletion of it does not appear to significantly impair the tRNA binding process *in vitro*. Potentially its major function is to create a pause site to coordinate with the co-transcriptional folding of the T-box.[186,187]

In conclusion, our study provides a comprehensive kinetic framework for describing tRNA recognition by the T-box riboswitch. The two-step binding process is driven by the specific structural elements of the T-box, and is kinetically beneficial for efficient, co-transcriptional recognition of the cognate tRNA ligand. Specific T-box structural elements drive the two-step binding process in a modular fashion, providing a guideline for synthetic biology design of RNA regulatory element. Finally, the *glyQS* T-box riboswitch represents one of the simplest members of this class of ribowitches. Other T-box

riboswitches are larger and have additional structural elements and can even appear in tandem arrangements[188]. While the two-step binding kinetics may be common to all T-box riboswitches, it is likely that the process in other T-box riboswitches shows differences modulated by the additional structural elements.

2.4 Materials and Methods

RNA purification and mutagenesis

RNA transcription was performed in vitro using His_6-tagged T7 RNA polymerase using standard protocols[189]. Cloning, design of a bicistronic DNA template encoding the *B. subtilis glyQS* T-box riboswitch and its cognate $tRNA^{Gly}$, and conditions for in vitro transcription were described before[176]. For the experiments, all the RNAs from the crude transcription reaction were purified on a 7.5% denaturing (8 M Urea) polyacrylamide gel. The RNAs of interest were located on the gel by UV shadowing, the bands were cut out, and the RNAs were eluted into 50 mM sodium acetate (pH 7.0) buffer containing 200 mM potassium chloride by overnight rocking at 4°C. The eluted RNAs were precipitated by adding 3 volumes of cold 100% ethanol and stored overnight at −20°C. The precipitated RNAs were pelleted by centrifugation for 15 min at 20,000 g. The RNA pellets were washed three times in cold 80% ethanol, dried in a Speedvac and re-suspended in water. The concentration was estimated by its absorbance at 260 nm and was kept frozen at −20°C for long term storage. All mutant constructs were made using a commercial site

directed mutagenesis protocol[190] (Quikchange, Stratagene) and the RNAs were produced and purified by using the same protocol described above. The sequence of all mutant T-box and tRNA constructs were confirmed by sequencing and validated by MFOLD[191] to ensure that the secondary structural elements were not affected by the mutations.

Fluorescent labelling of the RNA

For smFRET experiments, end labelling of RNA molecules was performed by modifying standard labelling protocols[123]. For 3' end labelling, 50 µg of RNA in 50 µL of reaction volume was incubated with 0.1 M Na-periodate in 0.1 M Na-acetate buffer at pH 5.2 for 90 min in the dark. The reaction was quenched by adding 5 µL of 2.5 M KCl and incubating on ice for 10 min. The resultant insoluble KIO_4 was removed by centrifugation at 20,000 g for 30 min and the supernatant was passed through a P6 column (Bio-Rad) to exchange the buffer to 0.1 M HEPES (pH 7.0), 40% DMSO. The RNA was incubated with Cy3 hydrazide (Lumiprobe) dye for 45 min with a final RNA/dye ratio of ~1:200. The RNA was then ethanol precipitated as described above. The precipitated RNA was pelleted by centrifugation for 15 min at 20,000 g, dried in a Speedvac and re-suspended in water. The final RNA solution was passed through a P6 column to remove any residual free dye.

5' end labelling of RNA was performed by N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC) – N-hydroxysuccinimide (NHS) coupling through activation of the 5' monophosphate of the RNA by EDC and imidazole[123]. To improve the

overall labelling efficiency of this method, a modified approach was used in which the 5' triphosphate of 100 µg of RNA was converted to 5' monophosphate in 100 µL reaction volume by incubating it with 100 units of RNA 5' Pyrophosphohydrolase (NEB) at 37°C for 1 hr. The enzyme was removed by phenol chloroform extraction and the supernatant was passed through a P6 column to exchange the buffer to 10 mM HEPES (pH 7.0), 150 mM NaCl, 10 mM EDTA. This was followed by addition of 12.5 mg of EDC to the RNA solution along with 50 µL of ethylene diamine and 200 µL of 0.1 M imidazole buffer (pH 6.0). The reaction was incubated for 3 hr at 37°C and the RNA was then ethanol precipitated as described above. The resultant RNA pellet was re-suspended in 0.1 M sodium carbonate buffer (pH 8.7) and residual EDC was removed by passing the solution through a P6 column. The resultant RNA solution in 0.1 M sodium carbonate buffer (pH 8.7) was incubated with Cy5 NHS (Lumiprobe) dye for 45 min with a final RNA/dye ratio of ~1:200. The RNA was ethanol precipitated as described above and re-suspended in water. The final RNA solution was passed through a P6 column to remove any residual free dye. The overall labelling efficiency of RNA constructs used in this study varied from 75% to 95%.

Isothermal titration calorimetry

The thermodynamic parameters associated with binding of $tRNA^{Gly}$ to *glyQS* T-box were determined at 25°C by using an ITC-200 Micro-Calorimeter (MicroCal). Prior to the experiment, the two interacting RNAs (in 50 mM HEPES pH 7.0, 100 mM KCl) were

refolded separately by first heating for 3 min at 90° C followed by incubation on ice for 2 min. At this point, $MgCl_2$ was added to attain a final concentration of either 1 mM or 10 mM. The RNA solution was then heated to 50° C for 10 min and 37° C for 30 min followed by cooling to 25°C. The refolded RNA was then concentrated using a 10 kDa cutoff Amicon filter and washed three times with the final ITC buffer (50 mM HEPES pH 7.0, 100 mM KCl and 10 mM $MgCl_2$ or 1 mM $MgCl_2$). For testing the binding of T-box with unlabeled and 5' Cy5 labelled tRNA in 10 mM $MgCl_2$ buffer, the sample cell was filled with 6.6 μM and 11 μM T-box respectively and the corresponding concentration of $tRNA^{Gly}$ in the syringe was 111 μM and 152 μM. For the ITC experiment in 1 mM $MgCl_2$ buffer, the cell was filled with T-box at a concentration of 6.7 μM and the syringe concentration of unlabeled $tRNA^{Gly}$ was 72 μM. For each ITC experiment, the titration was carried out by stepwise (2 μL) injection of $tRNA^{Gly}$ from a stirred syringe (1000 rev/min) into the sample cell. Successive injections were spaced by 150 s and values for the change in enthalpy (ΔH_b), association constant (K_b), change in enthalpy (ΔH_b) and stoichiometry (n) were determined by nonlinear least-squares fitting of the data using Origin 5.0 software (OriginLab).

Electrophoretic mobility shift assay

The T-box and tRNA were denatured and refolded using the same protocol as described for the ITC and smFRET experiments. For binding of tRNA, 5 μL of 4 μM of folded T-box and 5 μL of 2 μM folded tRNA samples were mixed together for 30 min in

final buffer containing 50 mM HEPES pH 7.0, 100 mM KCl and 15 mM $MgCl_2$. The folded T-box or T-box +tRNA mixture were loaded on to 6% native polyacrylamide gel containing 15 mM $MgCl_2$ and the gel was run at room temperature for 2–2.5 hr. RNAs were stained by SYBR™ green RNA staining dye (Invitrogen). The gel was imaged using ChemiDoc (Bio-rad) in SYBR green channel (for unlabeled T-box constructs), Cy5 channel (for tRNA-Cy5), or Cy3 channel (for labeled T-box constructs). Images were processed and analyzed by ImageJ[192].

Fluorophore conjugation of DNA oligos

DNA oligos that hybridize to 5' extension and 3' extension of the T-box construct were purchased from Integrated DNA Technologies with an amine modification at the 5' end and 3' end respectively. 13.5 µL of 100 µM DNA oligo was mixed with 1.5 µL of 1 M $NaHCO_3$ (pH 8.6). 25 µg of NHS conjugated fluorophore (Cy3 or Cy5) was dissolved with 0.5 µL DMSO and mixed with the DNA oligo solution. The mixture was incubated at 37°C overnight. 1.67 µL of 3 M NaOAc and 50 µL of pure ethanol was added to the mixture to precipitate the conjugated DNA oligo overnight at −20°C. The precipitated DNA oligo was pelleted by centrifugation for 30 min at 21000 g and re-suspended with 40 µL water. The DNA solution was passed through a P6 column to remove any residue free dye and salt. The overall fluorophore labeling efficiency is ~60%.

smFRET measurements

Slides containing microfluidic channels were prepared as previously described[193].

Slides and coverslips were coated with a mixture of poly-ethylene glycol (PEG, Mw = 500,000) and PEG-biotin (Mw = 500,000) according to previously published protocol[194]. The T-box and tRNA were denatured and refolded using the same protocol described for the ITC experiments with the final buffer containing 50 mM HEPES pH 7.0, 100 mM KCl, and 15 mM $MgCl_2$. T-box RNA was hybridized to the biotinylated DNA oligo at the extension during refolding, and immobilized via biotin-streptavidin interactions to the surface. tRNA was diluted in imaging buffer (50 mM HEPES pH 7.0, 100 mM KCl, 15 mM $MgCl_2$, 5 mM protocatechuic acid (PCA) (Sigma-Aldrich), 160 nM protocatechuate-3,4-dioxygenase (PCD) (Sigma-Aldrich), and 2 mM Trolox (Sigma-Aldrich)) and flowed into the microfluidic channels. smFRET measurements were performed with an objective based total internal reflection fluorescence (TIRF) microscope based on a Nikon Ti-E with 100X NA 1.49 CFI HP TIRF objective (Nikon). A 561 nm laser (Coherent Obis at a power density of 4.07×10^5 W/cm^2) was used for the FRET measurement. A 647 nm laser (Cobolt MLD at a power density of 5.88×10^5 W/cm^2) was used for direct excitation of Cy5 to check the presence of the acceptor. Emissions from both donor and acceptor were passed through an emission splitter (OptoSplit III, Cairn), and collected at different locations on an EMCCD (iXon Ultra 888, Andor). 1500 frames of time-lapse images were taken with 100 ms exposure time. Each independent measurement in our study should be considered as a technical replicate. The biological samples (i.e. the in vitro transcribed tRNAs and T-boxes) were generated once. In each measurement, T-box and tRNA were folded, smFRET images were recorded, and analysis were performed independently.

Lifetime analysis

Individual spots were picked from maximum intensity projection of Cy5 emission channel (**Figure S2.2.1**) using NIS Elements software, that is only the pixels with the highest intensity value of the same XY coordinates in time-lapse images are displayed in the maximum intensity projection image. Fluorescent intensity trajectories of these spots were generated from Cy3 and Cy5 channels of time-lapse images, and corrected for baseline and bleed-through in MATLAB as previously described[193]. FRET traces were generated by calculating I_{Cy5} / (I_{Cy5}+I_{Cy3}) at each time point from the intensity trajectories. smFRET traces were idealized by fitting with a Hidden Markov Model using vbFRET[195]. Dwell time of each FRET state before transition to another FRET state of individual traces was extracted from the idealized traces and the dwell time histograms were fit with exponential decay by Origin 7.0 (OriginLab) (**Figure S2.2.2**). In most of the cases, dwell time histograms fit well with a single exponential decay ($A \cdot \exp(-t/t_0) + y_0$). In a few cases, in which mixed populations with different lifetimes were expected, dwell time histograms were fit with a double exponential decay ($A_1 \cdot \exp(-t/t_1) + A_2 \cdot \exp(-t/t_2) + y_0$) and the population-weighted average lifetime was calculated by ($A_1 \cdot t_1 + A_2 \cdot t_2$) / ($A_1 + A_2$). The latter cases are explicitly mentioned in the text.

Determination of kinetic parameters

(I) k_{-1}

For all T-box constructs, k_{-1} was estimated from the dwell time of the partially bound state (0.4 FRET state of T-box$_{182}$, T-box$_{C56U}$ and T-box$_{SIII\text{-}\Delta 4bp}$) of tRNA$^{\Delta NCCA}$-Cy5 (**Figure**

S2.2.2). The dwell time of the partially bound state was fit with a single exponential decay. The lifetime of 0.4 FRET state ($\tau_{0.4\text{-}0}$) was extracted and k_{-1} was calculated using the following equation:

$$\boldsymbol{k_{-1} = \frac{1}{\tau_{0.4-0}}}$$

(II) k_1

We noticed that the lifetime of the unbound state ($\tau_{0\text{-}0.4}$) of tRNA$^{\Delta NCCA}$-Cy5 with Tbox$_{182}$-Cy3(3') was 31.3 ± 5.3 s, while the average smFRET trace length (τ_{trace}) was 70 ± 7 s under 4.07 × 10^5 W/cm^2 561 nm laser (**Figure 2.2-S2**), limited by Cy3 photobleaching. Therefore, there was a possibility of underestimating the lifetime of the unbound state, or overestimating k_1. We hence characterized the photobleaching effect on unbound state lifetime analysis using reduced 561 nm laser power (1.96 × 10^5 W/cm^2). $\tau_{0\text{-}0.4}$ was calculated to be 46 ± 4 s and τ_{trace} was 103.0 ± 0.6 s. To correct for the photobleaching effect, we estimated k_1, using the following equation[196,197]:

$$k_1 = \left(\frac{1}{\tau_{0-0.4}} - \frac{1}{\tau_{trace}}\right) \times \frac{1}{[tRNA]}$$

where [tRNA] represents the concentration of tRNA$^{\Delta NCCA}$-Cy5. k_1 values were consistent using both high laser power ((5.8 ± 0.2) x 10^5 $M^{-1}s^{-1}$) and low laser power ((4.1 ± 0.6) x 10^5 $M^{-1}s^{-1}$). In **Figure 2.6E**, we reported the mean ±S.D. of k_1 obtained with both high and low laser powers.

We also measured τ_{off} in the flow experiment and it was determined by the dwell time between the injection time and detection of the first binding event (**Figure S2.2.4C**). k_1 calculated for tRNA^{Gly} in the flow experiment was consistent with the value

for $tRNA^{\Delta NCCA}$ in the steady-state experiment. We therefore used k_1 of $tRNA^{\Delta NCCA}$-Cy5 binding to the T-box to approximate the rate constant for the first binding step in all T-box constructs.

For some T-box mutants (specifically, $T\text{-box}_{C56U}$-Cy3 and $T\text{-box}_{\Delta KT}$-Cy3), binding of tRNA-Cy5 became less efficient compared to the $T\text{-box}_{182}$-Cy3(3'), already a hint of a deficient first binding step. In this case, estimation of k_1 from the selected FRET traces will highly bias for the molecules that bind relatively fast within the photobleaching time of Cy3, therefore significantly overestimating k_1. We therefore applied a different approach to estimate k_1 for the T-box mutants. We counted trace percentage of the T-box mutants (number of FRET traces divided by the number of Cy3 spots) and normalized it to the trace percentage of $T\text{-box}_{182}$ with $tRNA^{\Delta NCCA}$-Cy5 to get the normalized trace percentage (f) (**Figure S2.6.4**).

We performed Gillespie simulations on the observed signal as a function of k_1 and k_{-1} during the first binding step[198–200]. The simulated signal mimicked the experiment and imaging processing. As our FRET traces were selected from the maximum intensity projection of the acceptor signal (**Figure S2.2.1**) regardless of the lifetime of the bound state, as long as there was a binding event during the imaging time window (70 s considering the photobleaching time of the donor dye), it would generate a signal (as a 'footprint') that showed in the output image of maximum projection (**Figure S2.6.5A**). For each parameter set, 2000 binding trajectories were simulated, and the fraction of the molecules that generated signal in the maximum intensity projection (referred to as 'f')

within the 70 s time window was calculated. The simulation shows that *f* is insensitive to k_{-1}, at least within the k_{-1} we observed for the various T-box constructs (**Figure S2.6.5B**), and that the dependence of *f* on k_1 is best described as

$$f = f_0 - ae^{bk_1}$$

where f_0 is the binding fraction at saturation, and *a* and *b* are constants (**Figure S2.6.5B**). By fitting $\ln(f_0-f)$ *vs.* k_1 with a linear function (**Figure S2.6.5C**) we estimated k_1 as

$$k_1 = \frac{0.45 - \ln(1.557 - f)}{72.09}$$

The standard deviations reported for k_1 were calculated by an error propagation function considering the standard deviation of *f*.

(III) k_2 and k_{-2}

Even though we observed that smFRET traces of $tRNA^{Gly}$ with $T\text{-}box_{182}$, $T\text{-}box_{C56U}$ and $T\text{-}box_{SIII\text{-}\Delta4bp}$ exhibit real-time binding, the majority of the traces show single-step binding from unbound to fully bound state while only a few traces sample the partially unbound state within 0.1 s (**Figure 2.2F, Figure S2.2.4B**, **Figure S2.6.1A and S2.6.2A**). For this reason, our calculation of k_2 is limited by the imaging time resolution and we estimate the transition from 0.4 to 0.7 is rapid with a k_2 larger than 10 s^{-1}.

We can also detect transition from 0.7 FRET to 0.4 FRET, and calculate the lifetime of 0.7 FRET to 0.4 FRET ($\tau_{0.7\text{-}0.4}$) (**Figure S2.2.2**). However, the fraction of traces showing this transition is small. Since most of the traces demonstrate long time 0.7 FRET, we reason that k_{-2} should be very small and the second binding step for these three

constructs is close to irreversible.

(IV) Kinetic model for T-box$_{\Delta KT}$

The kinetic model for tRNA binding to the K-turn mutant T-box$_{\Delta KT}$ is illustrated in **Figure S2.6.3C**, with kinetic parameters marked. Specifically, the partially bound state contains a relaxed conformation and a bent conformation at the original K-turn region. The interconversion rates between 0.2, 0.4 and 0.7 FRET states were determined by the lifetime of each FRET state before transitioning to the other FRET state. To compare better the kinetic parameters with other T-box constructs, we calculated the apparent k_2 considering both pathways (0.2→0.7 and 0.2→0.4→0.7) (**Figure 2.6.3C**) as below:

$$k_{2\_app} = \left(p_1 \frac{k_1' k_2'}{(k_1' + k_{-1}')} + p_2 k_2 \right)$$

Where p_1 and p_2 are the probabilities of the 0.2→0.4→0.7 and 0.2→0.7 pathways respectively. And we calculated the apparent k_{-2} considering both pathways (0.7→0.2 and 0.7→0.4) (**Figure 2.6.3C**) as shown below:

$$k_{-2\_app} = (p_1 k_{-2}' + p_2 k_{-2})$$

Where p_1 and p_2 are the probabilities of the 0.7→0.4 and 0.7→0.2 pathways, respectively. To compare with the k_{-1} of other T-box constructs, the apparent k_{-1_app} of T-box$_{\Delta KT}$ was estimated by considering both 0.4 and 0.2 FRET states.

2.5 Supplemental Information

Supplementary Tables

Table 2.1 E.coli expression plasmid used in this chapter

Name	Description	Benchling link
pUC19-GlyQS-TBox/tRNA	GlyQS Tbox and tRNA transcription template	https://benchling.com/emheideman/f/lib_yb0ZeUDy-t-box/seq_xXu1Ezed-puc19-glyqs-tboxtrna-/edit

Table 2.2 RNA template and DNA oligo sequences used in this chapter

Name	Description	sequence
SmFRET2	GlyQS Tbox with antiterminator and 5’ tail	GGGGGAAAUUAGGGGGUAUGUUGCAGUGAGAGAAAGAAG UACUUGCGUUUACCUCAUGAAAGCGACCUUAGGGCGGUG UAAGCUAAGGAUGAGCACGCAACGAAAGGCAUUCUUGAG CAAUUUUAAAAAAGAGGCUGGGAUUUUGUUCUCAGCAACU AGGGUGGAACCGCGGGAGAACUCUCGUCCCUA
SmFRET11	5’ tail + StemI + StemIII	GGGGGAAAUUAGGGGGUAUGUUGCAGUGAGAGAAAGAAG UACUUGCGUUUACCUCAUGAAAGCGACCUUAGGGCGGUG UAAGCUAAGGAUGAGCACGCAACGAAAGGCAUUCUUGAG CAAUUUUAAAAAAGAGGCUGGGAUUUUGUUCUCAGCAAC
SmFRET12	5’ tail + Stem I	GGGGGAAAUUAGGGGGUAUGUUGCAGUGAGAGAAAGAAG UACUUGCGUUUACCUCAUGAAAGCGACCUUAGGGCGGUG UAAGCUAAGGAUGAGCACGCAACGAAAGGCAUUCUUGAG CAAUUUUAAAAAAGAG
SmFRET13	GlyQS Tbox with C56U mutation and 5’ tail	GGGGGAAAUUAGGGGGUAUGUUGCAGUGAGAGAAAGAAG UACUUGCGUUUACCUCAUGAA***AGUGA***CCUUAGGGCGGUG UAAGCUAAGGAUGAGCACGCAACGAAAGGCAUUCUUGAG CAAUUUUAAAAAAGAGGCUGGGAUUUUGUUCUCAGCAACU AGGGUGGAACCGCGGGAGAACUCUCGUCCCUA
SmFRET15	GlyQS Tbox with StemIII truncation and 5’ tail	GGGGGAAAUUAGGGGGUAUGUUGCAGUGAGAGAAAGAAG UACUUGCGUUUACCUCAUGAAAGCGACCUUAGGGCGGUG UAAGCUAAGGAUGAGCACGCAACGAAAGGCAUUCUUGAG CAAUUUUAAAAAAGAGGCUUUUUGUAGCAACUAGGGUGGA ACCGCGGGAGAACUCUCGUCCCUA

Table 2.2, continued

SmFRET16	GlyQS Tbox with K-turn mutation and 5' tail	GGGGGAAAUUAGGGGGUAUGUUGC***UCA***AGAAAGAAGUAC UUGCGUUUACCUCAUGAAAGCGACCUUAGGGCGGUGUAA GCUAAGGAUGAGCACGCAACGAAAGGCAUUCUUGAGCAA UUUUAAAAAAGAGGCUGGGAUUUUGUUCUCAGCAACUAG GGUGGAACCGCGGGAGAACUCUCGUCCCUA
$tRNA^{Gly}$		GGGGAAGUAGUUCAGUGGUAGAACACCACCUUGCCAAGG UGGGGGUCGCGGGUUCGAAUCCCGUCUUCCCCUCCA
$tRNA^{\Delta NCCA}$		GGGGAAGUAGUUCAGUGGUAGAACACCACCUUGCCAAGG UGGGGGUCGCGGGUUCGAAUCCCGUCUUCCCCU
5' anchoring oligo	3' biotin and 5' amine modification	CCCCCTAATTTCCCCC
3' anchoring oligo	5' amine modification	GCAGCCGCAAACATT

Supplementary Figures

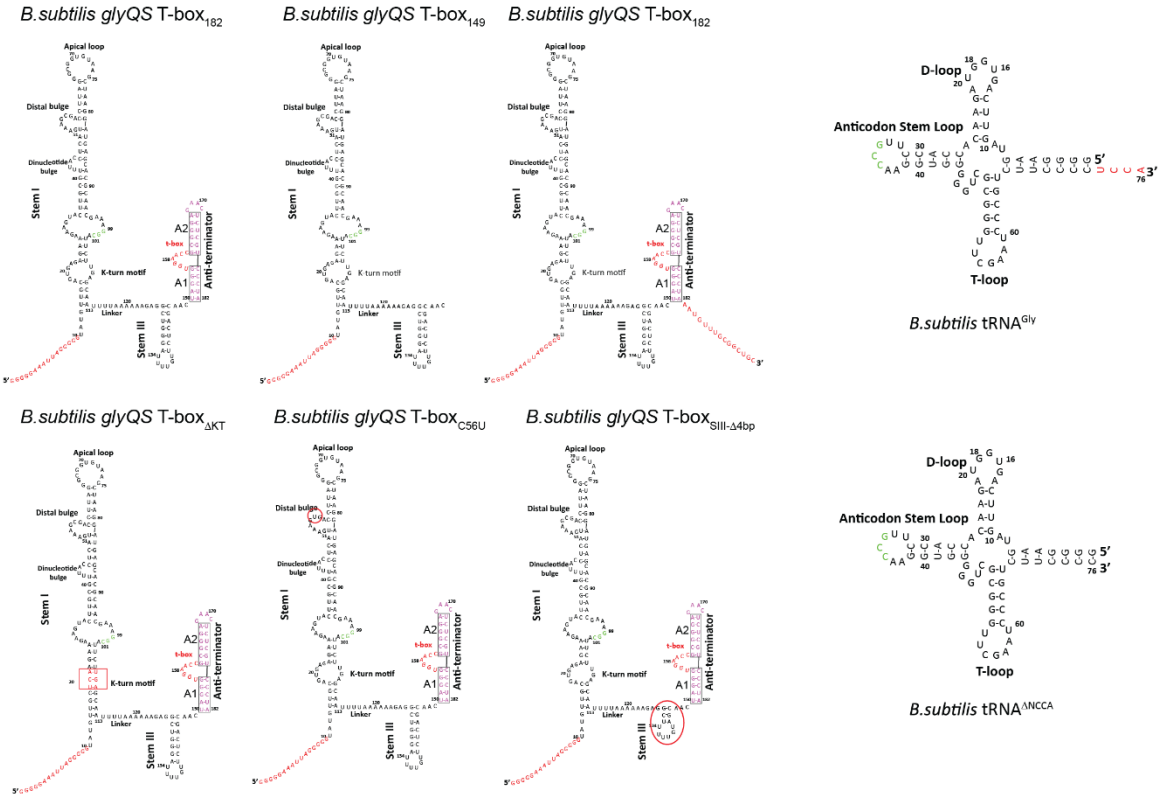


Figure S2.1.1. Secondary structure diagram of the constructs used for the smFRET studies. The diagrams show all the constructs used for the experiments. The 5' extensions added to attach the T-box constructs to the surface are shown in red. The $Tbox_{182}$ construct was made with only a 5'extension and also with extensions at both the 5' and 3' ends (red)..

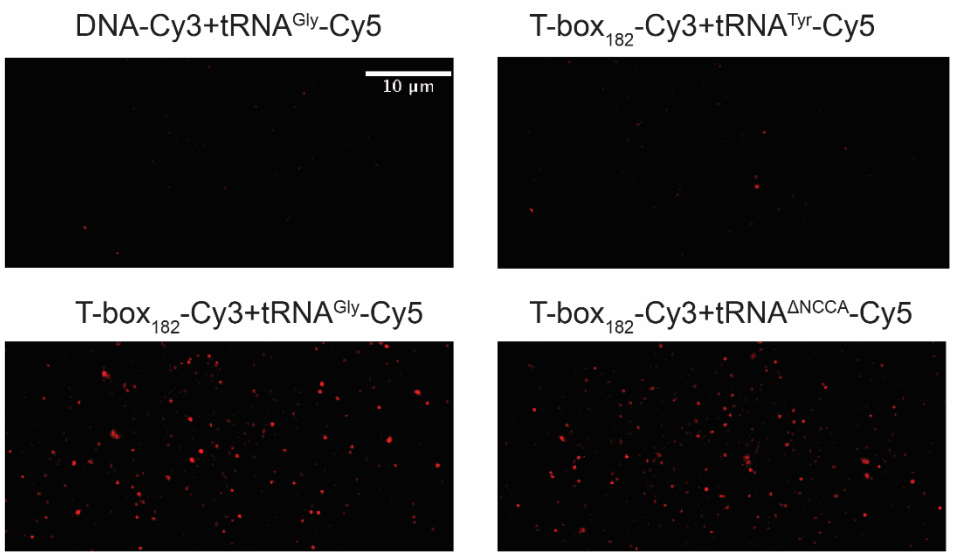


Figure S2.2.1. Representative images of smFRET data for T-box-Cy3(3') and tRNA-Cy5 binding. Images were created by maximum intensity projection of Cy5 emission of the time-lapse images and therefore report on the binding events of tRNA-Cy5 to the surface tethered T-box. A DNA oligo labeled with Cy3 is used as a negative control for non-specific binding or signal. Loading tRNA^{Tyr}-Cy5 to pre-immobilized T-box-Cy3(3') only generates background level of Cy5 signals in the maximum projection similar to the negative control, and these nonspecific Cy5 signals do not generate any smFRET traces.

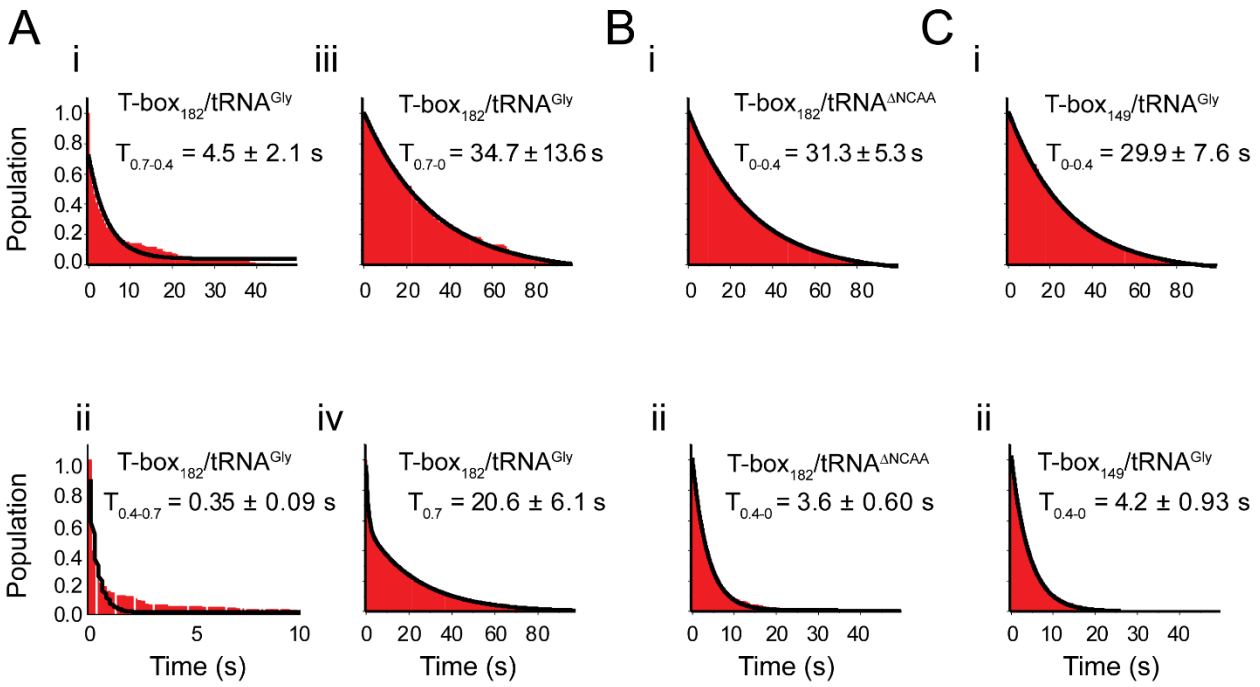


Figure S2.2.2. Representative lifetime analyses of *glyQS* T-box and tRNA interaction. (A) Dwell time of i) 0.7 FRET state to 0.4 FRET state, ii)0.4 FRET state to 0.7 FRET state, iii) 0.7 FRET state to 0 FRET state, and iv) 0.7 FRET state to other FRET states of T-box$_{182}$-Cy3(3') with tRNA^{Gly}-Cy5. Histograms of i), ii), and iii) are fit with a single-exponential decay function (black curve) and iv) is fit with a double exponential decay function to generate the population-weighted average lifetime of the 0.7 FRET state ($T_{0.7}$), as molecules can transit from 0.7 FRET state to both 0.4 FRET state occasionally, and 0 FRET state upon fluorophore photobleaching. **(B)** Dwell time of i) 0 FRET to 0.4 FRET state and ii) 0.4 FRET state to 0 FRET of T-box$_{182}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. Histograms are fit with a single-exponential decay function (black curve). **(C)** Dwell time of i) 0 FRET state to 0.4 FRET state and ii) 0.4 FRET state to 0 FRET state of T-box$_{149}$-Cy3(3') with tRNA^{Gly}-Cy5. Histograms are fit with a single-exponential decay function (black curve). Mean ± standard deviation (S.D.) are calculated from 3 independent measurements.

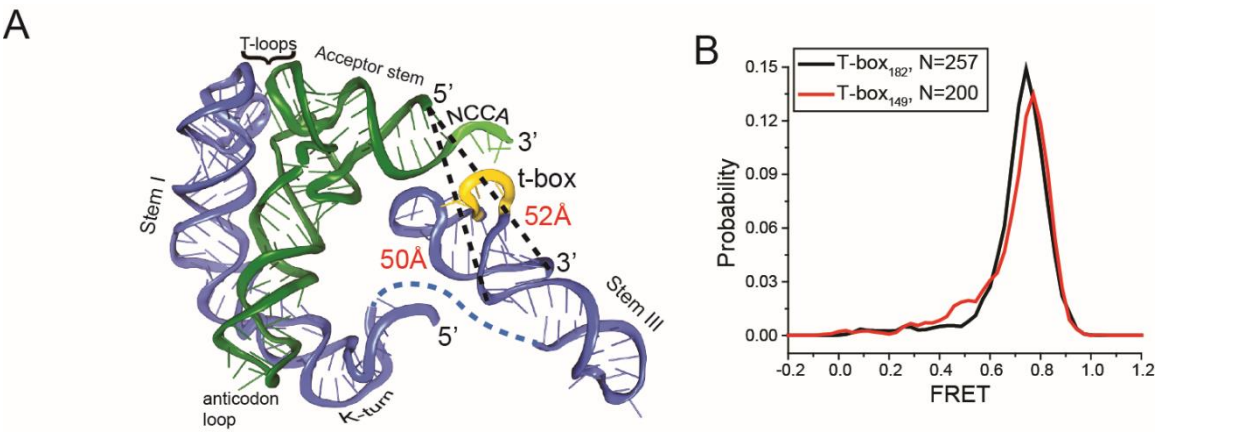


Figure S2.2.3. Intra-T-box FRET of T-box$_{182}$ and T-box$_{149}$. (A) Ribbon diagram of a complex of the *B. subtilis glyQS* T-box riboswitch and uncharged tRNA^{Gly}. Distances from the 5' end of the tRNA to the 3' end of the anti-terminator (T-box$_{182}$) and to the 3' end of Stem III (T-box$_{149}$) are marked. **(B)** One dimensional FRET histograms of intra-T-box pair of T-box$_{182}$ (black) and T-box$_{149}$ (red) with Cy3 attached directly to the 3' end of the T-box and Cy5 attached to the oligo hybridized to the 5' extension of the T-box. Only the first 50 data points of the FRET trajectories are used to plot the histogram to eliminate the zero FRET resulted from Cy5 photobleaching. "N" denotes the total number of traces in each histogram from 2 independent experiments.

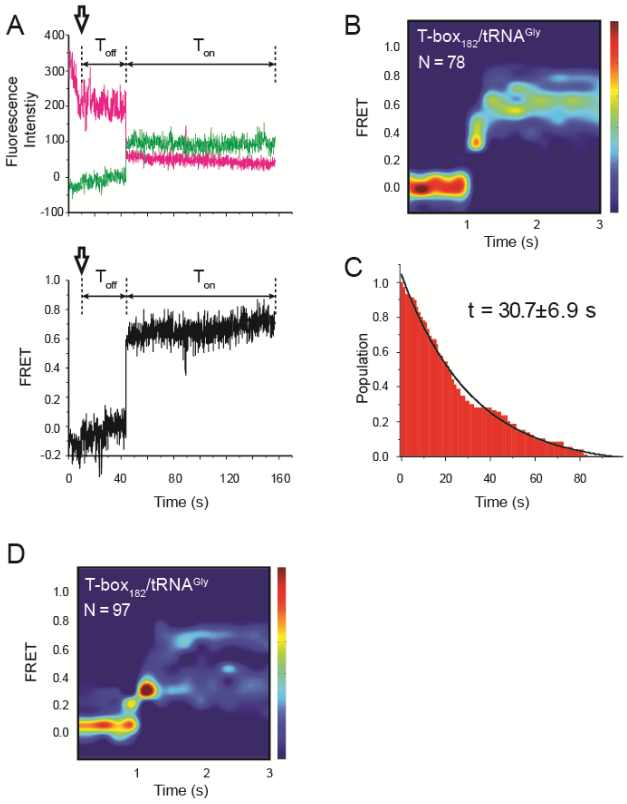


Figure S2.2.4. Real-time flow experiment of the T-box$_{182}$-Cy3(3') and tRNA^{Gly}-Cy5. (A) smFRET trajectories of T-box$_{182}$-Cy3(3') and tRNA^{Gly}-Cy5. Black arrows represent the time point when tRNA^{Gly}-Cy5 is flowed through the imaging area. Dwell time of unbound state (T_{off}) was calculated as the time between starting point and first FRET transition. **(B)** Surface contour plot of time-evolved FRET histogram of the flow experiment. Figure is plotted in the same way as **Figure 2F**. "N" denotes the total number of traces in each histogram from 2 independent experiments. **(C)** Histogram of T_{off} and its single-exponential decay fitting Mean ± S.D. are calculated from 2 independent measurements. **(D)** Surface contour plot of time-evolved FRET histogram of all traces showing real-time binding events, including the traces that are unable to reach the 0.7 FRET state. The same data sets are used in **Figure 2F**. "N" denotes the total number of traces in each histogram from 3 independent experiments, which are a subset of traces showing real-time binding event in the steady-state measurements. Total numbers of traces in each data sets are indicated in **Figure 2B**.

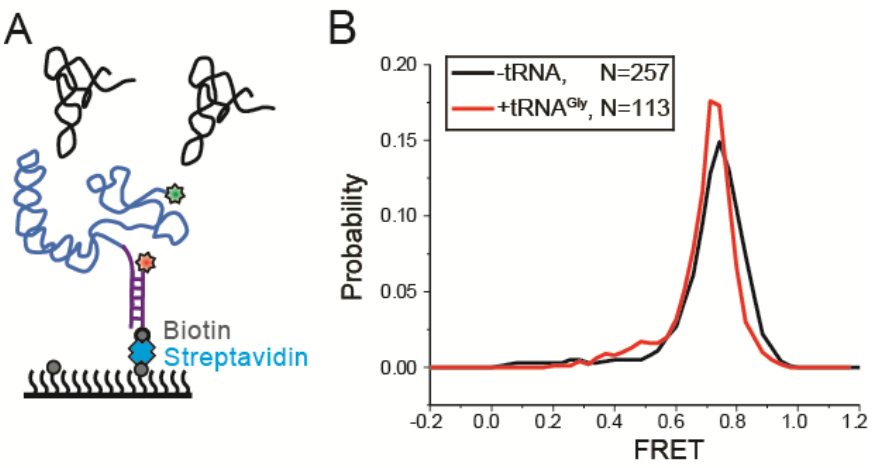


Figure S2.3.1. Intra-T-box FRET of T-box$_{182}$ in response to tRNA binding. (A) Cy5 (red star) and Cy3 (green star) are attached at the 5' extension of the T-box (blue) and the 3' end of the T-box, respectively. (B) One dimensional FRET histograms of intra-T-box pair of T-box$_{182}$ alone (black) and with tRNAGly (red). "N" denotes the total number of traces in each histogram from all independent experiments. For "-tRNA" case, two independent measurements were performed. For "+tRNA^{Gly}" case, as no difference was detected, we only performed the measurement once.

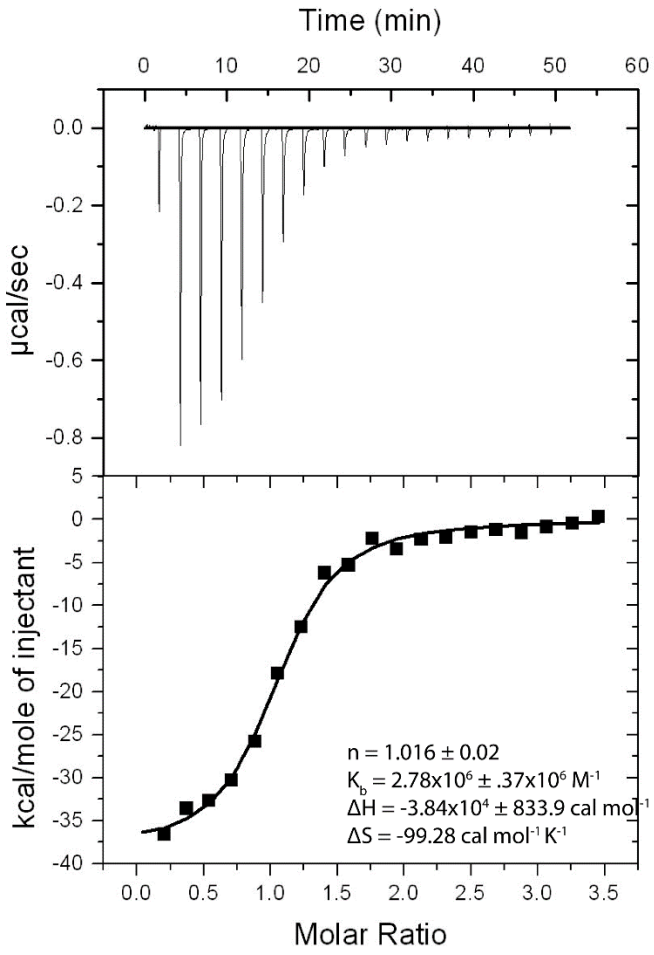


Figure S2.3.2. Isothermal Titration Calorimetry (ITC) of T-box$_{182}$ and tRNA^{Gly} for a T-box with extensions at both ends. The upper panel of the figure shows the heat change due to successive injections of tRNA^{Gly} to a T-box$_{182}$ constructs with extensions at both the 5' and 3' ends (**Supplement file 1**). The first injection of the titration was performed by injecting 0.5 µL of tRNA^{Gly} to minimize contribution of any artifact associate with loading the syringe. The lower panel of the figure shows the binding isotherm obtained by integrating the heat change associated with each injection and plotting it as a function of molar ratio of tRNA^{Gly} to T-box$_{182}$. A theoretical curve is fit to the integrated data using a single-site model with Origin 5.0 (OriginLab). Thermodynamic parameters is derived from a best fit curve ± minimized fitting error by non-linear regression analysis. The binding constant ($1/K_b$) of 360 nM is comparable to the one reported for a similar construct (209 nm)(Zhang and Ferré-D'Amaré 2013), but without the extensions. The experiment shows that the extensions have a negligible effect on binding.

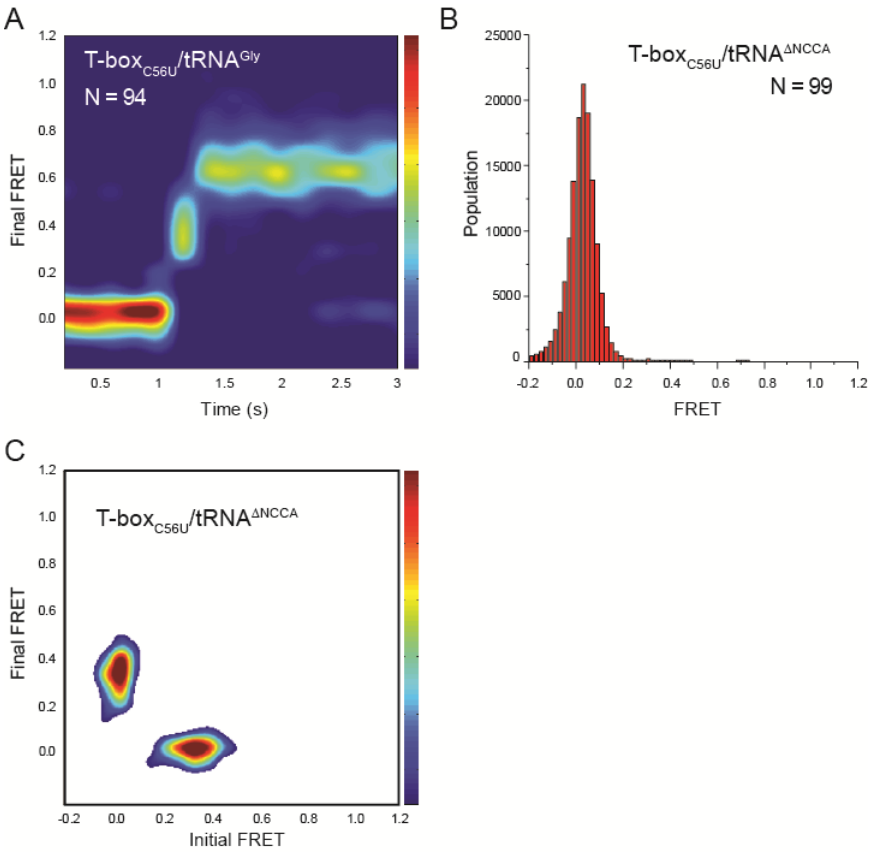


Figure S2.6.1. tRNA-Cy5 binding to T-box$_{C56U}$-Cy3(3'). (A) Surface contour plot of time-evolved FRET histogram of T-box$_{C56U}$-Cy3(3') with Trna^{Gly}-Cy5 with traces showing real-time binding. Figure is plotted in the same way as **Figure 2F**. "N" denotes the total number of traces in each histogram from 3 independent experiments, which are a subset of traces showing real-time binding event in the steady-state measurements. Total numbers of traces in each data sets are indicated in **Figure 6C**. **(B)** FRET histogram of T-box$_{C56U}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. "N" denotes the total number of traces in each histogram from 3 independent experiments. **(C)** TDP of T-box$_{C56U}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. TDPs are generated from all smFRET traces from 3 independent experiments.

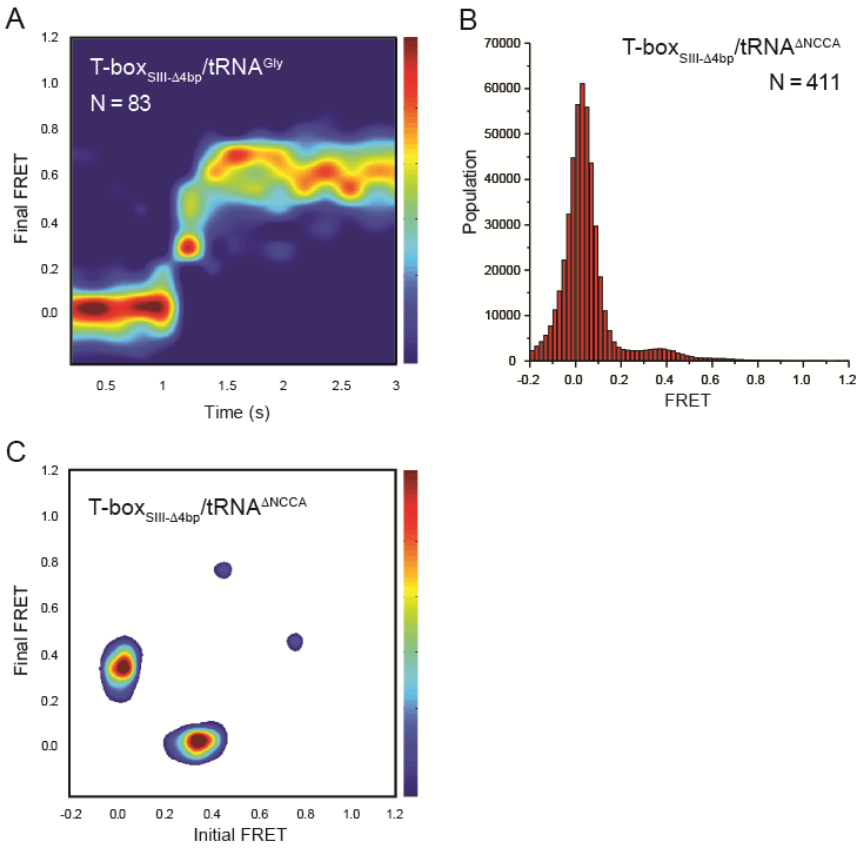


Figure S2.6.2. tRNA-Cy5 binding to T-box$_{SIII-\Delta4bp}$-Cy3(3'). (A) Surface contour plot of time-evolved FRET histogram of T-box$_{SIII-\Delta4bp}$-Cy3(3') with tRNA^{Gly}-Cy5 with traces showing real-time binding. Figure is plotted in the same way as **Figure 2F**. "N" denotes the total number of traces in each histogram from 3 independent experiments, which are a subset of traces showing real-time binding event in the steady-state measurements. Total numbers of traces in each data sets are indicated in **Figure 6C**. **(B)** FRET histogram of T-box$_{SIII-\Delta4bp}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. "N" denotes the total number of traces in each histogram from 4 independent experiments. **(C)** TDP of T-box$_{SIII-\Delta4bp}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. TDPs are generated from all smFRET traces from 4 independent experiments.

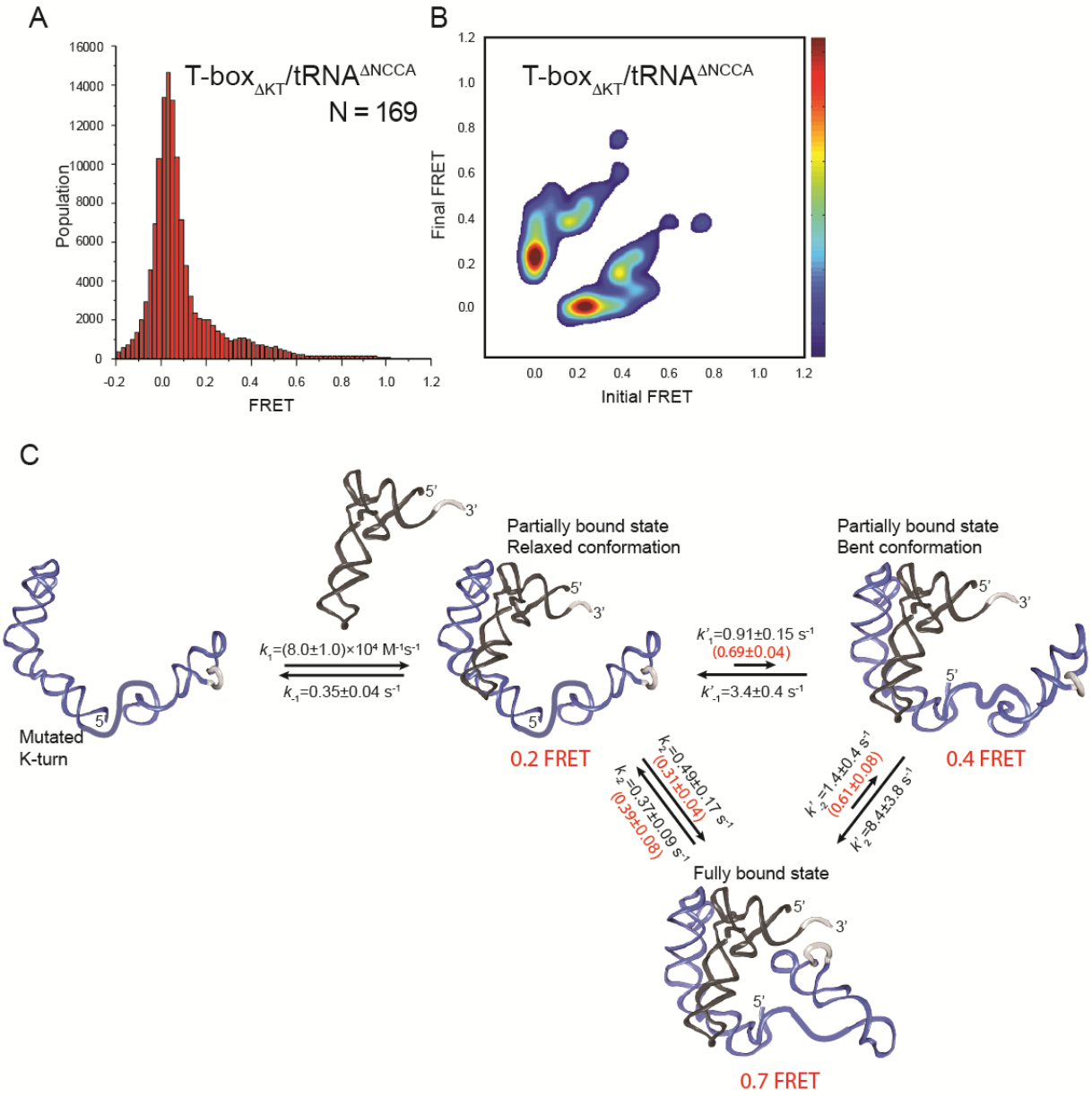


Figure S2.6.3. tRNA-Cy5 binding to T-box$_{\Delta KT}$-Cy3(3'). (A) FRET histogram of T-box$_{\Delta KT}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. "N" denotes the total number of traces in each histogram from 3 independent experiments. **(B)** TDP of T-box$_{\Delta KT}$-Cy3(3') with tRNA$^{\Delta NCCA}$-Cy5. TDPs are generated from all smFRET traces from 3 independent experiments. **(C)** Kinetic model of tRNA^{Gly} binding to T-box$_{\Delta KT}$. Transition rate constants are reported as mean ± S.D from 3 independent measurements. Probabilities of different pathways for transitioning into and out of 0.7 FRET states are marked in red in parenthesis.

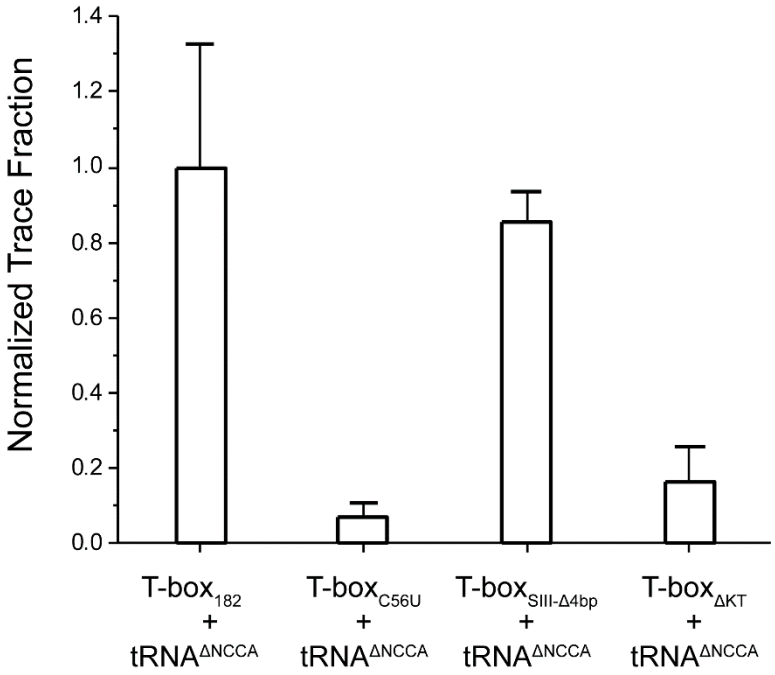


Figure S2.6.4. Normalized FRET Trace percentage of T-box-Cy3(3') and tRNA$^{\Delta NCCA}$-Cy5. The error bars correspond to the standard deviations from 3 or 4 independent measurements.

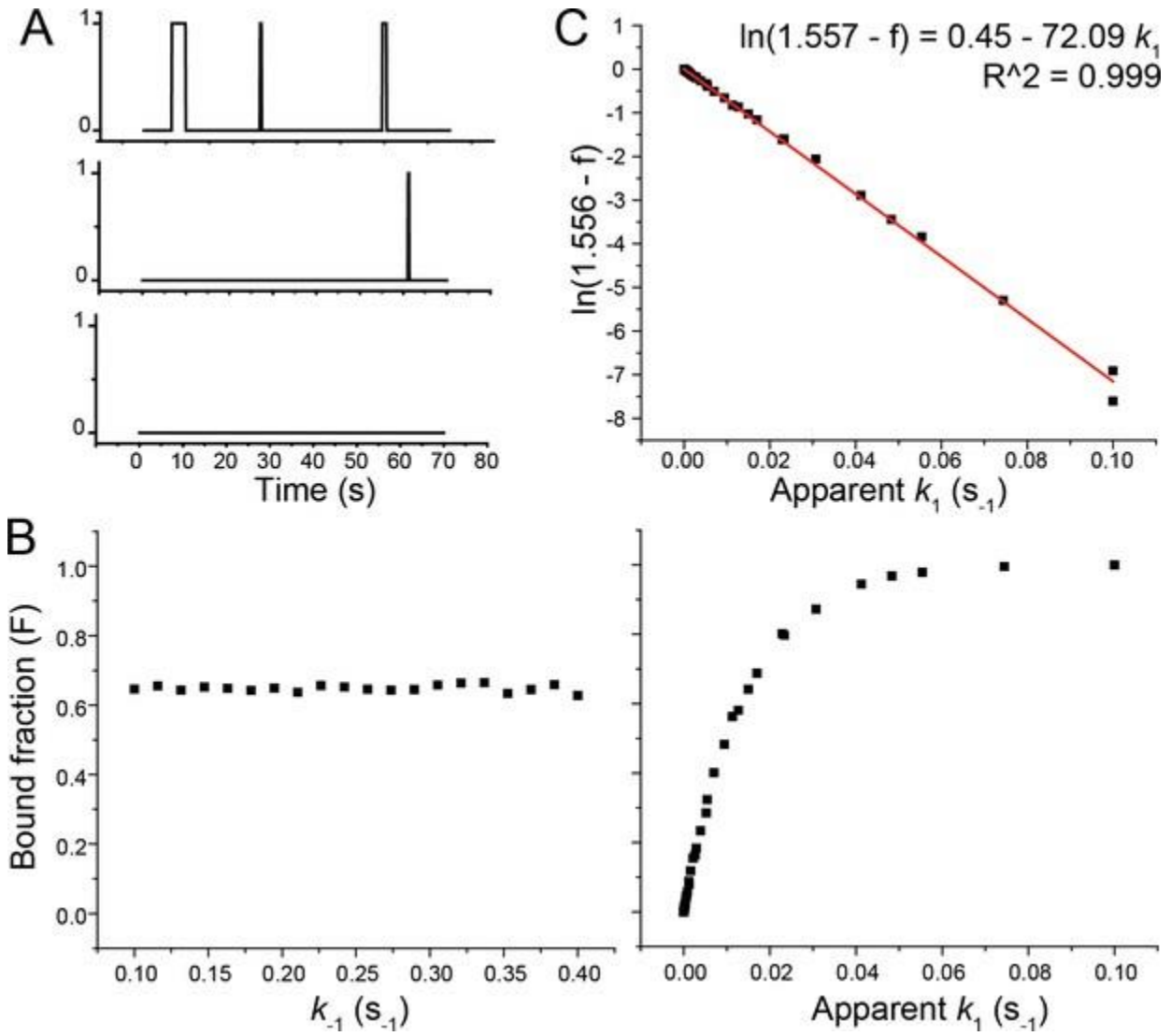


Figure S2.6.5. Determination of k_1. (**A**) Examples of binding traces simulated by the Gillespie algorithm. '1' represents a binding event, and '0' represents the unbound state. If there is a single binding event within the 70 s time window, the trace is included in the bound fraction (*F*). *F* is defined by the number of traces showing a binding event divided by the total number simulated traces. (**B**) Simulation of *F* as a function of k_{-1}, where apparent k_1 is set to be 0.015 s^{-1}, to be consistent with our measured apparent k_1 in the presence of 30 nM tRNA on WT T-box$_{182}$ (left panel). Simulation of *F* as a function of apparent k_1, where k_{-1} is set to be 0.25 s^{-1} to be consistent with our measured k_{-1}. (**C**) Normalized bound fraction (*f*) is *F* normalized to the value at apparent k_1 = 0.015 s^{-1}, to reflect the equivalent quantification of normalized trace fraction to the case of tRNA binding to WT T-box$_{182}$ (Figure 2.6—S4). Red line presents the linear fitting of ln(f_0-*f*) *vs.* apparent k_1.

CHAPTER 3

In vitro quantification and in situ imaging of m^{6}A modifications in mammalian cells

3.1 Introduction

Over 100 types of RNA modifications have been identified to date. Among them, N^6-methyladenosine (m^{6}A) is most prevalent in messenger RNA (mRNA) and various long noncoding RNA (lncRNA) in higher eukaryotes[201]. m^{6}A modifications are widely involved in post-transcriptional gene regulation. The complex and dynamic nature of m^{6}A-mediated regulation enables timely responses to signaling cues and large-scale modulation of gene expression. Therefore, m^{6}A has been shown to be essential for development, and associated with many human diseases[202,203]. The single methyl group is commonly deposited by either a methyltransferase writer complex composed of METTL3, METTL14, and WTAP or by METTL16 methyltransferase[204] and is removed by either FTO[205] or ALKBH5 demethylase[206]. Through its effects on RNA secondary structure and its interactions with m^{6}A binding proteins, m^{6}A modifications affect essentially all known steps during an RNA's lifetime, including alternative splicing, polyadenylation, RNA export, translation, and degradation[207,208]. Despite m^{6}A modification having a consensus DRACH motif (D=A, G or U; R=G or A; H=A, C or U)[41,52], the sub-stoichiometric nature of m^{6}A modification potentially creates large compositional heterogeneity in a single RNA

species, i.e., each RNA of the same species may selectively carry m^6A modification at one or a few DRACH motifs among all[209]. Being able to quantify the extent of m^6A modification at precise sites can greatly advance our current understanding of how changes in the m^6A modification pattern (the site and fraction) are modulated by signaling cues, and are then linked to various functional consequences.

Due to its important roles, techniques have been developed and applied to detect and quantify m^6A modification. Detection of m^6A modification is primarily facilitated by various high-throughput sequencing-based methods utilizing antibodies and chemical crosslinking [210425241]. Although these sequencing-based methods can map m^6A candidate sites at the transcriptomic level, they cannot provide the fraction of modification at each site, due to factors such as antibody binding efficiency, specificity and cross-linking reactivity[211]. Real-time quantitative PCR (qPCR) was previously applied for locus specific detection of pseudouridine (ψ) modification though chemical labelling of ψ residue, causing a shift in the melting peak of the resulting qPCR amplicons[212]. Similar quantitative methods were recently developed for detection of m^6A. These methods utilize enzymatic activities followed by qPCR, including differential ligation efficiency of T3 and T4 DNA ligases[213,214], differential reverse transcription activity of Tth and BstI reverse transcriptases[215,216], and a combination of selective elongation of DNA polymerase and ligation[217]. Although these polymer elongation and ligation-based methods are successful at modification discrimination and can report the relative m^6A abundance change, absolute quantification using these methods were only applied on *MALAT1*, an abundant

lncRNA. In addition, the potential sequence-dependence of these enzymatic activities requires caution for general applications to these methods[218,219]. Considering these potential pitfalls, absolute quantification using these methods would require calibration curves using fully modified and fully unmodified RNA for each target m^{6}A site, which is expensive. The only available qPCR-independent method that can provide absolute quantification of m^{6}A fraction site-specifically is SCARLET (site-specific cleavage and radioactive labeling followed by ligation-assisted extraction and TLC)[220]. However, the sophistication of the method and its requirement for radioactive labeling prevents its broad application. Very recently, endoribonuclease digestion-based sequencing methods have been developed, which rely on selective cleavage of unmethylated A at the ACA motif[221,222]. These approaches provide single-base resolution for identification of modifications site with relative quantitative information but are limited to m^{6}A sites carrying the ACA motif, as well as regions that contain relatively sparse ACA motifs[222].

In addition, since RNA localization is tied with their function, we believe the localization of m^{6}A modified RNA and unmodified RNA may also indicate its functionality. Neither antibody pull-down sequencing or SCARLET can reveal spatial distribution of the m^{6}A modification in the subcellular domains. And so far there is only one group reporting single-cell imaging m^{6}A modification based on proximal ligation. However, it still requires m^{6}A antibody and the various binding affinity against different m^{6}A consensus sequences together with the low efficiency limits its application and only suitable for abundant mRNA or lncRNAs.

To address these challenges, we present an easy-to-implement method for quantifying m^{6}A fraction at specific loci from extracted total RNA and an imaging strategy to imaging m^{6}A modifications of given sites of a target RNA. The first method utilizes a recently reported deoxyribozyme (DR) by Sednev *et al*., VMC10, to discriminate between A and m^{6}A containing RNA[223](**Figure 3.1A**). We chose VMC10 DR because it cleaves unmethylated A with reasonably high and robust efficiency while its cleavage of m^{6}A remains low even after a long incubation time[223]. Sednev *et al*. demonstrated that the remainder of intact RNA after VMC10 DR treatment is correlated with the known methylation level of specific sites of abundant endogenous RNAs[223]. However, quantification of the absolute m^{6}A fraction requires additional characterization and correction of potential false positive and false negative due to various factors. Without these additional characterizations, modification levels of different m^{6}A sites cannot be compared due to sequence dependence of DR. In our method, we quantify and correct for the potential errors caused by sequence-dependent incomplete cleavage of the DR, and the presence of non-target RNAs from the total RNA. We show that our method can be used to robustly quantify the absolute m^{6}A fraction at specific loci on endogenous RNAs with a broad range of cellular abundance. For the imaging strategy, we take advantage of the DR and primer extension and imaged m^{6}A sites of a few RNA targets in fixed cells. Finally, we extensively discuss the limitations of the method and factors that need to be considered when applying this method.

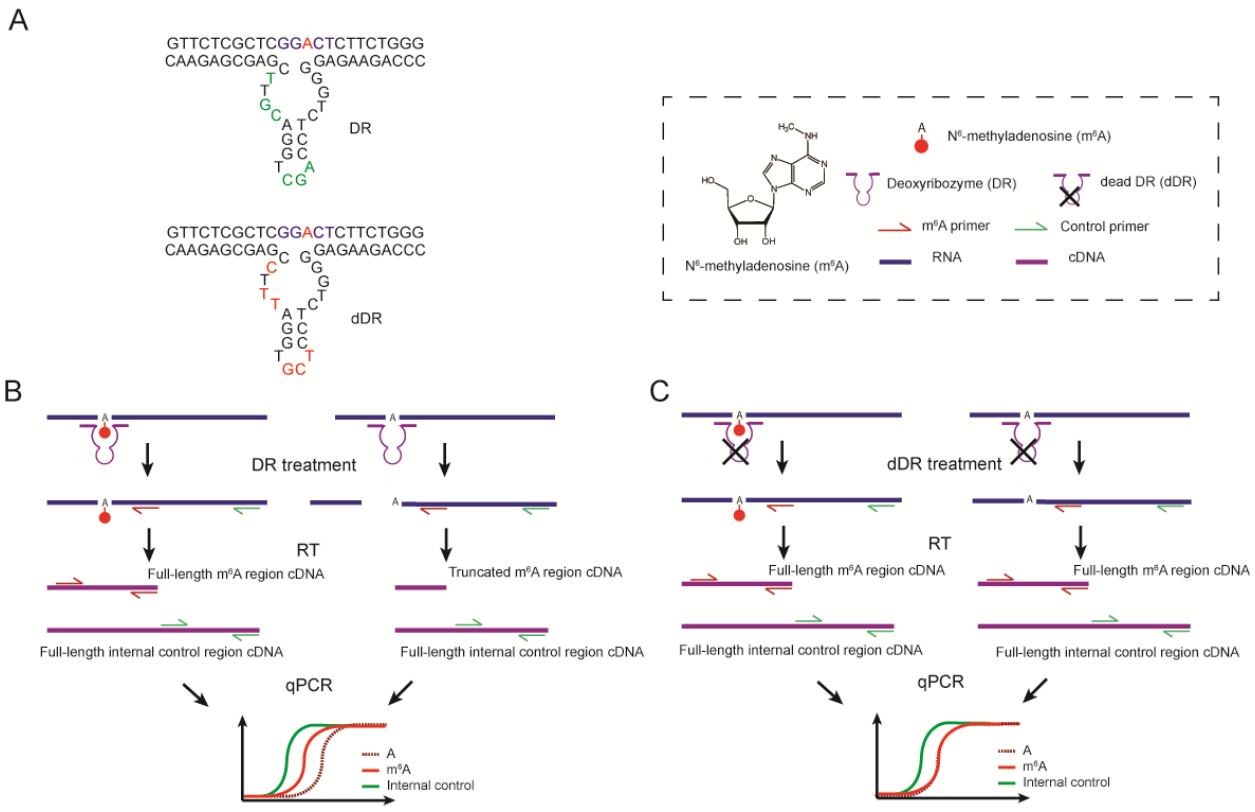


Figure 3.1. Workflow of the DR-based m^{6}A quantification method. (A) Representative schematic of the active deoxyribozyme (DR) and the inactive deoxyribozyme (dDR). (B) Unmodified RNA is selectively cleaved by DR upstream of the target site, while m^{6}A modified RNA remains uncleaved. The remaining uncleaved RNAs are then quantified using RT with gene-specific reverse primer and qPCR. To control for variations in RNA input, an adjacent region on target RNA is also quantified with RT and qPCR as an internal reference. (C) In the negative control sample, RNA is treated with a nonfunctional version of DR (dDR). Both m^{6}A modified and unmodified RNA targets remain uncleaved, and are subsequently quantified with RT and qPCR.

3.2 Results

3.2.1 Specificity and sequence dependence of DR cleavage efficiency

We first verified the cleavage efficiency of DR on a variety of fully modified or unmodified sites. For this purpose, we employed a 460-nt *in vitro* transcribed RNA from a gene block sequence with only one adenine in the sequence (referred as “GB RNA”

hereafter), and 35 to 41-mer synthetic RNA fragments with sequences around *MALAT1* 2515 site, *MALAT1* 2577 site, and *ACTB* 1216 site (**Table 3.1**). Each of these targets has either m^6A or A at the respective m^6A sites. The RNAs were treated with corresponding 40-mer VMC10 DR (referred as DR for simplification hereafter) and subsequently analyzed by denaturing polyacrylamide gel electrophoresis (PAGE). For all the targets, the RNA fragments with unmethylated A were cleaved with high efficiency, and the cleavage efficiencies of modified RNAs were consistently below 5% (**Figure 3.2A-D**). In addition, the cleavage efficiencies on the unmodified RNAs were sequence-dependent (**Figure 3.2E**), ranging from 50% to 82% for our tested cases.

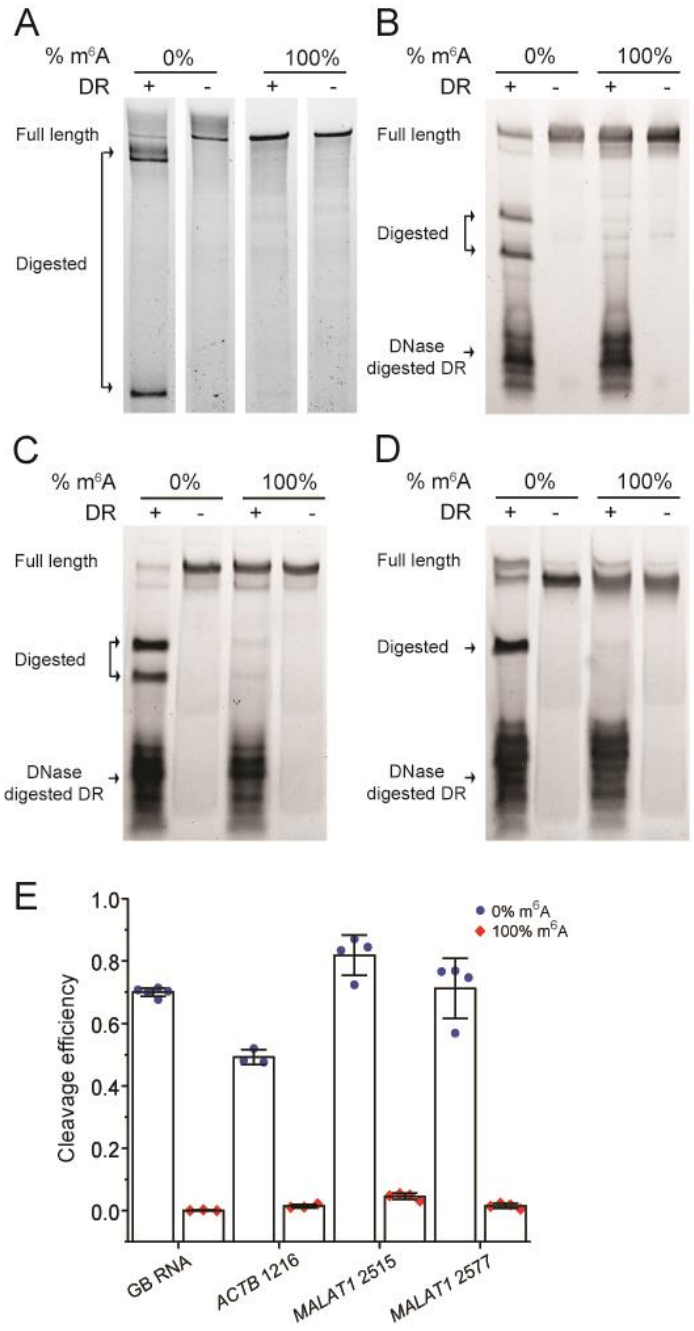


Figure 3.2. DR specifically cleaves unmodified RNAs, and its cleavage efficiency depends on sequence context around m^{6}A sites. PAGE showing DR cleavage of 0% and 100% modified (A) GB RNA, RNA fragments containing modification site of (B) *ACTB* 1216, (C) *MALAT*1 2515 and (D) *MALAT1* 2577. (E) Bar plot of the cleavage efficiencies of m^{6}A modified and unmodified target sites as quantified from PAGE. Error bars indicate mean ± s.d. for 3 to 4 independent cleavage reactions. The DR is digested with DNase after DR cleavage reaction, as the DR may migrate to the same location as the input or cleaved RNA and affect data interpretation.

3.2.2 Method for absolute quantification of m^{6}A fraction

We designed a quantification assay using reverse transcription (RT) and qPCR. As shown in Figure 3.1, DR is designed for each modification site based on a VMC10 construct. The total RNA is subjected to DR treatment during which only unmethylated RNAs upstream of the target site are cleaved. Thus, after the DR treatment, the amount of cleaved RNA should be inversely proportional to the methylation fraction (*Fm*) of RNA at the target site. The remaining RNA can be quantified using RT-qPCR. In order to control for any variations in the initial RNA input, we use RT-qPCR to also detect levels of adjacent uncleaved regions on the same target RNA as an internal reference.

Theoretically, the modified fraction calculated from qPCR can be written as

$$F_m = 2^{-\Delta\Delta Ct} \qquad (1),$$

in which

$$\Delta\Delta Ct = (Ct_{+DR-m6A} - Ct_{-DR-m6A}) - (Ct_{+DR-ref} - Ct_{-DR-ref}) \qquad (2)$$

where $Ct_{+DR-m6A}$ and $Ct_{+DR-ref}$ are the qPCR *Ct* values at the m^{6}A site and a nearby reference site in the DR treated sample, whereas $Ct_{-DR-m6A}$ and $Ct_{-DR-ref}$ are the *Ct* values at the m^{6}A and the reference site without DR digestion.

However, the measured $\Delta\Delta Ct$ only reflects the digested fraction of the RNA substrate, i.e., Eq. (1) only holds when digestion efficiency of the unmodified template is 100% and digestion efficiency of the modified template is 0%. Incomplete cleavage of unmodified A will lead to false positive, and cleavage of the m^{6}A will lead to false negative. Based on the previous study and our tested cases (**Figure 3.2**), the cleavage of VMC10

DR on m^{6}A sequence is minimal, leading to insignificant error caused by false negative (**Figure S3.1**). In addition, it is practically difficult and expensive to generate *in vitro* purified template containing 100% modified m^{6}A to account for the exactly false negative error at each m^{6}A site of interest, we therefore left out the correction factor for false negative error in our final calculation. On the other hand, the false positive error can be significant due to the sequence-dependent incomplete cleavage of unmodified RNA by the DR and needs to be corrected for each m^{6}A site of interest.

We therefore consider two major factors that may contribute to the false positive error due to incomplete digestion, and quantify the effect of the two factors to extract the true modification fraction: the intrinsic sequence-dependent digestion efficiency and the presence of a large amount of non-target RNAs from the total RNA extract. We define F_{DR} as a correction factor to count for the incomplete DR digestion efficiency, which has to be determined for each m^{6}A target (**Figure 3.2**). We can determine F_{DR} at each m^{6}A site of interest by performing the DR digestion followed by RT-qPCR using the *in vitro* transcribed unmodified RNA:

$$F_{DR} = 1 - 2^{-\Delta\Delta Ct} \qquad (3),$$

in which $\Delta\Delta Ct$ is determined as in Eq. (2). We define F_N as the ratio of DR digestion efficiency of an RNA target in total RNA over digestion efficiency of a pure RNA target, to account for the potential drop of DR efficiency due to the presence of non-target RNAs. We can determine F_N by performing DR digestion using the same *in vitro* transcribed unmodified RNA mixed with total RNA, and compare with F_{DR} from Eq. (3):

$$F_{DR}F_N = 1 - 2^{-\Delta\Delta Ct} \quad (4)$$

in which *ΔΔCt* is determined as in Eq. (2). With the quantification of F_{DR} and F_N, the corrected modification fraction follows:

$$(1 - F_m)\, F_{DR}F_N = 1 - 2^{-\Delta\Delta Ct} \quad (5).$$

We therefore can calculate F_m as:

$$F_m = \frac{2^{-\Delta\Delta Ct} + F_{DR}F_N - 1}{F_{DR}F_N} \quad (6).$$

3.2.3 Validation of the absolute quantification of m^{6}A fraction using pure RNA

To test the feasibility of the method to quantify the m^{6}A methylation fraction, we used GB RNA with methylation fractions ranging from 0% to 100%. We performed DR treatment on GB RNA and estimated the cleaved fractions by denaturing PAGE. The cleaved fraction was linearly dependent on the methylation fraction of the input RNA (**Figure 3.3A, B**). Next, we tested whether we can use RT-qPCR to quantify the absolute methylation fraction. As this quantification is performed on *in vitro* purified RNA, only F_{DR} is needed to correct for *Fm*. Based on Eq. (3), using the 100% unmodified GB RNA, we measured $Ct_{+DR-m6A}$ and $Ct_{+DR-ref}$ at the m^{6}A site and a nearby reference site in the DR treated sample, and $Ct_{-DR-m6A}$and $Ct_{-DR-ref}$ at the m^{6}A and the reference site using a negative control containing identical amount of RNA but without the DR. We found that in addition to the expected larger $Ct_{+DR-m6A}$ compared to $Ct_{-DR-m6A}$, there was a consistent difference between $Ct_{+DR-ref}$ and $Ct_{-DR-ref}$. We speculated that this

difference in *Ct* values at the reference site might be due to changes in the RNA secondary structure upon DR binding that can affect RT efficiency. To create a more accurate negative control, we designed a non-functional version of DR ("dead" DR or dDR) (**Figure 3.1A, C**), which has mutations in the AGC triplet, CG dinucleotide, and position 19 important for the catalytic activity of 8-17 family of enzymes (28). We tested the activity of dDR on multiple targets, for all of which digestion of RNA was undetectable (**Figure S3.2; Figure S3.3**). Indeed, using the dDR treated RNA as a negative control, the difference between $Ct_{+DR-ref}$ and $Ct_{-DR-ref}$ was eliminated (**Figure S3.4**). Based on Eq. (3), we determined F_{DR} of the synthetic RNA to be 0.49 ± 0.08 (mean ± s.d.). With the F_{DR} correction, we showed that the estimated *Fm* correlated well with the input m^{6}A methylation fractions (**Figure 3.3C; Figure S3.5**).

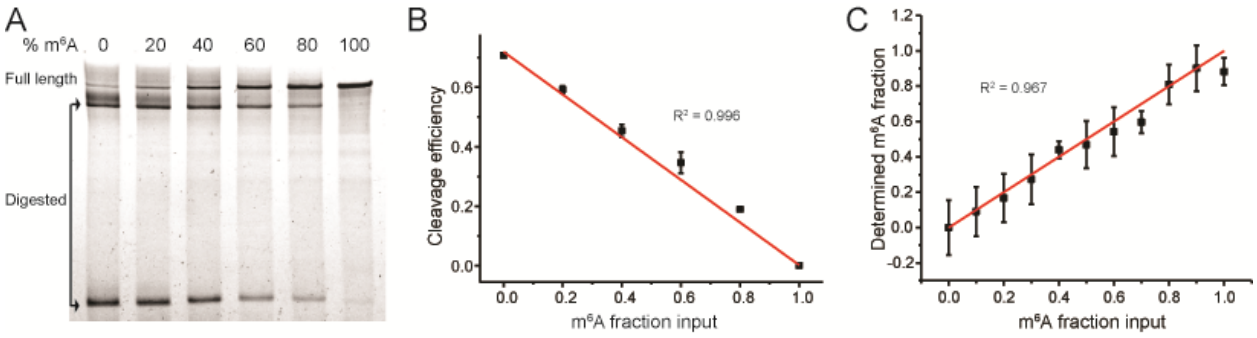


Figure 3.3. Validation of the method for absolute quantification of m^{6}A fraction. GB RNA containing varied m^{6}A fractions is used as a model system. (A) PAGE showing DR cleavage fraction of the GB RNA. (B) Linear relationship between the input m^{6}A fraction and the cleavage fraction of RNA by DR as quantified from the PAGE gel in (A). Error bars indicate mean ± s.d. for 3 biological replicates. (C) Estimated modification fraction as a function of input m^{6}A fraction for the GB RNA. Error bars indicate mean ± s.d. for at least 3 biological replicates.

3.2.4 DR cleavage efficiency in presence of non-target RNAs

Next, we evaluated how the presence of total RNA affects the cleavage efficiency of the DR. The presence of the large amount of non-target RNAs may compete for DR binding, consequently decreasing its cleavage efficiency at the target site in total RNA as opposed to purified RNA. We accounted for this potential decrease in efficiency with F_N correction factor, which we measured using three RNA transcripts: (1) the GB RNA used above, which is naturally missing in total RNA; (2) a *PLAC2* RNA fragment containing two target sites, which is of low abundance in HeLa cell line; and (3) an unmethylated A site in the endogenous *ACTB* mRNA. The A1165 site on *ACTB* mRNA was chosen as the unmethylated A site because it was not detected in the sequencing-based studies[46,224], nor contains the DRACH consensus motif. F_{DR} of these three RNAs were measured with *in vitro* transcribed RNAs based on Eq. (3) (**Figure 3.4A, B**). Then the *in vitro* transcribed GB RNA and the *PLAC2* RNA fragment were spiked into the total RNA respectively to determine F_N based on Eq. (4). For the unmethylated A site in the *ACTB* mRNA, F_N was determined by measuring the total RNA directly.

In order to increase the binding specificity of DR, we compared a 60-mer DR and a 40-mer DR. We found that F_{DR} values of 60-mer DR were higher than those of 40-mer DR (**Figure S3.2**), likely due to a higher hybridization efficiency by 60-mer DR. F_N values were consistently high for all tested RNAs, with the lowest F_N values being 0.78 ± 0.02 for 40-mer DR and 0.93 ± 0.02 for 60-mer DR, demonstrating that ability of our method to quantify m^6A status should not be compromised by the presence of total RNA, and that 60-mer can slightly outperform 40-mer DR (**Figure 3.4C**). Overall, the average F_N values

were determined to be 0.94 ± 0.1 for 40-mer DR and 0.98 ± 0.05 for 60-mer DR. In addition, we compared the effect of DR concentration on F_N values using *PLAC2* m^{6}A 2 DR as an example. We found that the cleavage efficiency is consistent at a wide range of DR concentrations as long as the molar concentration of DR is at least 10-fold higher than the molar concentration of RNA target (**Figure S3.6A**). Likewise, the F_N values stayed consistently at about 1 for all tested concentrations of DR that were saturated compared to target RNA. Given these results, we can simplify Eq. (6) to be

$$Fm = \frac{2^{-\Delta\Delta Ct} + F_{DR} - 1}{F_{DR}} \qquad (7).$$

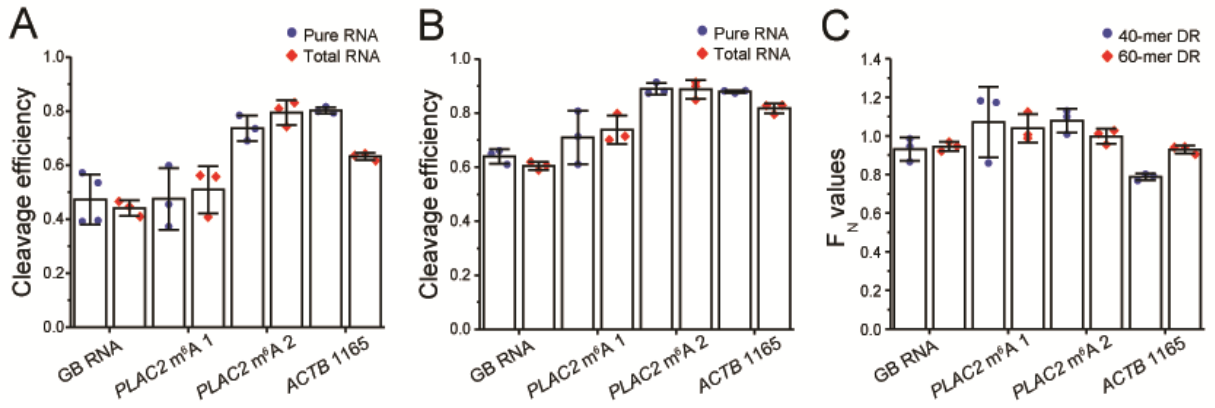


Figure 3.4. The cleavage efficiency of DR is not compromised by the presence of total RNA. The cleavage efficiencies of the GB RNA, two m^{6}A sites in *PLAC2*, and *ACTB* 1165 by (A) 40-mer DR and (B) 60-mer DR in presence and absence of total RNA are determined by RT and qPCR. (C) F_N correction values for the GB RNA, two m^{6}A sites in *PLAC2*, and *ACTB* 1165 for 40-mer and 60-mer DRs as determined from cleavage efficiencies in (A) and (B). All error bars report mean ± s.d. for 3 biological replicates.

3.2.5 Quantification of m^{6}A fraction of endogenous sites

Having developed and validated our method, we applied it to determine the methylation fraction of several endogenous sites that were identified as potential m^{6}A sites by RNA sequencing from more than one study: *MALAT1* 2515 (chr11 65500276), 2577 (chr11 65500338), and 2611 (chr11 65500372), *ACTB* 1216 (chr7 5527743), *LY6K* 1171 (chr8 142703380), *MCM5* 2367 (chr22 35424323), *SEC11A* 1120 (chr15 84669674), *INCENP* 912 (chr11 62130275), 967 (chr11 62130330), and 1060 (chr11 62130423), *LMO7* 2822 (chr13 75821377), and *MRPL20* 549 (chr1 1402080) (The genome position based on GRCh38.p13 Primary Assembly of m^{6}A site is indicated in parenthesis)[42,46,224]. The selected RNAs vary from low to high abundance in Hela cells, and some of them contain more than one modification site. Specifically, four of the targets (*MALAT1* 2515, *MALAT1* 2577, and *MALAT1* 2611 and *ACTB* 1216) were previously measured using the SCARLET assay[220], therefore serving as an additional validation of our method. To apply our method, a DR and a dDR were designed for each site. Due to the higher F_{DR} and F_N values with 60-mer DR, we chose to use the 60-mer DR for all endogenous RNAs. For each target site, we first generated *in vitro* transcribed RNAs containing the m^{6}A sites of interest, and performed DR digestion on these *in vitro* transcribed unmethylated RNAs to get F_{DR} for each site. The F_{DR} values were all greater than 0.49 and again varied among different RNAs (**Figure 3.5A; Figure S3.3**).

The methylation fractions of the endogenous sites were determined to range from 0.13 to 0.92 (**Figure 3.5B**). Notably, our results show comparable methylation fractions for *MALAT1* 2515, *MALAT1* 2577, and *MALAT1* 2611 and *ACTB* 1216 as in the SCARLET

assay[220]. While the generally consistent results between our methods and SCARLET assay help validate our assay, we did notice that values measured in our assay are slightly higher than those from SCARLET. One possible explanation for this slight variation can be the splint ligation step used in the SCARLET assay, in which the DNA oligo needs to be ligated to the RNase H cleaved RNA carrying either unmodified A or m^6A at the 5' end[220]. It is possible that the splint ligation is less efficient for the m^6A containing RNA, and therefore, underestimates the m^6A fraction in SCARLET assay.

Having validated that the method is able to quantify m^6A fractions one site at a time, we investigated whether the method can be utilized in a multiplexed way. As a proof of concept, we chose to remeasure three m^6A sites with varying levels of methylation: *MALAT1* 2611, *ACTB* 1216, and *INCENP* 912. For each biological replicate, the three corresponding active DRs were combined in one reaction and the three dDRs were combined in another reaction. Similarly, RT reactions were also performed with combined RT primers for all three targets at either m^6A site or internal control site, followed by separate qPCR reaction for each target RNA. The multiplexed measurements of the methylation fractions were comparable to the ones we previously obtained in individual measurements (**Figure 3.5C**). These results support that method is accurate at measuring m^6A fractions in a multiplexed fashion and can be further adapted for high throughput measurements.

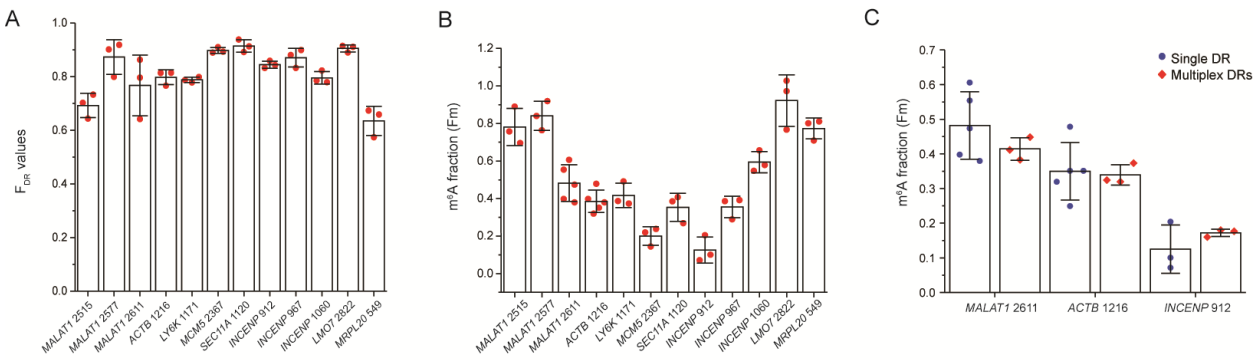


Figure 3.5. Determination of m^{6}A fraction of endogenous sites. (A) The cleavage efficiencies (F_{DR}) of the in vitro transcribed RNA by 60-mer DR as determined by RT and qPCR. (B) Determined m^{6}A modification fractions of the 12 endogenous sites. (C) Determined m^{6}A modification fractions for three endogenous targets using single and multiplexed measurements. All error bars report mean ± s.d. for at least 3 biological replicates.

3.2.6 Effect of the nearby modifications on the DR cleavage efficiency

M^{6}A modifications often exist in clusters[41]. In addition, other type of RNA modifications are identified in mRNAs and lncRNAs[202]. Therefore, the possible effect of the nearby modifications needs to be considered when applying this method. We designed synthetic RNA containing a nearby m^{6}A, m^{1}A or ψ and measured their effects on cleavage efficiency of DR by PAGE analysis (**Figure 3.6, Figure S3.7**). The cleavage efficiency was unaffected by the presence of m^{6}A modification 2 and 4 nt upstream and downstream of the target site, suggesting that the method can be used to quantify m^{6}A fraction in RNA that contain m^{6}A modifications in clusters. Furthermore, ψ modification had a very minimal decrease in the cleavage efficiency of DR 2 nt away from target site and had no effect on the cleavage when present 4 nt away from the target site. Finally, m^{1}A modification, which can affect the Watson-Crick base pairing, significantly decreased

the cleavage efficiency at 2 nt away from the target site and moderately decreased the cleavage efficiency at 4 nt away from the target site. Overall, the results indicate that other nearby RNA modifications that do not affect the base pairing with the DR are not likely to affect the DR cleavage efficiency even when placed as close as only 2 nt away from the target site. However, nearby RNA modifications that weaken the base pairing with the DR will have a larger effect on the DR activity, but the effect decreases when the modification is more distal from the target site.

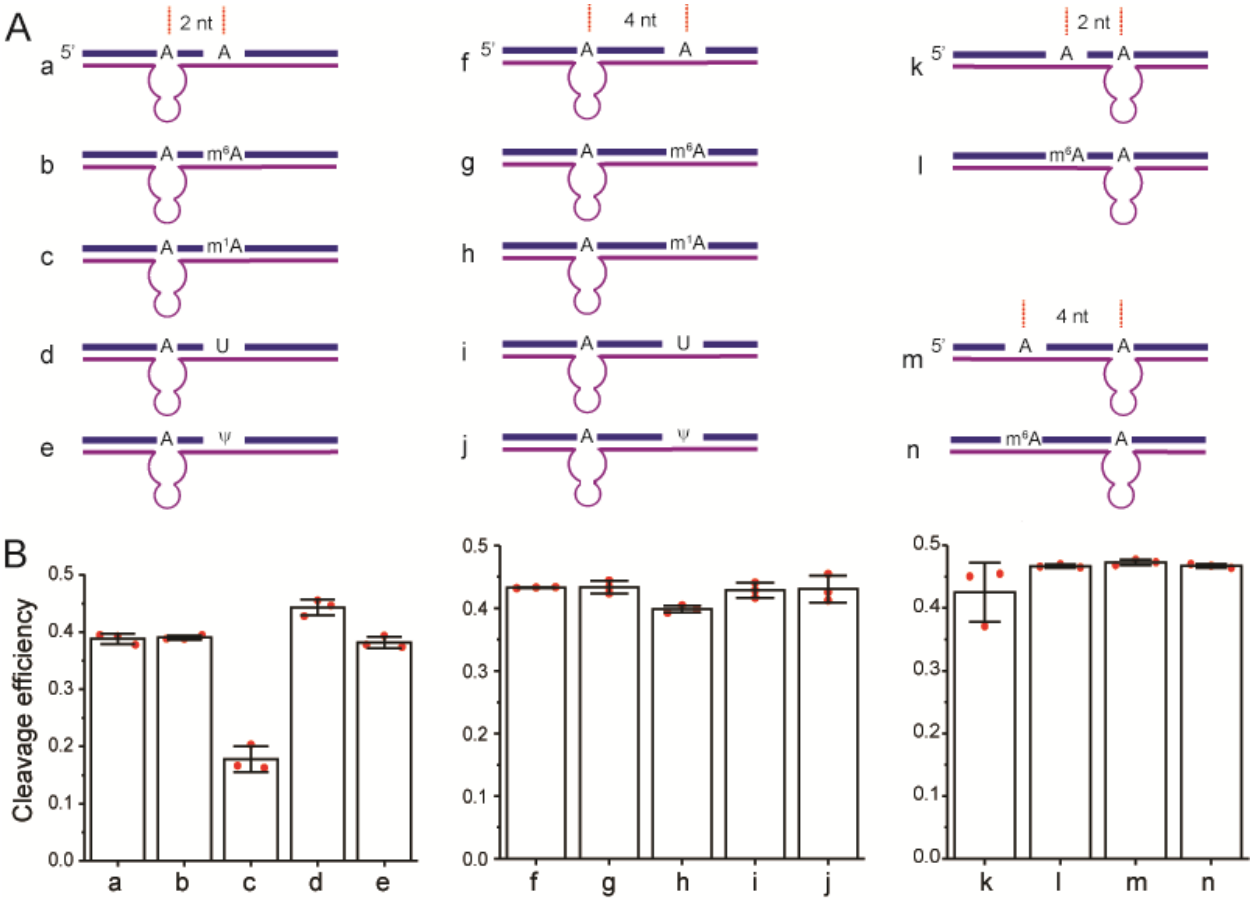


Figure 3.6. The effects of nearby RNA modifications on cleavage efficiency (F_{DR}) of DR. (A) Scheme of 35-nt synthetic RNA containing m^6A, m^1A, and ψ modifications. (B) Bar plot of the cleavage efficiencies of synthetic RNAs as quantified from PAGE. Error bars indicate mean ± s.d. for 3 independent DR cleavage reactions.

3.2.7 Design of imaging m^6A modification in situ by DR

So far, all of the m^6A quantification methods are performed on purified total RNA transcripts. However, disruption of the cells limits the spatial and functional relevance of these data. In order to understand the biological function of m^6A modifications, we propose an imaging assay to leverage the DR based method to reveal the localization of a given m^6A site inside fixed cells. The DR is first introduced to cells, hybridize to the target RNA site and digest the unmodified RNA. Second, the DR is washed away and a fluorophore (Alexa-647) conjugated primer of the target RNA is hybridized. Reverse transcription (RT) is then performed and the primer on modified RNA template can be elongated and we can wash out the non-extended primer while keeping the extended one. By rehybridize Cy3 labeled FISH probes of the target RNA, the m^6A modified target RNA will have both primer and FISH probe signal whereas the unmodified ones only contain FISH signal (**Figure 3.7**).

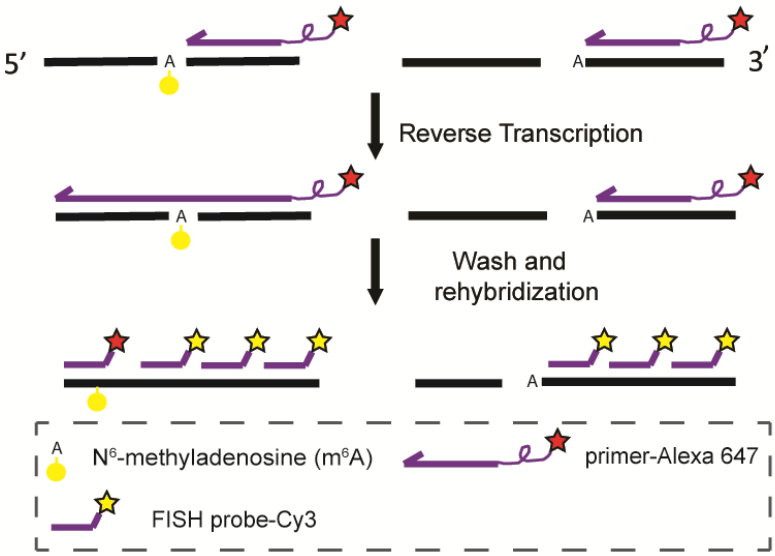


Figure. 3.7 Work flow of DR imaging assay. After washing DR in fixed cell, Alexa 647 labeled primer is hybridized to the RNA target site followed by reverse transcription. The primer on the modified RNA template is extended and the non-extended on can be washed out. Cy3 labeled FISH probes are hybridized to the target RNA after the wash.

3.2.8 Validation of the imaging method using reported m^{6}A sites

We first access the DR digestion and RT efficiency in fixed cell. To compare the DR cleavage efficiency in fixed cell and in vitro, we chose to check unmodified site (1165) on *ACTB* mRNA by the RT-qPCR assay mentioned above. The qPCR results showed that DR has similar efficiency in fixed cells compared with total RNAs (**Figure 3.8A**). In order to evaluate the RT efficiency in situ, we chose to image unmodified site (2546) on *MALAT1*. *MALAT1* RNA is predominantly localized inside nuclear speckle and we estimated the RT efficiency by measuring the primer intensity between samples treated with or without RT after the wash step (**Figure 3.8B, C**).

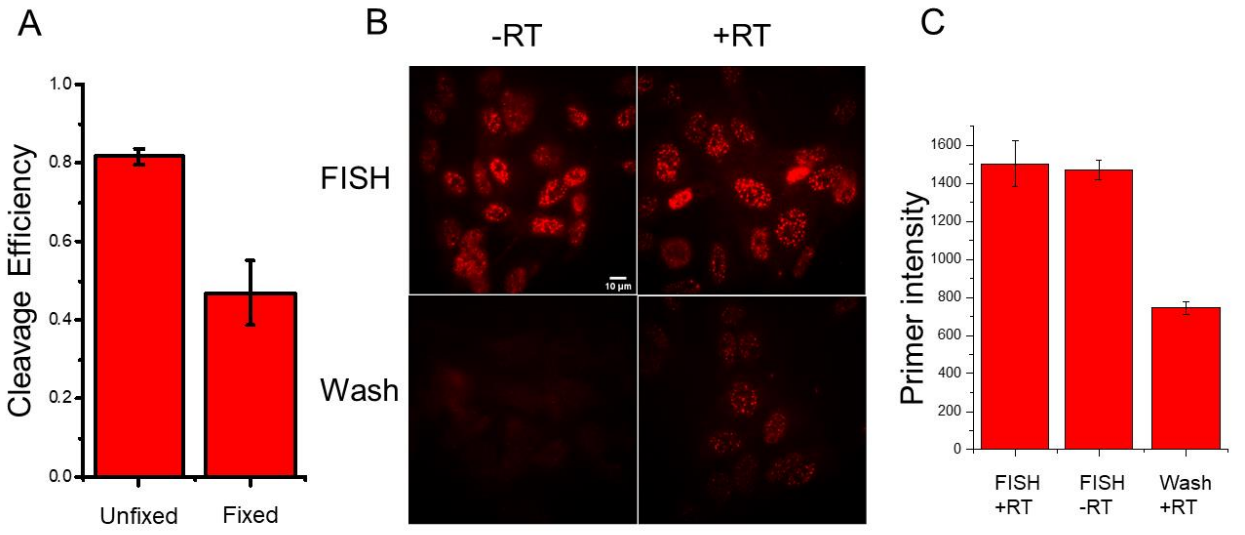


Figure. 3.8 Validation of DR cleavage and reverse transcription (RT). (A) DR cleavage efficiency of ACTB1165 site from total RNA purified from unfixed and fixed cells. (B) Representative images of RT test. (C) Primer intensity before and after wash of RT treated and untreated samples. Error bars indicate mean ± s.d. of 3 biological replicates.

Having validated the efficiency of each step, we first chose the unmodified site (1165) of *ACTB* mRNA in HeLa cells to check whether the imaging assay can perform efficiently. Here, we introduced a non-functional deoxyribozyme (dDR) as a control for the DR hybridization and cleavage step. By calculating the colocalization between the primer and FISH signal, we were able to observe ~50% decrease on the DR treated sample, consistent with the DR digestion efficiency in situ. In order to know whether m^6A modification can compromise the colocalization decrease, we chose to image the previously reported site (1216) on *ACTB* mRNA. The DR treated sample showed a higher colocalization compared to the unmodified site indicating that our hypothesis is correct. Mettl3 functions as m^6A modification writer and we reason that knocking down Mettl3 can

lead to decrease m^{6}A level at *ACTB* 1216 site. With our DR imaging assay, the results showed the DR treated sample has a similar colocalization as the unmodified site (**Figure 3.9**). These experiments have demonstrated that DR imaging assay can be used to measure the m^{6}A level on *ACTB* mRNA.

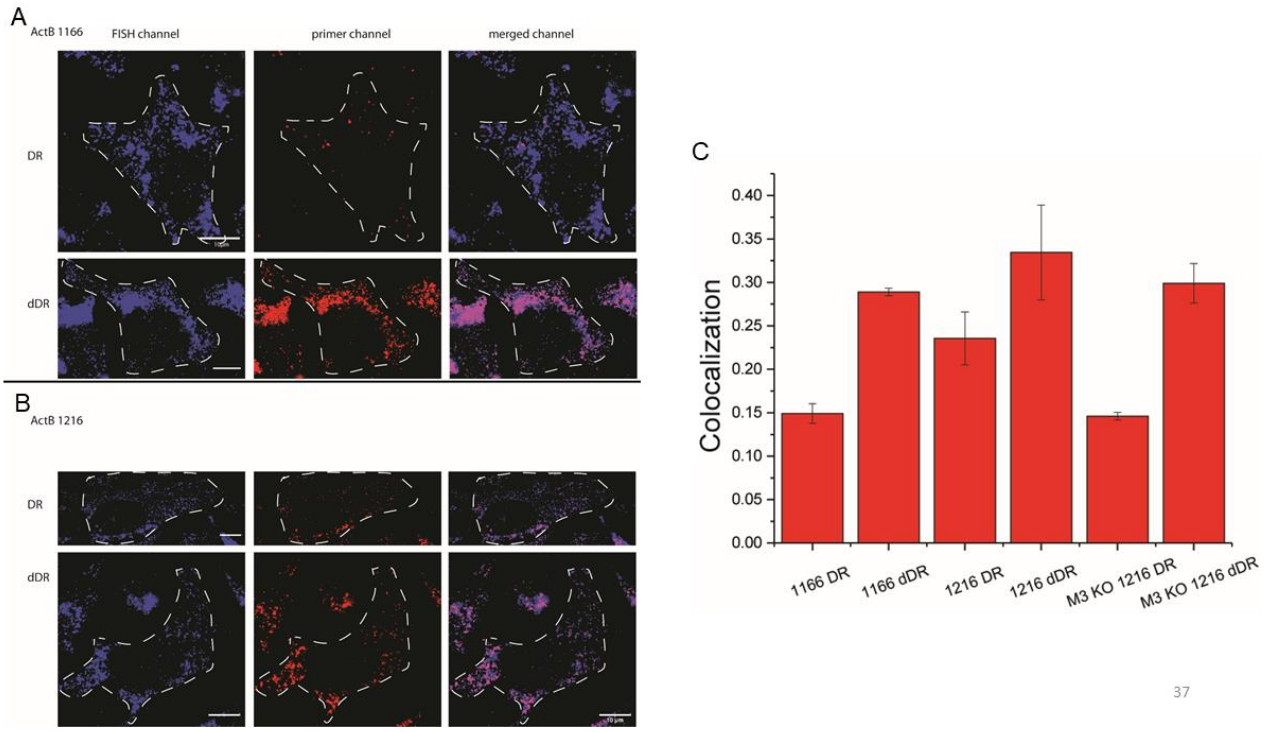


Figure 3.9 DR imaging assay on ACTB mRNA. (A) & (B) Representative two color FISH images of ACTB 1166 and 1216 sites after DR assay. (C) Colocalization analysis of primer and FISH channel. Scale bar = 5μm. Error bars indicate mean ± s.d. of 3 biological repeats.

The *ACTB* 1216 site was reported only to be 21% m^{6}A modified and we decide to select some highly modified sites for validation. *MALAT1* 2515 and 2577 sites were reported to contain 61% and 80% m6A modification respectively and we performed DR

imaging assay on these two sites and a non-modified site (2546) as a control. We quantified the m^6A level by calculating the ratio between primer and FISH signal and surprisingly, both modified sites showed similar ration as the unmodified site (**Figure 3.10**). We wondered if the complexity of the speckle components affects the digestion result and tried different fixation methods but all of them gave the same results (data not shown). These experiments indicate DR can digest highly modified *MALAT1* RNA sites for a reason that is unknown yet.

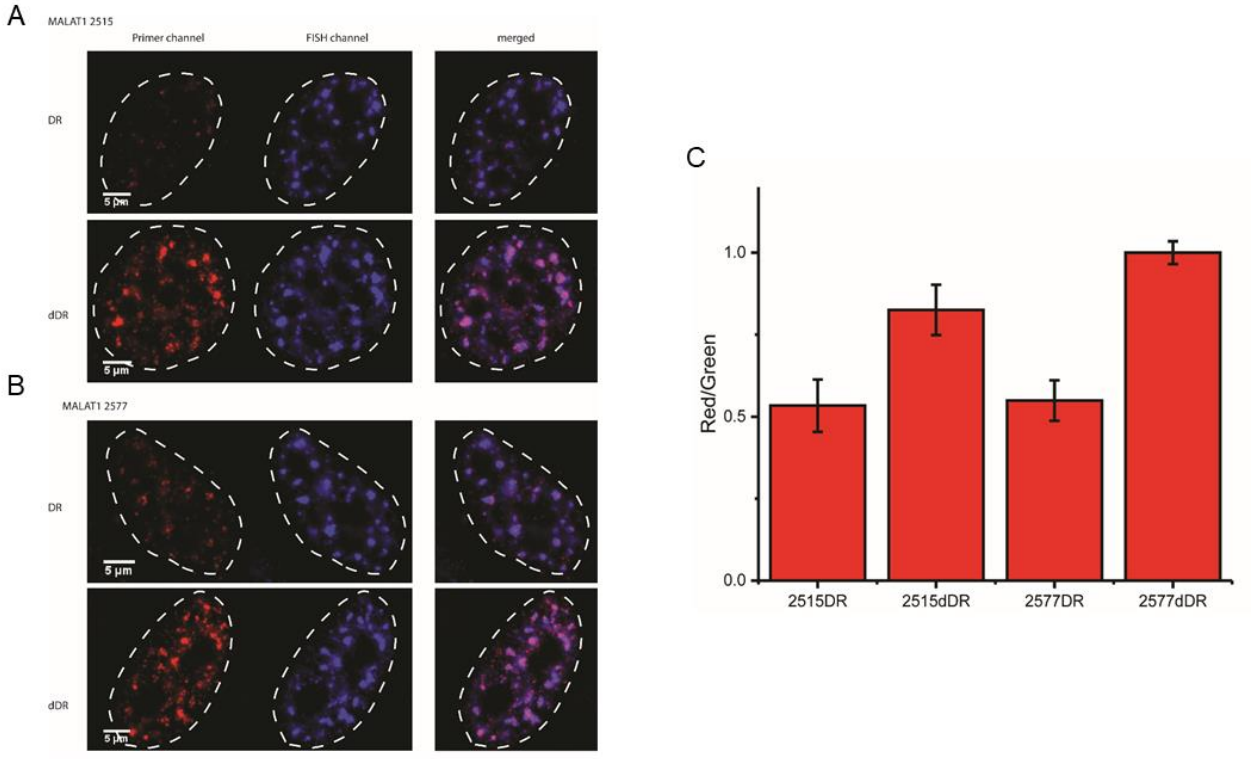


Figure 3.10. DR imaging assay on MALAT1 RNA. (A) & (B) Representative two color FISH images of MALAT1 2515 and 2577 sites after DR assay. (C) Average intensity of primer (red) versus FISH (green) channels after DR imaging assay. Scale bar = 5μm. Error bars indicate mean ± s.d. of 3 biological repeats.

3.3 Conclusion and discussion

In summary, here we present a method for quantifying the absolute methylation fraction of potential m^{6}A sites using a previously developed VMC10 DR[223], expanding the toolkit for site-specific quantification of m^{6}A. In addition, the method can be adjusted for high throughput quantification of m^{6}A sites. Furthermore, as the VMC10 DR selectively cleaves the unmodified A, it can potentially be used to discriminate other modifications, such as m^{1}A[225]. We, therefore, expect the DR-based quantification method can be easily applied to site-specific absolute quantifications of other RNA modifications. However, the DR cleavage activity varies among different RNAs in fixed cells. The cleavage efficiency of DR is compromised at the *ACTB* 1216 site, however, it can also digest highly m^{6}A modified *MALAT1* RNA.

While this method is easy to implement, there are several limitations that need to be considered. Firstly, the assay utilizes VMC10 DR, which has high cleavage efficiencies only on DGACH sequences, limiting its application on a subset of m^{6}A sites with the DAACH sequences[223]. Secondly, DR digestion efficiency varies among different sequences. Although low DR cleavage efficiency can be corrected by determining F_{DR} for each modification site of interest using *in vitro* transcribed RNA, low DR efficiency can lead to less accurate quantification due to two reasons. (1) A higher digestion efficiency leads to a larger $\Delta\Delta Ct$ that reduces the measurement variation by qPCR. Conversely, low digestion efficiency will make the $\Delta\Delta Ct$ too small to be accurately detected by qPCR. (2)

The <5% cleavage efficiency on the modified RNA can lead to underestimation of the m^6A fraction, and the percentage of underestimation depends on F_{DR} (**Figure S3.1**). A lower F_{DR} will result in a larger underestimation. When F_{DR} is 50%, a 5% cleavage of the modified RNA will result in a 10% underestimation of the m^6A. Finally, the presence of a nearby modified nucleotide may affect the DR cleavage efficiency depending on the type of the modification (**Figure 3.6**).

To improve the accuracy of the measurement, there are also a few factors to note. Firstly, for the synthetic RNA, we observed equal quality of F_m estimation using samples treated with dDR or samples lacking any DR as a negative control (**Figure 3.2C; Figure S3.5**). Nevertheless, we still recommend using dDR treated sample as a negative control, because it corrects for potential changes in the RNA secondary structure caused by DR binding that can affect RT efficiency. Secondly, we recommend using 60-mer DR for quantification, as 60-mer DR overall has higher digestion efficiencies of unmethylated RNAs potentially due to a higher hybridization efficiency. Thirdly, the quality of the primers used for RT and qPCR should be verified by performing calibration curves. Finally, we noticed that the largest source of technical variability in measurements originates from the RT step (comparing error bars in **Figure 3B and C**). We, thereby, recommend performing multiple RT reactions for each DR treated sample for a more accurate characterization.

DR based m^6A imaging in fixed cells includes multiple biochemical steps and the efficiency of each step contributes to the overall detection efficiency lineally. DR digestion

efficiency, as mentioned above, varies among different sequences. Ineffective wash of DR leads to compromised primer binding. Even though we have tested the wash conditions is stringent enough to wash out labeled DR (data not shown), the primer hybridization is not comparable to the no DR control (data not shown) with an unclear clue. In order to establish a digestion calibration curve with different percentages of m^6A modifications on GB RNAs, we modified GB RNAs with capping and polyadenylation and introduced the GB RNA into the cells through transfection. However, due to the low primer hybridization efficiency, the primer signal is barely observed from the images (data not shown). Another limitation is that the primer sequences cannot be easily modified to achieve an optimal melting temperature since they have to hybridize to a very close region to the m^6A sites of the target RNA, which may result in unspecific hybridization. For some cases, the unspecific signal can be washed away by RT or the wash step following RT, yet multiple primers are still required to be tested against one m^6A site.

3.4 Materials and Methods

Cell culture and RNA extraction

HeLa cells were cultured in DMEM medium (Gibco) supplemented with 10% (v/v) Fetal Bovine Serum (Thermo Fisher Scientific) and 100units/mL penicillin and 100μg/mL of streptomycin. The cells were grown at 37 ºC under humidified conditions with 5% CO_2. The total RNA was extracted using RNeasy Mini Kit (Qiagen) according to the

manufacturer's instructions.

In vitro transcription of endogenous RNA target fragments

The dsDNA templates for *in vitro* transcription were prepared by PCR with primers (Integrated DNA Technologies) that contain T7 promoter sequence and cDNA generated from total HeLa RNA. 1 µg of dsDNA templates were added to 100 µl reactions containing final concentrations of 2.5 mM each rNTP (New England Biolabs), 1x T7 Polymerase reaction buffer (New England Biolabs), 4 mM $MgCl_2$, 0.5 U/µl SUPERase-In RNase Inhibitor (Thermo Fisher Scientific), and 14 U/µl T7 polymerase (a kind gift from Dr. D. Bishop's Group). The reactions were incubated at 37 ºC for 1 hour. The transcript products were treated with DNase I recombinant (Roche) at 37 ºC for 30 min and purified by Phenol-chloroform extraction and ethanol precipitation. Primers are listed in **Table 3.2**.

In vitro transcription of 0% and 100% methylated GB RNA

A gene block containing 460 nt random sequence with 51% GC content and one adenosine was purchased from Genewiz. The gene block sequence is listed in Supplemental Table S1. The dsDNA template was amplified with primers containing an upstream T7 promoter sequence (**Table 3.2**). The *in vitro* reactions were carried out in the same conditions as for endogenous RNA targets, except that N^6-methyladenosine-5'-triphosphate (Trilink Biotechnologies) was used instead of rATP for the generation of 100% methylated RNA. The transcript products were treated with DNase I recombinant (Roche)

at 37 ºC for 30 min and purified by phenol-chloroform extraction, 7% denaturing polyacrylamide gel, and ethanol precipitation.

Synthesis of RNA oligo

Unmodified phosphoramidites were purchased from Glen Research. Phosphoramidite of N^6-methyladenosine was synthesized by following previously published procedure[213]. RNA oligos were synthesized using Expedite DNA synthesizer at 1 umol scale. After deprotection, RNA oligos were purified by PAGE.

Deoxyribozyme digestion in vitro

Total RNA (500 ng- 2 µg) or *in vitro* transcribed RNA fragments (50 nM) were mixed with 55.6 µM of either DR or dDR (Integrated DNA Technologies), 55.6 mM Tris-HCl (pH 7.5), and 166.7 mM NaCl in a final volume of 9 µL. The annealing of DR to the target site was facilitated by 5 min incubation at 95 ºC, followed by slow cooling to room temperature. After annealing, 1 µL of 200 mM $MgCl_2$ was added to each reaction and incubated at 37 ºC for 12 hours. The final concentrations of the reagents in the incubation buffer are 50 µM of DR, 50 mM Tris-HCl (pH 7.5), and 150 mM NaCl, and 20 mM $MgCl_2$ in 10 µL reaction. The DR treatment of 35-40 nt *MALAT1* 2515, *MALAT1* 2577, and *ACTB* 1216 was carried out following the same protocol, except with step-wise cooling (95 ºC for 5 min and 25 ºC for 10 min) instead of slow cooling. To remove the DR after the digestion, 1.33 µL of 10x TURBO DNase buffer and 2 µL of TURBO DNase (Thermo Fisher Scientific)

were added to the 10 µL DR reactions. The samples were incubated at 37 ºC for 2 hours. Subsequently, the DNase enzyme was inactivated by addition of EDTA (pH 7.5) to 15 mM final concentration and incubated at 75 ºC for 10 min. The DR sequences are listed in **Table 3.3**.

RNA digestion analyzed by PAGE

DR digestion reactions containing 50 ng to 100 ng of *in vitro* transcribed RNA or 35-40 nt synthetic RNAs were run on either 7% or 15% denaturing (7 M urea) polyacrylamide gel electrophoresis (PAGE), respectively. The gels were stained with SYBR Green II RNA Gel Stain (Thermo Fisher Scientific) for 10 min and imaged with ChemiDoc™ Imaging System (Bio-Rad). The cleavage efficiencies were analyzed with ImageJ using intensities of bands corresponding to the full-length RNA and the longer cleaved product.

Reverse transcription in vitro

For each DR treated sample, separate reverse transcription (RT) reactions were performed with gene-specific reverse primers for the m^6A region and internal reference site. Due to the presence of excess EDTA after DNase inactivation, the reactions were either significantly diluted or extra $MgCl_2$ was added for maximum reverse transcriptase activity. The RNA was denatured at 70 ºC for 5 min and then added to freshly prepared RT buffer with final concentration of 1 mM dNTPs (Thermo Fisher Scientific), 10% DMSO (Thermo Fisher Scientific), 10 mM DTT (Sigma-Aldrich), 250 nM of gene-specific reverse

primer (Integrated DNA Technologies), and 20-fold dilution of reverse transcriptase from iScript cDNA Synthesis Kit (Bio-Rad). The reactions were incubated at 25 ºC for 5 min, at 46 ºC for 20 min, and heat-inactivated at 95 ºC for 1 min. All primers are listed in **Table 3.4**.

qPCR

1 µL of cDNA was added into reaction mixture, containing 250 nM of each forward and reverse primers and 1x SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad) in a final volume of 20 µL. The qPCR reactions were performed with CFX real-time PCR system (Bio-Rad), using pre-incubation of 95 ºC for 30 s, followed by 40 cycles of 95 ºC for 10 s and 60 ºC for 30 s. The reactions were then subjected to melting curve analysis: 95 ºC for 10 s, 65 ºC for 5 s increment by 0.5 ºC to 95 ºC for 5 s. The data was analyzed with the supporting Bio-Rad CFX Maestro software. All primers are listed in Supplemental Table S4. All error bars in the figures are mean ± standard deviation (s.d.) of multiple biological replicates. For *in vitro* prepared GB RNA with different input m6A fraction, biological replicates are defined as independently mixed GB RNA samples. For the cases of endogenous mRNAs, biological replicates are defined as independently extracted total RNA samples. The m^6A fraction calculated for each biological replicate is from the average values of multiple technical replicates defined by independently performed RT reactions for each RNA sample.

Deoxyribozyme based imaging Assay

Fluorophore conjugation of primer and DNA probes

Same as Chapter 2

Cell fixation and permeabilization

Hela cells were seeded on 8-well imaging chamber (Cellvis) and grown to 60-80% confluency before fixation. Cells were washed with 1xPBS once (all wash steps are performed at room temperature for 3 min if not mentioned specifically) and incubated with fixation buffer at room temperature for 10 min. Cells were washed with 1xPBS twice. The solution was replaced by permeabilization buffer and the cells were incubated on ice for 10min. The cells were wash with 1xPBS 3 times.

DR hybridization and cleavage

Permeabilized cells were washed with 2xSSC once and 10% wash buffer once. DR was diluted to 500nM in hybridization buffer. The hybridization mix was added to the cells and the cells were incubated at 37 ºC for at least 6 h. Cells were washed with 10% wash buffer at 37 ºC for 30 min. Cells were washed with 2xSSC once and annealing buffer twice. The buffer was replaced with digestion buffer and the cells were incubated at 37 ºC for 12 h.

DR wash and primer hybridization

Cells were wash with 2xSSC twice and washed twice with 30% wash buffer at 42 ºC

for 30 min. Cells were washed with 2xSSC once and 10% wash buffer once. Alexa-647 labeled primer was diluted to 20 nM in hybridization buffer. The hybridization mix was added to the cells and the cells were incubated at 37 ºC for at least 6 h. Cells were washed with 10% wash buffer at 37 ºC for 30 min. Cells were then washed with 2xSSC once.

Reverse transcription in situ, primer wash and FISH probe rehybridization

Cells were washed with 1xRT buffer twice. 500 μM dNTPs (New England Biolabs), 10 mM DTT (Sigma-Aldrich) and 400 U of Superscript IV reverse transcriptase (Thermo Fisher Scientific) were added to 100 μL of 1xRT buffer. The buffer was added to the imaging chamber and incubated at 37 ºC overnight. The primer was washed with 30% wash buffer once at 42 ºC for 30 min. Cells were washed with 2xSSC twice and 10% wash buffer once. Each single Cy3B labeled FISH probes were diluted to 5-10 nM in hybridization buffer. The hybridization mix was added to the cells and the cells were incubated at 37 ºC for at least 6 h. Cells were washed with 10% wash buffer at 37 ºC for 30min followed by wash 2xSSC twice. The sample can be stored with 4xSSC at 4 ºC for 1 week.

Cell Imaging and image analysis

Imaging buffer was added to the cells before imaging. Epi Imaging was performed on Nikon TiE microscope with 100X NA 1.49 CFI HP TIRF objective. A 561nM laser (20% ~4.34×10^5 W/cm^2) was used for checking FISH channel and a 647nm laser (40% ~4.34

$\times 10^5$ W/cm^2) was used for checking primer signal. A 7-frame z-stack image with 300nm interval was taken on both channels. Colocalization analysis of background subtracted images was done on ImageJ with JACoP plugin. The threshold for both channels were set manually but kept the same for the same batch of experiment. M1 and M2 values were exported for statistical analysis. For primer and FISH signal quantification, the images were first background subtracted on both channels and a manual threshold was applied on FISH channel and a mask image was created. Average intensity values of the mask regions on both channels and the size of the mask regions were exported. The average ratio of primer and FISH signal of each image was calculated by $\sum_1^n(S_{pn} \times I_{pn}) / \sum_1^n(S_{fn} \times I_{fn})$, where S_{pn} is the area of primer signal, I_{pn} is average intensity of the primer signal, S_{fn} is the area of FISH signal and I_{fn} is the intensity of FISH signal.

3.5 Supplemental Information

Supplementary Tables

Table 3.1 Synthetic DNA and RNA sequences used in this chapter.

Name	Sequence
GB DNA template	GGTTGCGTTGGGTGTTCCTGTTTCTTTTGGCCTTTGTCTCTGTT TCTTTCCTTTCTCCTCCTTGTCGTGTCTGTTCGTTTCGTCGCTTT CCTCCTTCCTTGTTCTCGCTCGGACTCTTCTGGGCTCTTTCTGC GTTCGCCCTTCTTGTTCTCCCTTCTCTGTGGGTCCTGTTCTTGT GCTGGTTGTGCTCCCTCCTCTTGGTGCTCCTCCTTTCTGTGGC TGCGCTGGTGTTTCTTTCTCTCGGCTGCTCTGTTTGTTGGGTCT TTGTTGTGTGTTGTTGTTCTTGTGTGCTGCGTTTTGGTGGTGTC GGTTCTGTGCTGTCTTTCGGCCTGTCGTTTTCCTTCTCGTGTTC CGTCCTGTTTTGCGTGTCTCTCCCTGTGTTCCCGCTTTCCGTGT

Table 3.1, continued

	TGGCTGTGCTTGGTGTCTTTCGCTTGTTGGTTGGTCTCCTGTCT CCTGTGCTCGTCGGTCTTGTGG
41-nt synthetic *ACTB* 1216	CGCAAAUGCUUCUAGGCGGACUAUGACUUAGUUGCGUUACU
41-nt synthetic *MALAT1* 2515	AGUUUGAAAAAUGUGAAGGACUUUCGUAACGGAAGUAAUUU
32-nt synthetic *MALAT1* 2577	AACUUAAUGUUUUUGCAUUGGACUUUGAGUUA
(a) 35-nt synthetic A 2-nt A control 1	GCCUUGUUCUCGCUCGGACUAUUCUGGGCUCUUUC
(b) 35-nt synthetic A 2-nt m^6A	GCCUUGUUCUCGCUCGGACUm^6AUUCUGGGCUCUUUC
(c) 35-nt synthetic A 2-nt m^1A	GCCUUGUUCUCGCUCGGACUm^1AUUCUGGGCUCUUUC
(d) 35-nt synthetic A 2-nt U control	GCCUUGUUCUCGCUCGGACUUUUCUGGGCUCUUUC
(e) 35-nt synthetic A 2-nt ψ	GCCUUGUUCUCGCUCGGACUψUUCUGGGCUCUUUC
(f) 35-nt synthetic A 4-nt A control 1	GCCUUGUUCUCGCUCGGACUCUACUGGGCUCUUUC
(g) 35-nt synthetic A4-nt m^6A	GCCUUGUUCUCGCUCGGACUCUm^6ACUGGGCUCUUUC
(h) 35-nt synthetic A 4-nt m^1A	GCCUUGUUCUCGCUCGGACUCUm^1ACUGGGCUCUUUC
(i) 35-nt synthetic A 4-nt U control	GCCUUGUUCUCGCUCGGACUCUUCUGGGCUCUUUC
(j) 35-nt synthetic A 4-nt ψ	GCCUUGUUCUCGCUCGGACUCUψCUGGGCUCUUUC
(k) 35-nt synthetic A 2-nt A control 2	GCCUUGUUCUCGCUAGGACUCUUCUGGGCUCUUUC
(l) 35-nt synthetic m^6A 2-nt A	GCCUUGUUCUCGCUm^6AGGACUCUUCUGGGCUCUUUC
(m) 35-nt synthetic A 4-nt A control 2	GCCUUGUUCUCGAUCGGACUCUUCUGGGCUCUUUC
(n) 35-nt synthetic m^6A 4-nt A	GCCUUGUUCUCGm^6AUCGGACUCUUCUGGGCUCUUUC

Table 3.2 Primers used for generating templates for *in vitro* transcription.

Name	Sequence
Forward primer GB RNA DNA template	TAATACGACTCACTATAGGGTTGCGTTGGGTGTTCCTG
Reverse primer for GB RNA DNA template	CCACAAGACCGACGAGCACA
Forward primer for *ACTB* DNA template	TAATACGACTCACTATAGGCCAACACAGTGCTGTCTGGC
Reverse primer for *ACTB* DNA template	CTGCTGTCACCTTCACCGTTCC

Table 3.2, continued

Forward primer for *PLAC2* DNA template	TAATACGACTCACTATAGCAAGCAAAGTGAACACGTCG
Reverse primer for *PLAC2* DNA template	GTACTGACGTCGGCATCGAT
Forward primer for *MALAT1* DNA template	TAATACGACTCACTATAGGCTACTAAAAGGACTGGTGT
Reverse primer for *MALAT1* DNA template	TTCACCACCAAATCGTTAGC
Forward primer for *LY6K* DNA template	TAATACGACTCACTATAGGCAGGCCATACCACGCAGAAG
Reverse primer for *LY6K* DNA template	CCAAGACCCTGGGAAGTCAAA
Forward primer for *MCM5* DNA template	TAATACGACTCACTATAGGGAGATGCTGAGCCGCATC
Reverse primer for *MCM5* DNA template	CAGCAGGACACTACAGCTCC
Forward primer for *SEC11A* DNA template	TAATACGACTCACTATAGGGTCTGTGATTGGTGGAATGG
Reverse primer for *SEC11A* DNA template	AAGACTTACGACCACCTCAG
Forward primer for *INCENP* DNA template	TAATACGACTCACTATAGATAACCACACCCAGTGCCAG
Reverse primer for *INCENP* DNA template	TGCGGACAACACTTTCCTGT
Forward primer for *LMO7* DNA template	TAATACGACTCACTATAGGAAATGCTGCAGGACAGGGA
Reverse primer for *LMO7* DNA template	TGAGAGCCAAAGGGTCTTGG
Forward primer for *MRPL20* DNA template	TAATACGACTCACTATAGGCCGCTACTTTCGGATCCAGG
Reverse primer for *MRPL20* DNA template	GGCCATCCCTCATGTCTGTT

Table 3.3 Deoxyribozyme sequences used in this chapter.

Name	Sequence
GB RNA DR 40-mer	CCCAGAAGAGGGGTCTCCAGCTGGACGTTCGAGCGAGAAC
GB RNA dDR 40-mer	CCCAGAAGAGGGGTCTCCTCGTGGATTTCCGAGCGAGAAC
GB RNA dDR 60-mer	GCAGAAAGAGCCCAGAAGAGGGGTCTCCTCGTGGATTTCCGAGC GAGAACAAGGAAGGAG
GB RNA DR 60-mer	GCAGAAAGAGCCCAGAAGAGGGGTCTCCAGCTGGACGTTCGAG CGAGAACAAGGAAGGAG
PLAC2 m^{6}A 1 DR 40-mer	CCTCTGAGTGGGGTCTCCAGCTGGACGTTACTCCTGCCCC
PLAC2 m^{6}A 2 DR 40-mer	TGGGAAAATGGGGTCTCCAGCTGGACGTTCTGGGCAAGAG
PLAC2 m^{6}A 1 dDR 40-mer	CCTCTGAGTGGGGTCTCCTCGTGGATTTCACTCCTGCCCC
PLAC2 m^{6}A 2 dDR 40-mer	TGGGAAAATGGGGTCTCCTCGTGGATTTCCTGGGCAAGAG
PLAC2 m^{6}A 1 DR 60-mer	AGCGGAAGTGCCTCTGAGTGGGGTCTCCAGCTGGACGTTACTCC TGCCCCTTCTGTGCTT
PLAC2 m^{6}A 2 DR 60-mer	AAGGTGTGGCTGGGAAAATGGGGTCTCCAGCTGGACGTTCTGGG CAAGAGCGGAAGTGCC
PLAC2 m^{6}A 1 dDR 60-mer	AGCGGAAGTGCCTCTGAGTGGGGTCTCCTCGTGGATTTCACTCC TGCCCCTTCTGTGCTT
PLAC2 m^{6}A 2 dDR 60-mer	AAGGTGTGGCTGGGAAAATGGGGTCTCCTCGTGGATTTCCTGGG CAAGAGCGGAAGTGCC
ACTB 1165 DR 40-mer	CTCGTCATACGGGTCTCCAGCTGGACGTTCTGCTTGCTGA
ACTB 1165 dDR 40-mer	CTCGTCATACGGGTCTCCTCGTGGATTTCCTGCTTGCTGA
ACTB 1165 DR 60-mer	AGGGGCCGGACTCGTCATACGGGTCTCCAGCTGGACGTTCTGCT TGCTGATCCACATCTG
ACTB 1165 dDR 60-mer	AGGGGCCGGACTCGTCATACGGGTCTCCTCGTGGATTTCCTGCTT GCTGATCCACATCTG
ACTB 1216 DR 60-mer	GTAACGCAACTAAGTCATAGGGGTCTCCAGCTGGACGTTCGCCTA GAAGCATTTGCGGTG
ACTB 1216 dDR 60-mer	GTAACGCAACTAAGTCATAGGGGTCTCCTCGTGGATTTCCGCCTA GAAGCATTTGCGGTG
MALAT1 2515 DR 60-mer	AATTACTTCCGTTACGAAAGGGGTCTCCAGCTGGACGTTCTTCAC ATTTTTCAAACTAAG
MALAT1 2515 dDR 60-mer	AATTACTTCCGTTACGAAAGGGGTCTCCTCGTGGATTTCCTTCACA TTTTTCAAACTAAG
MALAT1 2577 DR 60-mer	AAAATAATCTTAACTCAAAGGGGTCTCCAGCTGGACGTTCAATGCA AAAACATTAAGTTG
MALAT1 2577 dDR 60-mer	AAAATAATCTTAACTCAAAGGGGTCTCCTCGTGGATTTCCAATGCA AAAACATTAAGTTG

Table 3.3, continued

MALAT1 2611 DR 60-mer	CAGCTGTCAATTAATGCTAGGGGTCTCCAGCTGGACGTTCTCAGG ATTTAAAAAATAATC
MALAT1 2611 dDR 60-mer	CAGCTGTCAATTAATGCTAGGGGTCTCCTCGTGGATTTCCTCAGG ATTTAAAAAATAATC
LY6K DR 60-mer	GAAGGCTCAGTCTGTGGCAGGGGTCTCCAGCTGGACGTTCCGTG GCTCAAGACAGGCTGA
LY6K dDR 60-mer	GAAGGCTCAGTCTGTGGCAGGGGTCTCCTCGTGGATTTCCCGTG GCTCAAGACAGGCTGA
MCM5 DR 60-mer	CAGAGGTCCCAGCAACATTGGGGTCTCCAGCTGGACGTTAATGG CAGGCAGCGGCAGGAG
MCM5 dDR 60-mer	CAGAGGTCCCAGCAACATTGGGGTCTCCTCGTGGATTTCAATGGC AGGCAGCGGCAGGAG
SEC11A DR 60-mer	GCTGCATTTTCATTTACAAGGGGTCTCCAGCTGGACGTTTCTGTAG GCACTTTAGAAGTG
SEC11A dDR 60-mer	GCTGCATTTTCATTTACAAGGGGTCTCCTCGTGGATTTCTCTGTAG GCACTTTAGAAGTG
INCENP 912 DR 60-mer	CTTAGACGCAGACCGCCCCGGGGTCTCCAGCTGGACGTTCCGAC CCCTTGACCCTTGGGG
INCENP 912 dDR 60-mer	CTTAGACGCAGACCGCCCCGGGGTCTCCTCGTGGATTTCCCGAC CCCTTGACCCTTGGGG
INCENP 967 DR 60-mer	AATCTGGAAAGGCTGGCGAGGGGTCTCCAGCTGGACGTTCCGTG GGCCAGGGGAGACCTG
INCENP 967 dDR 60-mer	AATCTGGAAAGGCTGGCGAGGGGTCTCCTCGTGGATTTCCCGTG GGCCAGGGGAGACCTG
INCENP 1060 DR 60-mer	TGTGCCGCACCGATTGAGAGGGGTCTCCAGCTGGACGTTCGTGC GAGAGCCCGTGGGCGT
INCENP 1060 dDR 60-mer	TGTGCCGCACCGATTGAGAGGGGTCTCCTCGTGGATTTCCGTGC GAGAGCCCGTGGGCGT
LMO7 DR 60-mer	GAATTTCAGTTGTTACACGGGGGTCTCCAGCTGGACGTTCTCTCT TTTTGCAAAAGTGGT
LMO7 dDR 60-mer	GAATTTCAGTTGTTACACGGGGGTCTCCTCGTGGATTTCCTCTCTT TTTGCAAAAGTGGT
MRPL20 DR 60-mer	CCTAATCAATACAGCAACAGGGGTCTCCAGCTGGACGTTCTCAGT GGTACTGCACCACTC
MRPL20 dDR 60-mer	CCTAATCAATACAGCAACAGGGGTCTCCTCGTGGATTTCCTCAGT GGTACTGCACCACTC

Table 3.4 Primers for reverse transcription and qPCR used in this chapter

Name	Sequence
Forward primer GB RNA m^{6}A region	GGTTGCGTTGGGTGTTCCTG
Reverse primer for GB RNA m^{6}A region	GGGAGAACAAGAAGGGCGAA
Forward primer for GB RNA control region	CGTCCTGTTTTGCGTGTCTC
Reverse primer for GB RNA control region	CCACAAGACCGACGAGCACA
Forward primer for *ACTB* m^{6}A region	CCTTCCAGCAGATGTGGATC
Reverse primer for *ACTB* m^{6}A region	GCCATGCCAATCTCATCTTG
Forward primer for *ACTB* control region	CAGGATGCAGAAGGAGATCAC
Reverse primer for *ACTB* control region	CGATCCACACGGAGTACTTG
Forward primer for *PLAC2* m^{6}A region	AAGAGAAGCACAGAAGGGGC
Reverse primer for *PLAC2* m^{6}A region	ACGGCTTGGGCAAAGGTGTG
Forward primer for *PLAC2* control region	CAAGCAAAGTGAACACGTCG
Reverse primer for *PLAC2* control region	TCACTTTAACTTGCACTTTACTGC
Forward primer for *MALAT1* m^{6}A region	GGCAGAAGGCTTTTGGAAGAGT
Reverse primer for *MALAT1* m^{6}A region	CTGGGTCAGCTGTCAATTAATGC
Forward primer for *MALAT1* control region	CAGCAGCAGACAGGATTCCA
Reverse primer for *MALAT1* control region	TCCTATCTTCACCACGAACTGC
Forward primer for *LY6K* m^{6}A region	GGCCTCAGCCTGTCTTGA
Reverse primer for *LY6K* m^{6}A region	AATGCAACAGGTGACAACGG
Forward primer for *LY6K* control region	TGACTGTGCACCTTTGAGCA
Reverse primer for *LY6K* control region	ACCGAGAGAAGGCAATCACG
Forward primer for *MCM5* m^{6}A region	TCACTGGACTCATGGACTCG
Reverse primer for *MCM5* m^{6}A region	AAGTTCGAGGGCTGCAGT
Forward primer for *MCM5* control region	GAGCACAGCATCATCAAGGA
Reverse primer for *MCM5* control region	TGCATGCGATGCTGGATCT
Forward primer for *SEC11A* m^{6}A region	CAAAGCCCCCAGTGTTTGTA
Reverse primer for *SEC11A* m^{6}A region	CGTGCAGAGCTGCATTTTCAT
Forward primer for *SEC11A* control region	CACTCGAGGGGACTTTCAGT
Reverse primer for *SEC11A* control region	GGCTTTGGCTCAACCTTTTAAT
Forward primer for *INCENP* 912 and 967 m^{6}A region	TGAGCTCCCTGATGGCTACA
Reverse primer for *INCENP* 912 and 967 m^{6}A region	CTCCCGCCATGGAGAATCTG
Forward primer for *INCENP* 1060 m^{6}A region	CTCCCATCCTGCCGGATAAC
Reverse primer for *INCENP* 1060 m^{6}A region	TGGGCTAAGACTTGGGGACT
Forward primer for *INCENP* control region	CATCAGTGAGCGCCAGAATG
Reverse primer for *INCENP* control region	TGATGTCGGGATGCCCTG
Forward primer for *LMO7* m^{6}A region	GAGAGAGTAGAAGAGAAGGG
Reverse primer for *LMO7* m^{6}A region	CAAAGAGGCTGGGCTTTGTTC
Forward primer for *LMO7* control region	TCACGGAGCACACAAATGGA

Table 3.4, continued

Reverse primer for *LMO7* control region	TGAGAGCCAAAGGGTCTTGG
Forward primer for *MRPL20* m^6A region	GGGAAGGAACCTGAAGGCAT
Reverse primer for *MRPL20* m^6A region	TGCAAATTACTCTGTCTCTTTTCC
Forward primer for *MRPL20* control region	CCAAAGCCCGATACCTGAAGA
Reverse primer for *MRPL20* control region	GCTCCACCTGGCACTTAACTA

Table 3.5 List of oligos for DR imaging assay used in this chapter

Name	Sequence
ACTB 1165 primer	/5AmMC12/AGGGGCCGGACTCGTCATACT
ACTB 1216 primer	/5AmMC12/GTGTAACGCAACTAAGTCATAGT
MALAT1 2515 primer	/5AmMC6/ATCTTGAATTACTTCCGTTACGAAAG
MALAT1 2577 primer	/5AmMC6/TAACTCAAAGTCCAATGCAAAA
MALAT1 2546 primer	/5AmMC6/CCAATGCAAAAACATTAAGTTGGTAATTAC

Table 3.6 List of buffers for DR imaging assay used in this chapter

Name	Components
Fixation buffer	4% paraformaldehyde (Electron Microscopy Sciences) in 1xPBS
Permeabilization buffer	0.5xTriton-X (Thermo Fisher Sciences), 2mM Ribonucleoside vanadyl complexes (Sigma-Aldrich) in 1xPBS
10% wash buffer	10% formamide (Ambion) in 2xSSC
Hybridization buffer	10% formamide, 10% dextran sulfate (Sigma-Aldrich D8906) in 2xSSC
5x Annealing buffer	250 mM Tris-HCl, 750 mM NaCl pH 7.5
Digestion buffer	250 mM Tris-HCl, 750 mM NaCl, 20mM $MgCl_2$ pH 7.5
5x RT buffer	250 mM Tris-HCl, 375 mM KCl, 15 mM $MgCl_2$ pH 8
Imaging buffer	10 mM NaCl, 50mM Tris pH=8.5, 4% glucose in 2xSSC

Supplementary Figures

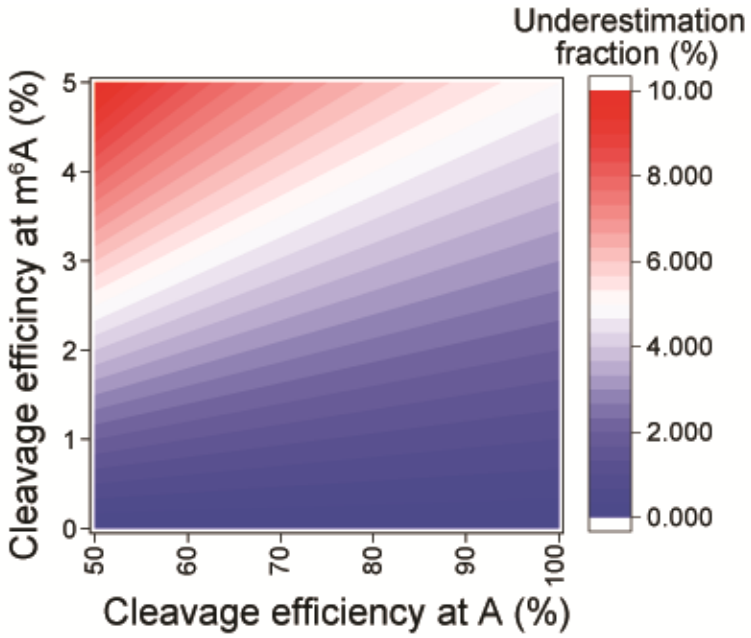


Figure S3.1. **Percentage of underestimation of m^{6}A fraction due to cleavage of DR on m^{6}A sequence.** The trace amount of cleavage of DR at m^{6}A containing sequence will cause false negative signal in the cleavage reaction and underestimation of the m^{6}A percentage. Considering DR cleavage efficiencies of the unmethylated A and m^{6}A sequence are F_{DR} and F'_{DR} respectively, and the true m^{6}A fraction is F'_m,

$$(1 - F'_m)F_{DR} + F_m{}'F'_{DR} = 1 - 2^{-\Delta\Delta Ct} \ ,$$

where $\Delta\Delta Ct$ is determined in Eq (2). Comparing F'_m to F_m from Eq (7) determined without the correction of F'_{DR}, the percentage of underestimation is

$$\frac{F'_m - F_m}{F'_m} = \frac{F'_{DR}}{F_{DR}} \ .$$

The heat map shows the percentage of underestimation as a function of F_{DR} and F'_{DR}, which increases when F_{DR} decreases, assuming F'_{DR} remains lower than 5%. When F_{DR} is 50% and F'_{DR} is 5%, the error in m^{6}A fraction is 10%.

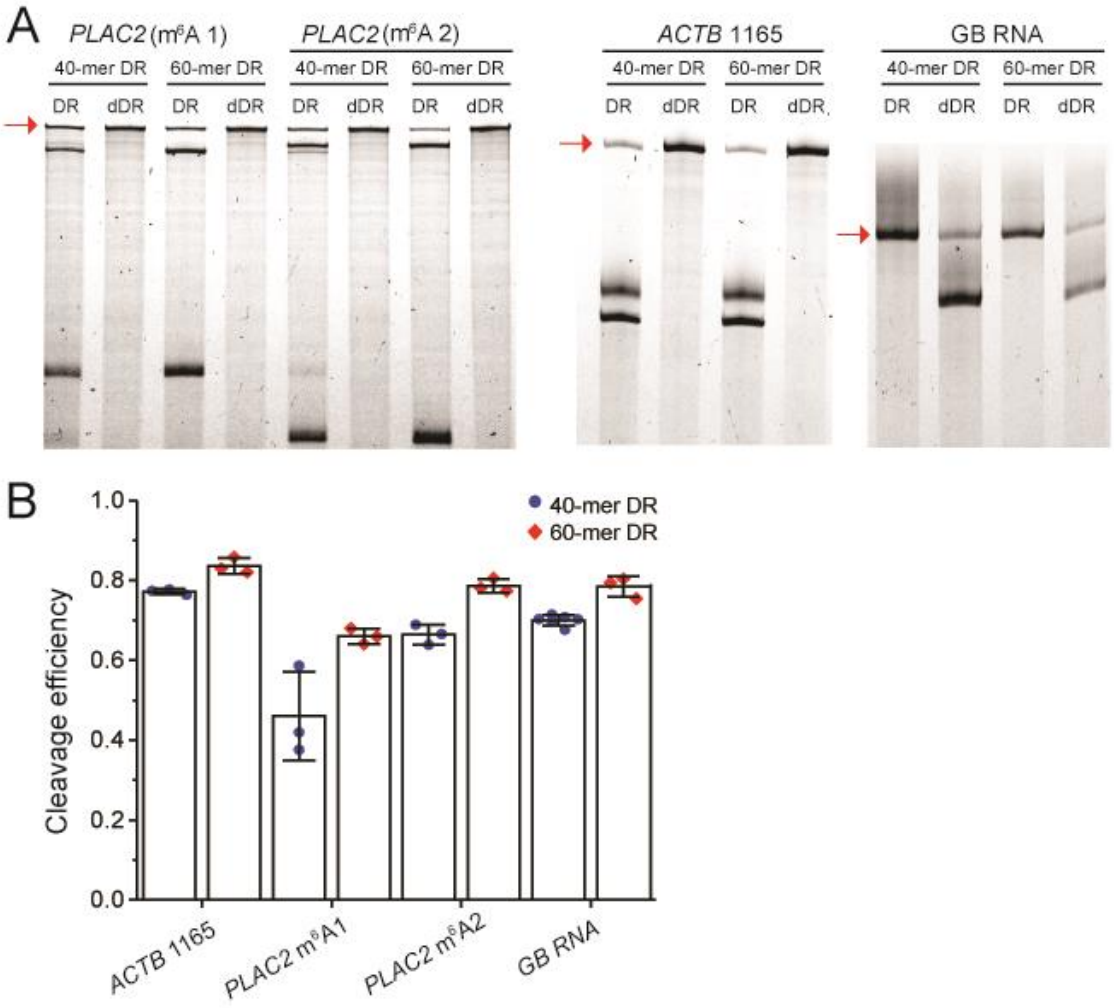


Figure S3.2. **Cleavage efficiencies (F_{DR}) of unmodified *in vitro* transcribed RNA by 40-mer and 60-mer DR and dDR.** (*A*) PAGE showing the DR cleavage of *PLAC2* m^{6}A1 and m^{6}A2 sites, *ACTB* 1165, and GB RNA by 40-mer and 60-mer DR. (*B*) Bar plot of the cleavage efficiencies of *PLAC2* m^{6}A1 and m^{6}A2, *ACTB* 1165, and the GB RNA by 40-mer and 60-mer DR as quantified from (*A*). Error bars indicate mean ± s.d. for 3 independent DR cleavage reactions. Red arrows point to full-length uncleaved RNA fragments.

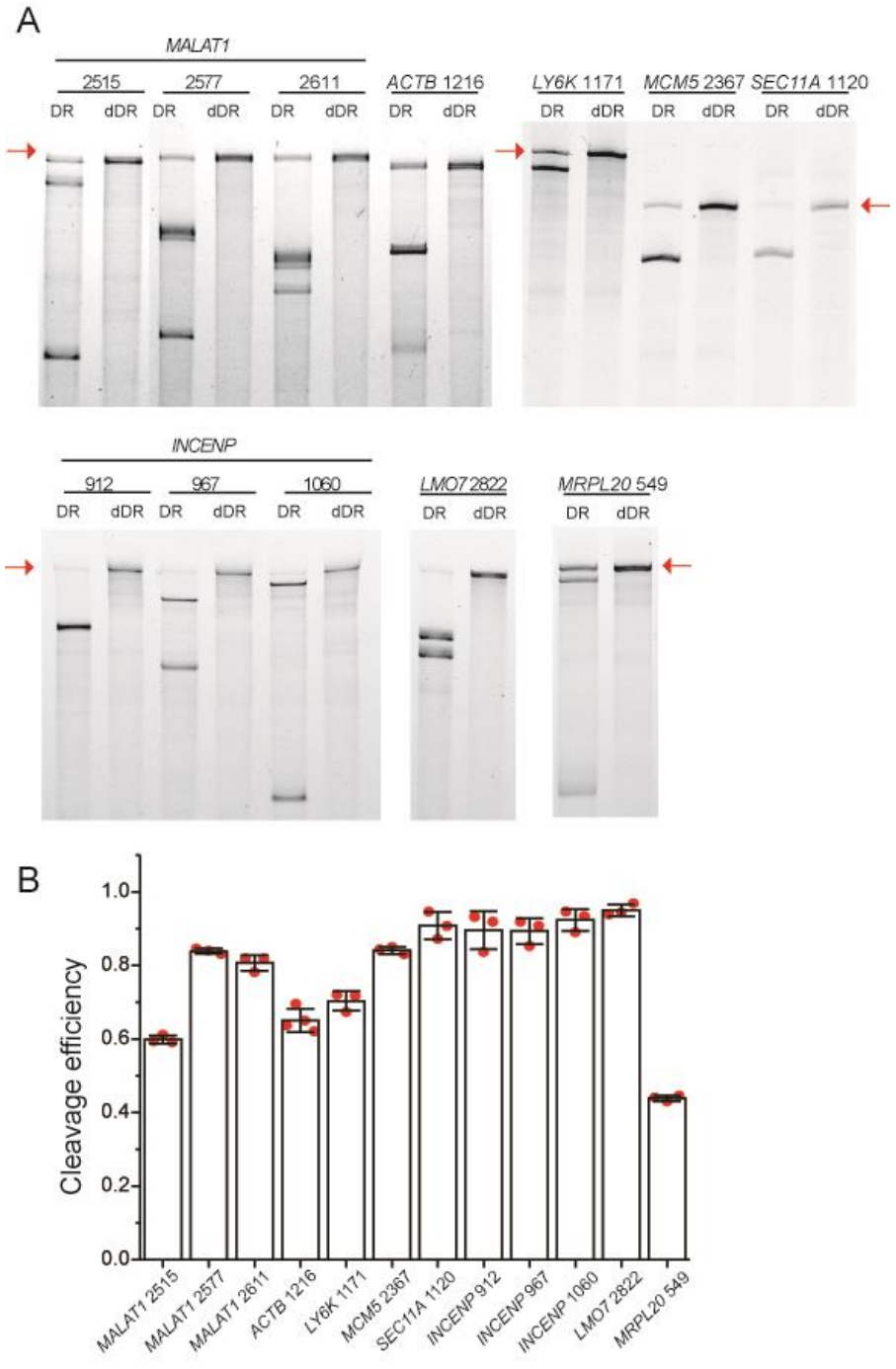


Figure S3.3. **Cleavage efficiencies (F_{DR}) of unmodified *in vitro* transcribed RNA by 60-mer DR and dDR**. (*A*) PAGE showing the DR cleavage of the seven endogenous targets: *MALAT1* 2515, 2577, and 2611, *ACTB* 1216, *LY6K* 1171, *MCM5* 2367, *SEC11A* 1120, *INCENP* 912, 967 and 1060, *LMO7* 2822, and *MRPL20* 549. (*B*) Bar plot of the cleavage efficiencies of endogenous targets as quantified from (*A*). Error bars indicate mean ± s.d. for 3 DR cleavage reactions. Red arrows point to full-length uncleaved RNA fragments.

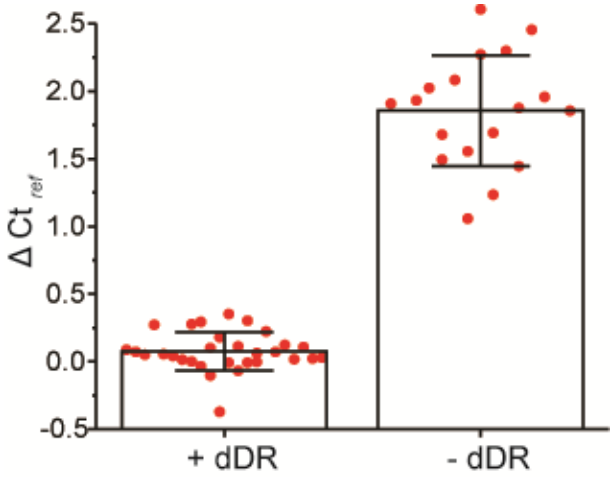


Figure S3.4. **Negative control without DR and with non-functional version of DR (dDR)**. When the negative control does not contain DR, there is a consistent difference between $Ct_{+DR-ref}$ and $Ct_{-DR-ref}$ ($\Delta\ Ct_{ref}$), with $Ct_{-DR-ref}$ being larger. Use of dDR in the negative control eliminates $\Delta\ Ct_{ref}$. Data comes from quantification of GB RNA with 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0 m^{6}A fraction input. Error bars indicate mean ± s.d.

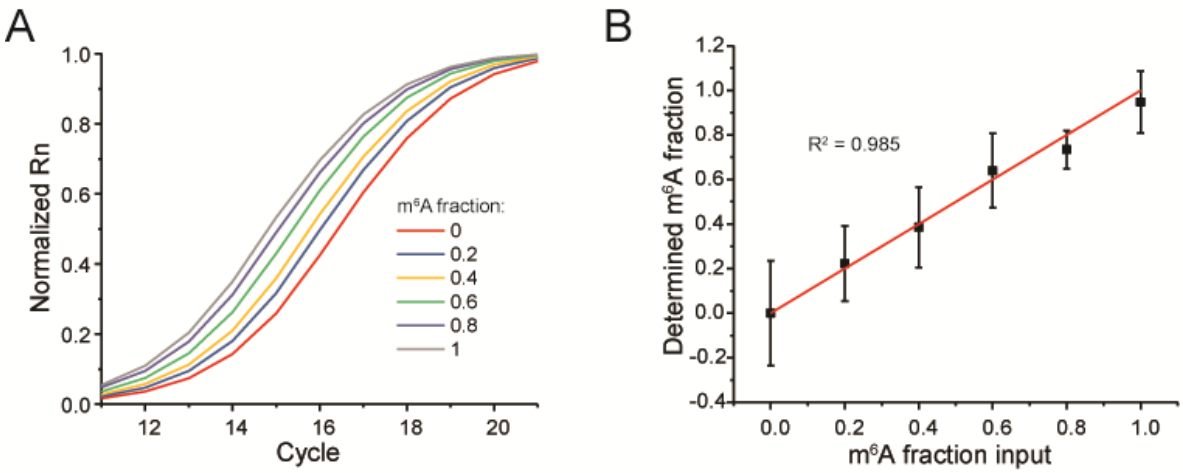


Figure S3.5. **Validation of the method for absolute quantification of m^{6}A fraction of the GB RNA without dDR.** (*A*) Normalized real-time fluorescence amplification curves for the DR cleaved synthetic RNAs with primers amplifying the m^{6}A site. (*B*) Estimated modification fraction as a function of input m^{6}A fraction for the GB RNA. Error bars indicate mean ± s.d. for 3 biological replicates.

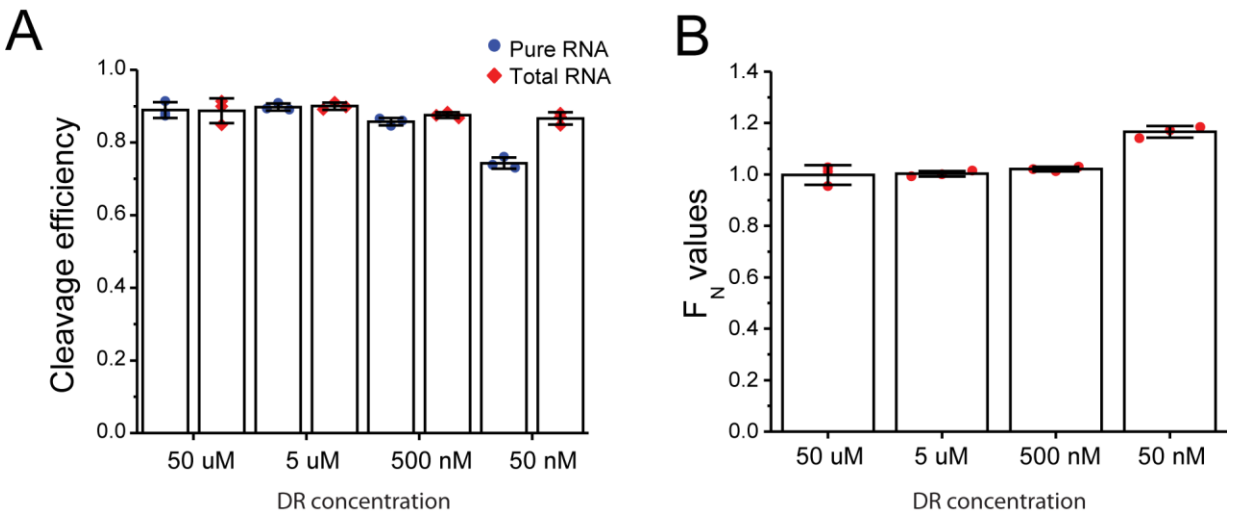


Figure S3.6. The dependence of cleavage efficiency on the concentration of DR. (A) The cleavage efficiencies of *PLAC2* m^{6}A 2 DR in presence and absence of total RNA with varying concentrations of DR. The concentration of the pure RNA was fixed at 50 nM, whereas the concentration of spike-in RNA in to the total RNA was fixed at 1 ng *PLAC2* RNA in 500 ng total RNA to better mimic the abundance of endogenous RNA. (B) F_N correction values for varying concentrations of *PLAC2* m^{6}A DR as determined from cleavage efficiencies in (A). All error bars report mean ± s.d. for 3 biological replicates. The cleavage efficiency of pure RNA is lower when using 50nM of DR, because the reaction was performed in the presence of 50 nM pure RNA, suggesting that the DR needs to be in over molar excess of the target RNA to ensure efficiency hybridization. The decrease in the cleavage efficiency for pure RNA results in a slightly larger than 1 F_N value at 50 nM DR concentration.

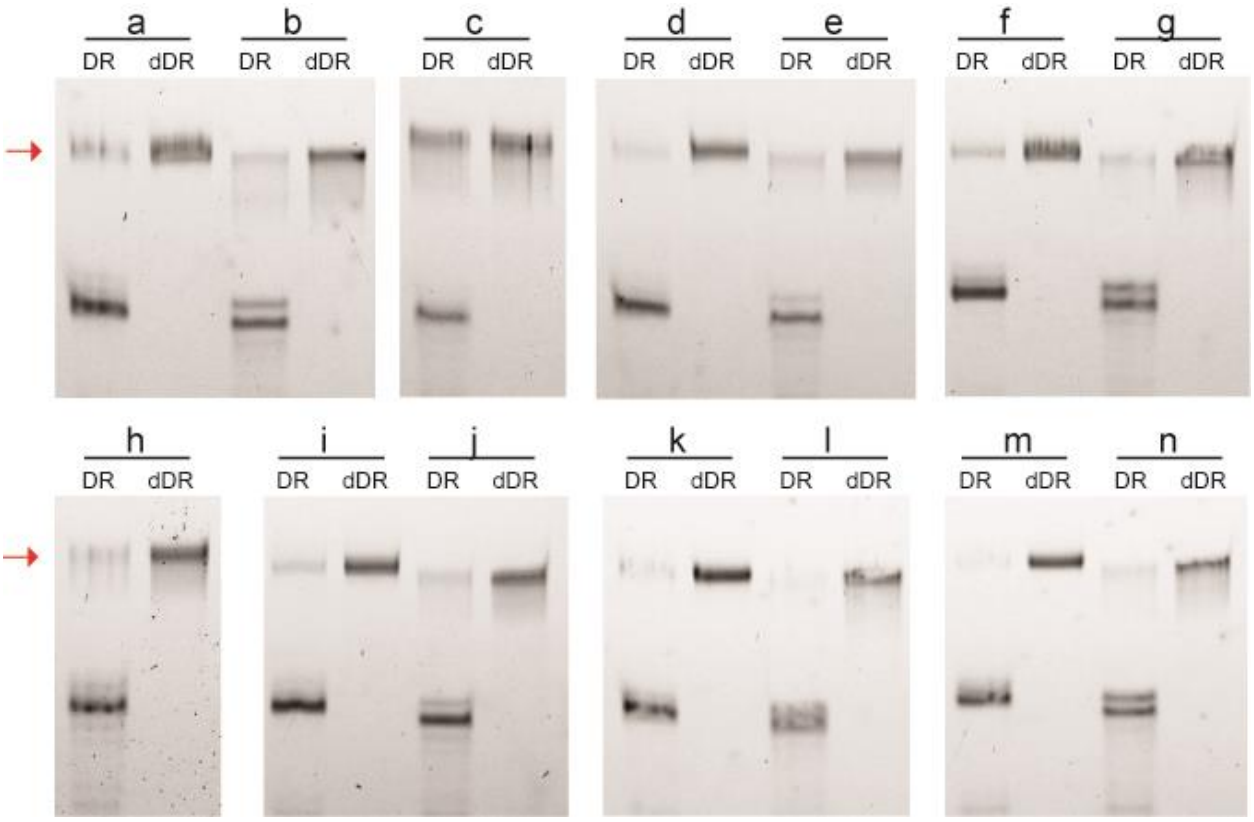


Figure S3.7. **PAGE showing the DR cleavage efficiency in absence and presence of nearby m^6A, m^1A, and ψ modifications.** Labels a-n correspond to 35-nt synthetic RNAs shown in Figure 3.6A and listed in **Table 3.1**. Red arrows point to full-length uncleaved RNA.

CHAPTER 4

Sequence dependence of pre-mRNA localization and splicing outcome

4.1 Introduction

As I have mentioned in Chapter 1.1.1, alternative splicing is regulated by both canonical slicing sequence and cis-acting splicing regulatory elements (SRE) and trans-acting splicing factor proteins and certain splicing factor protein have exhibit distinct localization in nuclear speckles. We have previously shown that the nuclear speckles are actually multilayered complex, in which spliceosome component U2 RNA is enriched at the periphery of the structure while proteins like SRSF2 are more concentrated inside the core[226] (**Figure 4.1**). We speculate that the localization of the splicing factor proteins can serve as a bridge between SRE sequences and splicing consequences. In other words, we argue that the sequence elements determine spatial organization which in turn determines splicing consequences. In this chapter, I will introduce our preliminary work on testing the hypothesis by massively parallel reporter assay (MPRA) and colocalization imaging.

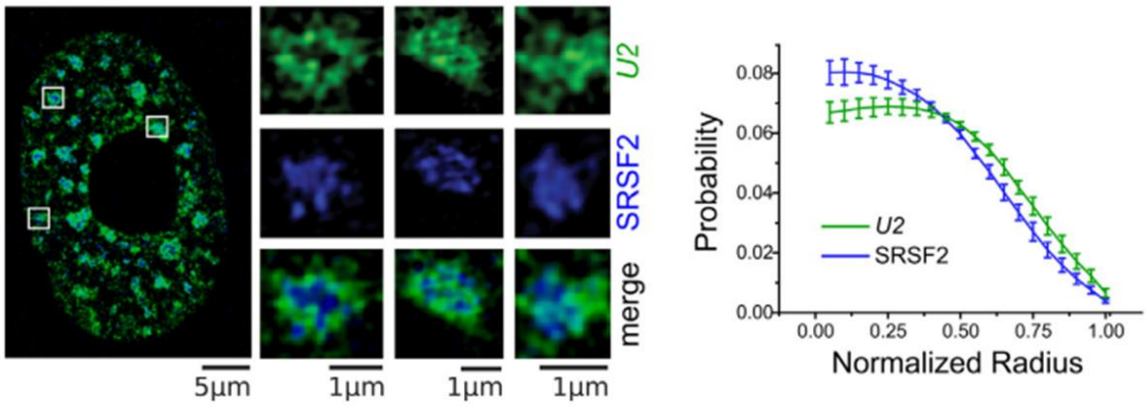


Figure 4.1. **Multilayer structure of the NS.** On the left, FISH image of *U2* and immunostaining of SRSF2 in U2OS cells. On the right, probability distribution of *U2* and SRSF2 inside the nuclear speckle. Images are adapted from Fei et al. 2017.

4.2 Results

4.2.1 Splicing outcome is determined by the balance between inclusion motifs and skipping motifs

In order to generate more SRE combinations to comprehensively evaluate the sequence effect on exon skipping which is a major form of alternative splicing, our collaborator Oded Regev at NYU performed massively parallel reporter assay (MPRA) that enables simultaneous measurement of $>10^6$ constructs. Over 350000 minigene constructs, each containing a different 70nt random sequence in the middle exon and a unique barcode in the last exon were generated as a pool library. The rest of the sequences, including promoter sequences, splicing core sequences, the first and last exons are identical in all the constructs. The library was transiently transfected in to HeLa cells and the RNAs were purified, reverse transcribed to amplicons. The amplicons were

sequenced on Illumina NextSeq platform. The results showed decent read coverage and consistency on exon inclusion/skipping among different biological replicates (**Figure 4.2**).

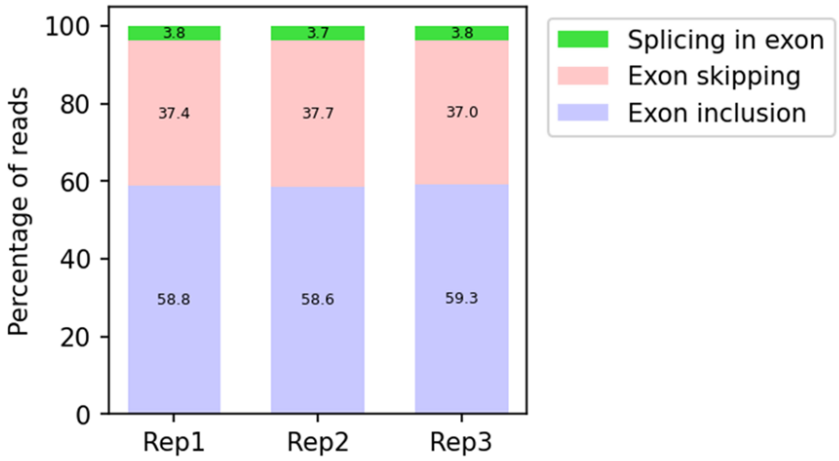


Figure 4.2. Splicing outcomes from RNA amplicon sequencing. Splicing outcomes of 3 biological replicates quantified by percentage of reads corresponded to exon skipping, exon inclusion and splicing in exon. The image was a courtesy from Oded.

The MPRA data was further used for training model using 'Adam' optimizer and the parameters of the model were tuned until the output matches the experimental outcomes. The model has identified a short list of ~20 motifs that affect splicing (**Figure 4.3A**). Remarkably, the identified motifs are in agreement with well-known RNA binding protein motifs and with the reported contribution to splicing: binding motifs of SR proteins appear as inclusion motifs and binding motifs of hnRNP proteins appear as skipping motifs. Ther splicing outcome is determined by the balance between the inclusion and skipping motifs. To be more specific, if the total strength of SR motifs is greater than that of hnRNP motifs, the exon is included and on the other wise, the exon is skipped (**Figure 4.3**B).

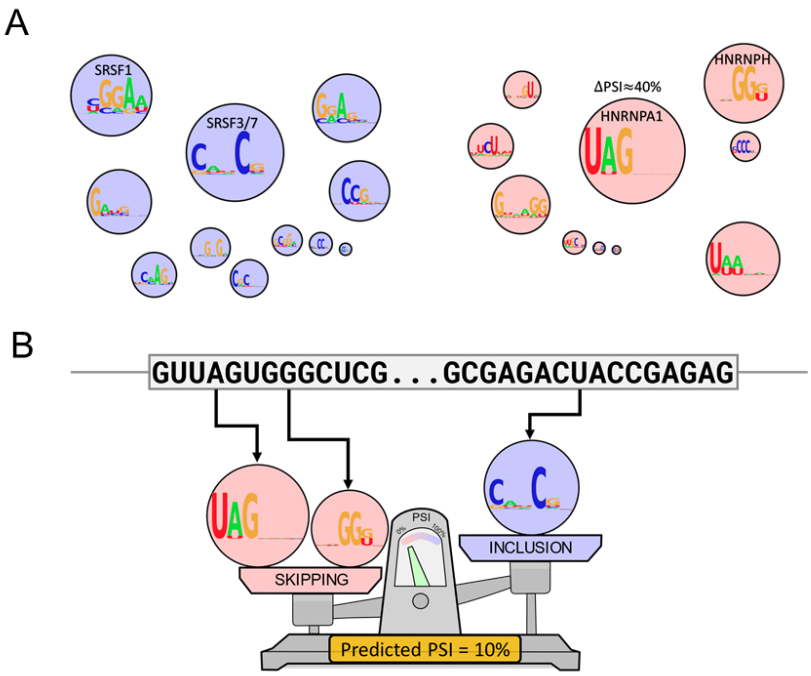


Figure 4.3. Splicing motifs detected by the machine learning model. (A) Splicing motifs detected by the machine learning model. (B) The splicing outcome is determined by the balance between the total strength of inclusion and skipping motifs. PSI (percent spliced in) indicates the fraction of reads showing exon inclusion. Images were a courtesy from Oded.

4.2.2 Sequence dependent RNA localization inside nuclear speckles

To test whether specific combinations of RNA motifs regulate RNA localization inside the cells, we constructed expression reporters containing tandem repeats of the motifs identified from the MPRA. We first picked SRSF1 and hnRNPA1 motifs and introduced these motifs into two different building blocks of the reporters with neutral reference sequences flanked by. No undesired RNA motifs were inserted at the rest of the sequences. Three different types of reporters were constructed first (**Figure 4.4A**). The expression of all the reporters is controlled by Tet-On promoter and same polyA tails are

added to the 3' end of the reporters. The type I reporter only contains the two building blocks connected by 5' splice site. In type II reporter, there are three exons bridged by two intron sequences. The SRSF1 and hnRNPA1 building blocks are on the second exon and intron respectively. In type III reporter, only the SRSF1 and hnRNPA1 position are replaced while the other sequences remain the same as that of the type II reporter. Fluorescent in situ hybridization (FISH) probes were designed at the tandem repeats of the RNA motifs.

We first transfected the type II reporter to HeLa cells that are constructed with Tet-On expression system (Tet-On HeLa) to check the RNA expression and splicing inside the cell since the type II reporter mimics pre-mRNA better compared with the type I reporter. The reporter was induced by doxycycline at different time points and we observed different RNA localization at different time points. At 30 min induction, the FISH signal showed good colocalization of both RNA motifs. The reporter RNAs formed granule and were associated with nuclear speckle protein marker (**Figure 4.4B**). At 2 h induction, the SRSF1 was more exported to the cytoplasm while the hnRNPA1 remained inside the nuclear speckles indicating the reporter pre-RNAs underwent splicing and mature RNAs were exported to the cytosol. Since we are more interested in the RNA localization during splicing, we decided to select 30 min as the induction time.

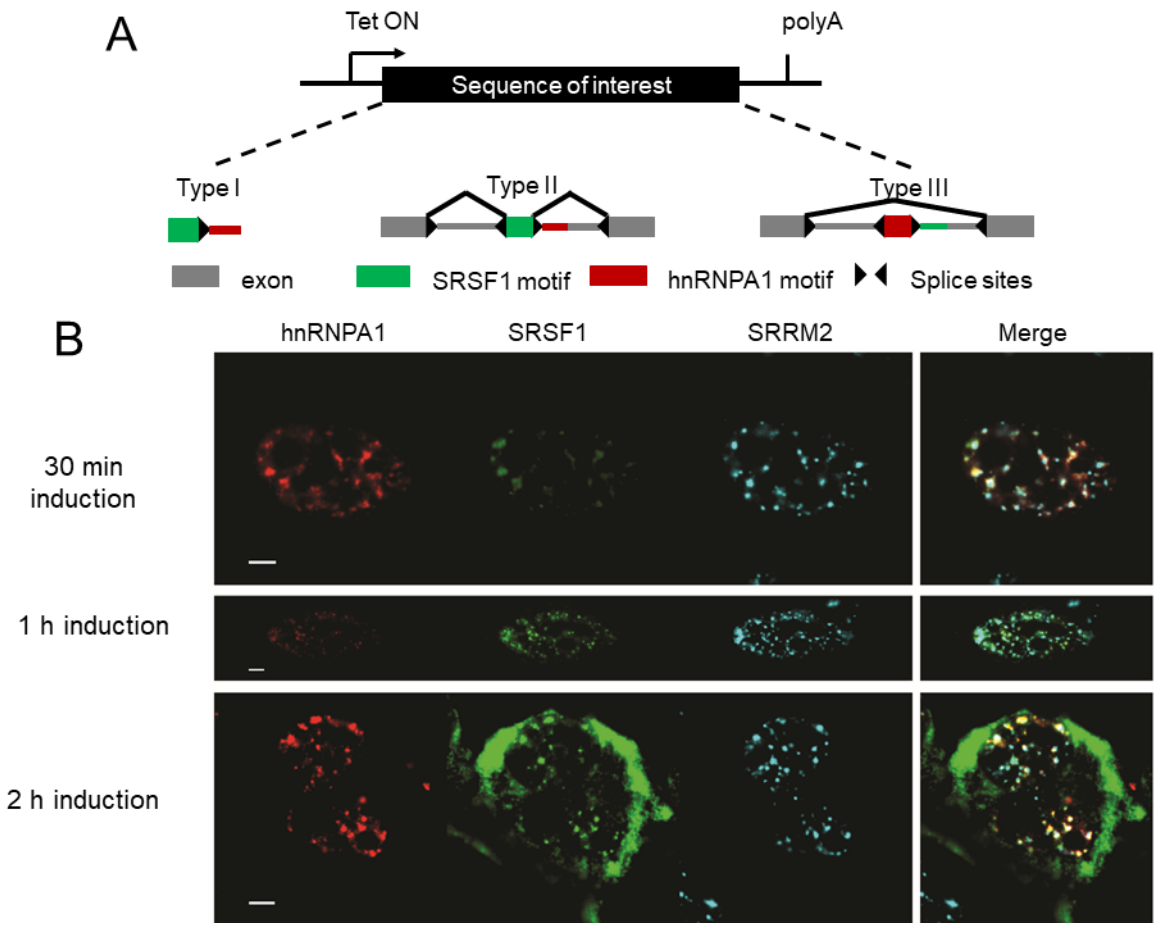


Figure 4.4. Reporter design and imaging results of type II reporter. (A) Cartoon scheme of the reporters. (B) Representative images of the type II reporter at different induction time. Scale bar = 5μm.

Then we transfected both the type I and type III reporters to the Tet-On HeLa cells separately. Interestingly, the type I reporter RNA could also form granules but they were not associated with nuclear speckles (**Figure 4.5A**). Since the type I reporter only contains 5' splice site of the core splicing sequences, this result indicates that the RNA motifs alone are not sufficient for nuclear speckle localization but core sequences are still required for mRNA association with nuclear speckles. However, for the type III reporters, the RNAs were distributed around the nucleoplasm with a few granules formed but not associated with nuclear speckle (**Figure 4.5A**). With 2 h induction, the type III reporter

RNA still localize in the nucleoplasm instead of being exported to cytosol (**Figure 4.5A**). The signal inside the nucleus can be either unspliced pre-mRNA or spliced mRNA lariat. We tested the splicing outcome of the type III reporter RNA by PCR and found out the second exon was skipped, which is different from type II reporter (**Figure 4.5B**). These results demonstrated that the RNA motifs can affect the pre-mRNA localization and the splicing outcome.

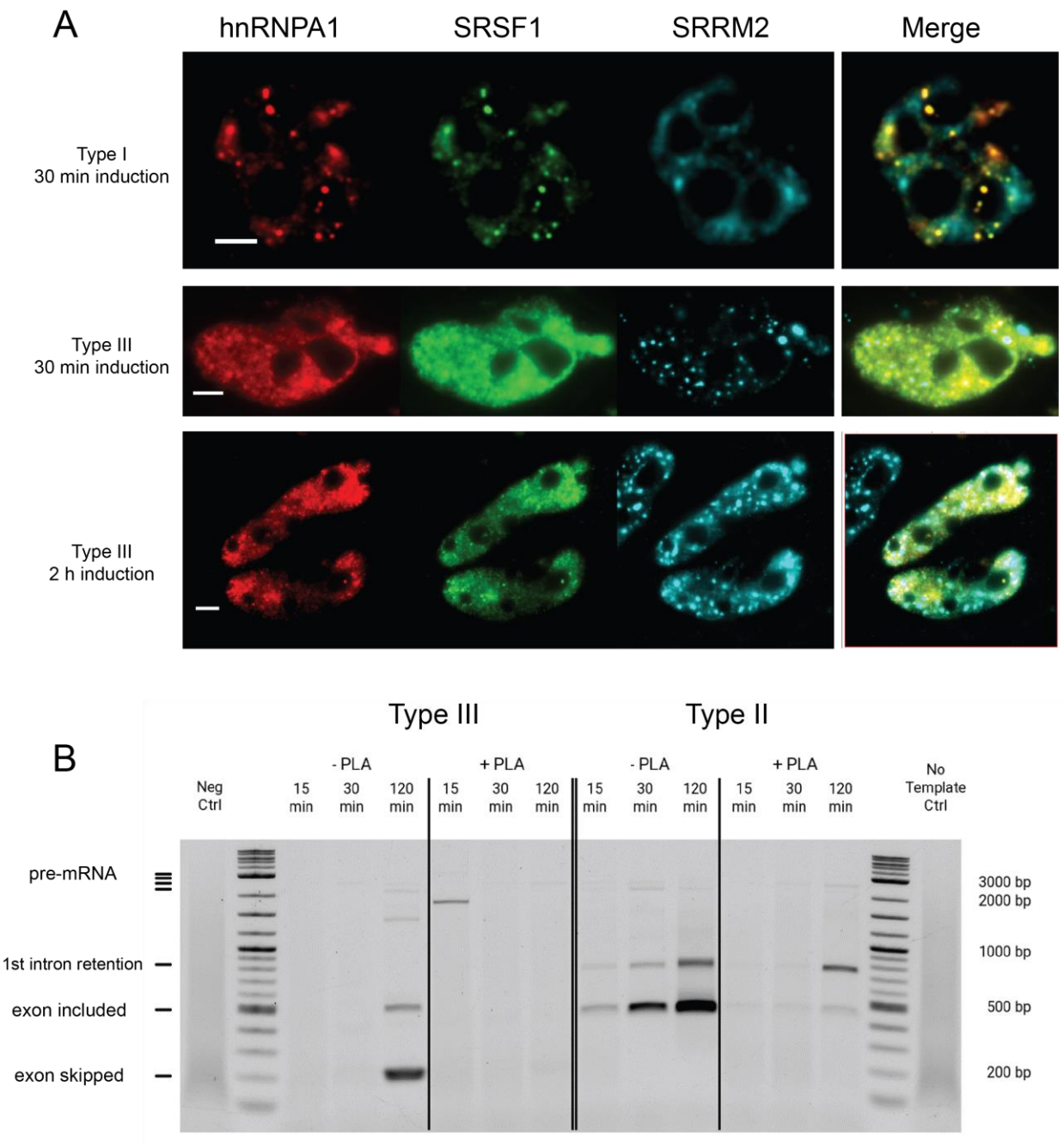


Figure 4.5. **Sequence dependence of splicing outcome.** (A) Representative images of type I and type III reporters at different induction time. Scale bar = 5μm (B) Gel electrophoresis result of the splicing outcome of Type III and Type II reporters at different induction time and the effect of pladienolide B (PLA). Figure B was a courtesy from Mauricio.

4.2.3 RNA localization is determined before splicing

The previous results have shown that pre-mRNAs that associated with nuclear speckle can be fully spliced. Next, we investigated what is the causation between speckle localization and RNA splicing. In order to do that, we added pladienolide B, an RNA splicing inhibitor that inhibits splicing factor 3B subunit (SF3B1) at early stage of splicing, to the cells expressing type II reporter RNAs. The expression of the reporter RNA was not affected by pladienolide B, but the RNAs could not be exported to the cytosol compared with the DMSO control, indicating no mature RNA was generated (**Figure 4.7A**). The PCR assay also confirmed the inhibited RNA splicing process (**Figure 4.5B**). The reporter RNA was more concentrated inside the nuclear speckles. Interestingly, the size of the nuclear speckles was dramatically larger than the control and this is possibly because the reporter RNAs could not get exported and accumulate inside the speckles while recruiting more protein to the speckles (**Figure 4.7B**). Since inhibition of the RNA splicing process does not affect the reporter RNA localization, the RNA localization should be determined before RNA splicing occurs.

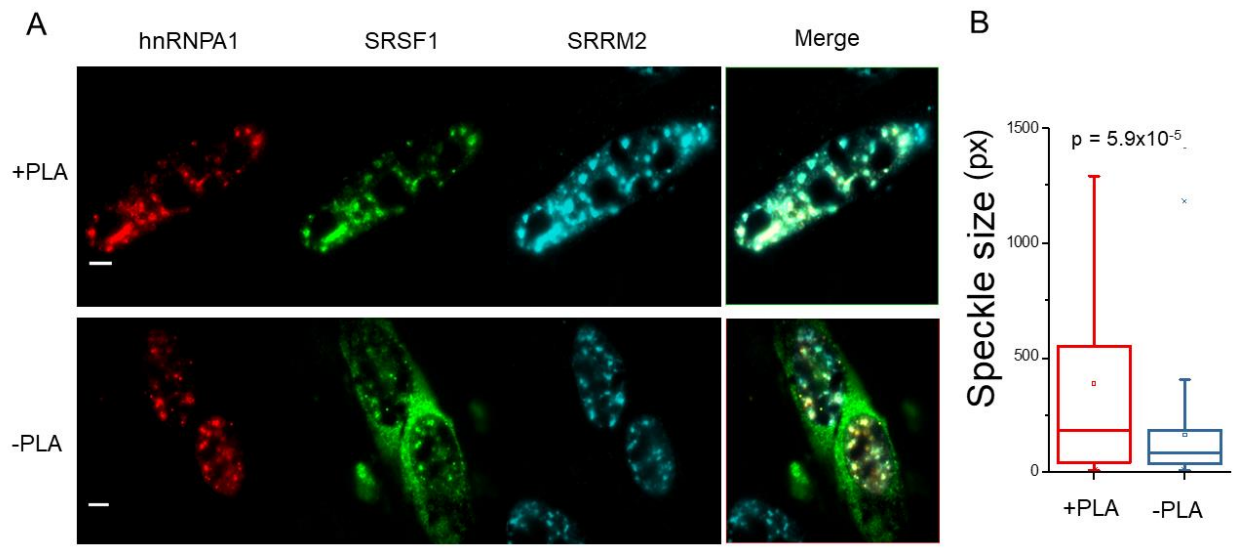


Figure 4.6. RNA localization inside NS does not require splicing. (A) Representative image of the type II reporter treated with or without pladienolide B (PLA). Scale bar = 5μm (B) Quantification of the speckle sizes.

4.2.4 SRSF1 and hnRNPA1 motifs exhibited different spatial distribution inside nuclear speckles

Since SRSF1 binding protein is enriched inside the nuclear speckle and hnRNPA1 protein are more likely to be in the nucleoplasm, we hypothesized that interaction with the RNA motif binding proteins can direct the reporter RNA sub-localization inside the nuclear speckle. To be more specific, the SRSF1 motif is expected to be more into the core of the speckle while the hnRNPA1 motif is expected to be around the periphery of the speckle. However, epi images cannot provide high resolution to visualize the spatial difference of RNA motifs within tens of nanometers. Therefore, we performed Stochastic Optical Reconstruction Microscopy (STORM) on the type II reporter and distance analysis to measure the normalized distance (ND) of the RNA motif to the center of the nuclear

speckle. Surprisingly, for RNA motifs with ND>0.5, the ND value of hnRNPA1 motif was significantly higher than that of SRSF1 (**Figure 4.6**), indicating that hnRNPA1 motif is more to the outside of the nuclear speckle. Considering that one of the spliceosome components, U2 snRNA, localizes at the periphery of the nuclear speckle (**Figure 4.1**), our super-resolution imaging results also indicate that the SRSF1 and hnRNPA1 motifs together with their binding proteins facilitate the RNA positioning within its splice sites around the periphery of the nuclear speckle and the RNA undergoes splicing at the interface of the nuclear speckle.

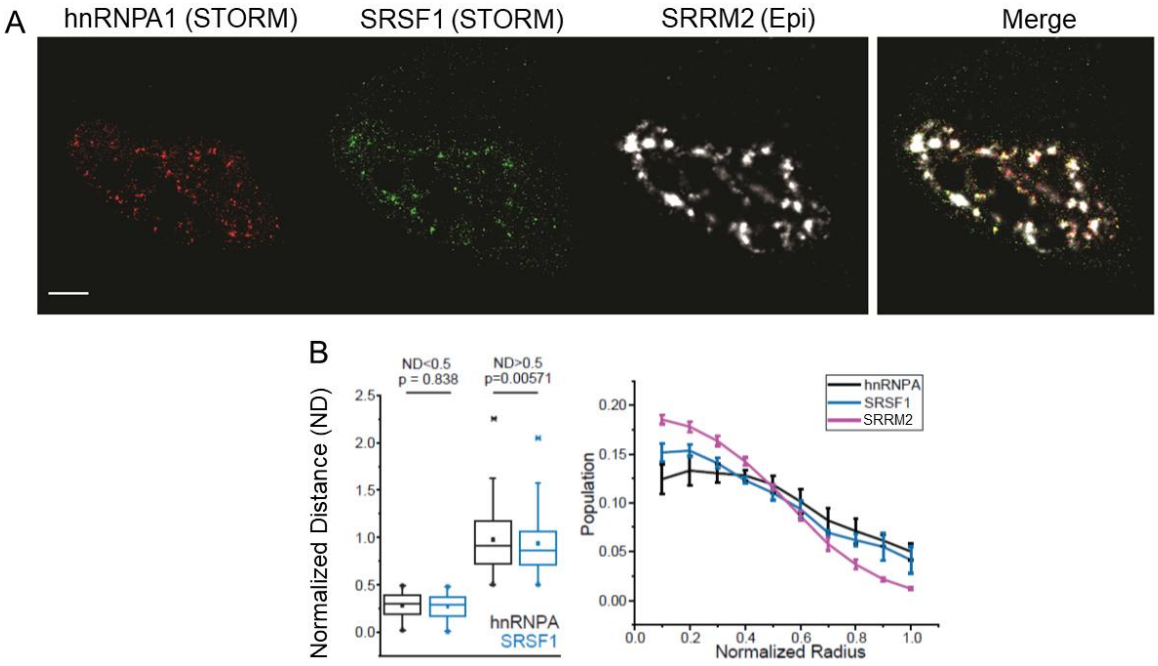


Figure 4.7. **Sub-localization of RNA motifs inside NS.** (A) Representative STORM image of the type II reporter. Scale bar = 5μm (B) Left, normalized distance of hnRNPA1 and SRSF1 motif to the center of the nuclear speckle; Right, population distribution of hnRNPA1 and SRSF1 and SRRM2 inside the nuclear speckle.

4.3 Conclusion and Discussion

In summary, we have created a model that identifies RNA motifs affecting splicing outcome and constructed two of the RNA motifs, SRSF1 and hnRNPA1 which have the opposite function on exon inclusion, to multiple reporters. With fluorescence microscopy, we have observed that the both core sequences and the RNA motifs are associated with RNA localization inside the nuclear speckles. The two RNA motifs are not positioned randomly in the nuclear speckle. The hnRNPA1 motif is more distributed on the periphery of the speckles whereas the SRSF1 motif is more enriched inside the speckle. By adding splicing inhibitor, we have shown that the RNA localization inside the nuclear speckle is determined by the sequence but not splicing event. The super-resolution imaging results demonstrated significant sub-localization difference between SRSF1 and hnRNPA1 motifs inside the NS.

We have observed RNA granules formation for all of the three types of reporters. Since type I and III have no association with nuclear speckle, the RNA granule formation may be caused by the tandem repeats of the RNA motifs or the active transcription at transcription sites. We will perform DNA FISH on the reporter plasmid to verify the association of the RNA granules with the transcription sites. Different strength of the RNA motifs would have different effect on RNA localization and the splicing outcome, therefore we will generate and test RNA reporters with different RNA motifs such as SRSF3 or SRSF7. The different sub-localization of the hnRNPA1 and SRSF1 motifs inside the

nuclear speckle is possibly driven by the RNA binding proteins. Since the spliceosome RNAs are at the periphery of the nuclear speckles, we speculate that the sub-localization of the RNA motif may indicate an interfacial splicing model, where the splicing factor proteins position the reporter RNA along the radius of the NS and bring the splicing site close to the interface of the NS, at which the splicing happens.

4.4 Materials and Methods

Fluorophore conjugation to secondary antibodies

26 μL of secondary antibodies was mixed with 3 μL of 1 M $NaHCO_3$ (pH 8.6). 1 μL of 1 μg/μL NHS modified fluorophore (Alexa-488, Thermo Fisher Scientific) dissolved in DMSO was added to the secondary antibody solution. The mixture was incubated at room temperature for 0.5-1 h. The conjugated antibody was passed through a P6 column (Bio-rad) that was pre-equilibrated with 1xPBS. The absorbance of the secondary antibody and the fluorophore was measured with spectrometer and the concentration and antibody/dye ratio was calculated based on lamber beer law.

Cell culture and transfection

Tet-ON HeLa cells (TaKaRa) were cultured in DMEM medium (Gibco) supplemented with 10% (v/v) Fetal Bovine Serum (Thermo Fisher Scientific) and 100units/mL penicillin and 100μg/mL of streptomycin. The cells were grown at 37 ºC under humidified conditions with 5% CO_2. Cells were seeded on 8-well imaging chamber 1 day before transfection. The confluency should reach 60-80% right before transfection. For transfection of each

well, 0.6μL of Lipofectamine® 3000 (Thermo Fisher Scientific) was diluted with 15 μL Opti-MeM (Thermo Fisher Scientific) and the mixture was vortexed. 200ng of plasmid DNA and 0.4 μL of P3000 was diluted together with 15 μL Opti-MeM and the mixture was vortexed. The two mixtures were mixed well and incubated at room temperature for 15 min. After the incubation, the culture medium was replaced in the chamber with pre-warmed Tet-free medium. The transfection mixture was added into the imaging chamber. After 6-8 h incubation, the medium was replaced with fresh Tet-free medium and incubate overnight.

Drug treatment and Induction

100nM Pladienolide B (Cayman) was added for at least 2 h before induction. For induction, 2 μg/mL doxycycline was added to the cells and cells were incubated for at least 30 min.

Fixation, permeabilization and FISH

Same as Chapter 3

Secondary antibody fluorophore conjugation

26 μL of secondary antibody was mixed with 3 μL 1M $NaHCO_3$ and 1μg NHS modified fluorophore. The mixture was incubated at room temperature for 30 min to 1 h followed by P6 column purification according to the manufacture instructions. The dye/antibody

label efficiency was measured by the spectrometer and calculated based on Lambert - Beer Law. Purified antibody was stored in -20 ºC

Immunostaining

FISH treated cells were post-fixed with 4% PFA in 2xSSC at room temperature for 10min. Cells were washed with 1xPBS twice. Then cells were incubated with blocking buffer at room temperature for 30 min. After blocking, primary antibody (Sigma S4045) was diluted 400 fold with blocking buffer and the mixture was added to the cells and incubatde at room temperature for 1 h. After primary antibody incubation, cells were washed with blocking buffer three times. Fluorophore conjugated secondary antibody was diluted 200 fold with blocking buffer and the mixture was added to the cells after wash. The incubation was done at room temperature for 1 h. The cells were further washed with 1xPBS three times and the sample is ready for imaging.

Cell Imaging and image analysis

Epi imaging was performed the same way as in Chapter 3.

2D STORM imaging was done on Nikon TiE microscope with 100X NA 1.49 CFI HP TIRF objective. Tetraspeck beads (Life Technologies) were diluted 400 fold with STORM imaging buffer and the mixture was added into the sample. Make sure the density of beads is optimal and replace the buffer with new STORM imaging buffer. STORM movies were taken on both Alexa-647 and CF-568 channels sequentially using JOBS module on

NIS software. For a brief summary, 647nm or 561nm laser was used to excited Alexa-647 and CF-568 fluorophore respectively. The acquisition was performed with 3 frames of 647/561 nm laser excitation followed by 1 frame of 405nm laser excitation. The laser power of 405nm laser was tuned during the acquisition to maintain a reasonable density of 'blinking-on' spots. The acquisition was stopped after a total number of 15000 frames of movie was taken. The collected images were exported to tiff files before analysis.

STORM image analysis was performed and reconstructed on ImageJ with ThunderStorm plugin. The reconstructed images (SRSF1 and hnRNPA1 FISH channel) were aligned with the SRRM channel and the distance analysis was performed using home written Matlab codes.

4.5 Supplementary Information

Supplementary Tables

Table 4.1. List of FISH probe sequences used in this chapter

Name	Sequences
hnRNPA1	/5AmMC6/CCCTAAttgtttggTTCCC
SRSF1	/5AmMC6/tcttgtttggcgttcctc

Table 4.2. List of buffers used in this chapter

Name	Components
Blocking buffer	0.1 mg/mL ultrapure BSA (Thermo Fisher Scientific) in 1xPBS
Tet-free medium	DMEM medium with 10% Tet system Approves FBS (TaKaRa)
STORM imaging buffer	10 mM NaCl, 50mM Tris pH=8.5, 10% glucose, 0.5mg/mL glucose oxidase and 67 μg/mL catalase in 2xSSC

References

1. Brown, T. A. *Genomes*. (Wiley-Liss, 2002).

2. von Hippel, P. H. An integrated model of the transcription complex in elongation, termination, and editing. *Science* **281**, 660–665 (1998).

3. Mackie, G. A. RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* **11**, 45–57 (2013).

4. Cheng, Z.-F. & Deutscher, M. P. An important role for RNase R in mRNA decay. *Mol. Cell* **17**, 313–318 (2005).

5. Yao, S., Sharp, J. S. & Bechhofer, D. H. Bacillus subtilis RNase J1 endonuclease and 5’ exonuclease activities in the turnover of DeltaermC mRNA. *RNA* **15**, 2331–2339 (2009).

6. Ramanathan, A., Robb, G. B. & Chan, S.-H. mRNA capping: biological functions and applications. *Nucleic Acids Res.* **44**, 7511–7526 (2016).

7. Bard, J. *et al.* Structure of yeast poly(A) polymerase alone and in complex with 3’-dATP. *Science* **289**, 1346–1349 (2000).

8. Bentley, D. Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* **11**, 347–351 (1999).

9. Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11–23 (2001).

10. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigó, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* **579**, 1900–1903 (2005).

11. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).

12. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).

13. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).

14. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55**, 1119–1150 (1986).

15. Cáceres, J. F. & Kornblihtt, A. R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186–193 (2002).

16. Woodley, L. & Valcárcel, J. Regulation of alternative pre-mRNA splicing. *Brief. Funct. Genomic. Proteomic.* **1**, 266–277 (2002).

17. Blencowe, B. J. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* **25**, 106–110 (2000).

18. Fairbrother, W. G., Yeh, R.-F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).

19. Singh, R. & Valcárcel, J. Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* **12**, 645–653 (2005).

20. Sanford, J. R., Ellis, J. & Cáceres, J. F. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem. Soc. Trans.* **33**, 443–446 (2005).

21. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).

22. Neugebauer, K. M. & Roth, M. B. Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.* **11**, 1148–1159 (1997).

23. Wansink, D. G. *et al.* Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.* **122**, 283–293 (1993).

24. Mattern, K. A., van der Kraan, I., Schul, W., de Jong, L. & van Driel, R. Spatial organization of four hnRNP proteins in relation to sites of transcription, to nuclear speckles, and to each other in interphase nuclei and nuclear matrices of HeLa cells. *Exp. Cell Res.* **246**, 461–470 (1999).

25. Ilık, İ. A. & Aktaş, T. Nuclear speckles: dynamic hubs of gene expression regulation. *FEBS J.* (2021) doi:10.1111/febs.16117.

26. Chen, Y. & Belmont, A. S. Genome organization around nuclear speckles. *Curr. Opin. Genet. Dev.* **55**, 91–99 (2019).

27. Brugiolo, M., Herzel, L. & Neugebauer, K. M. Counting on co-transcriptional splicing. *F1000Prime Rep.* **5**, 9 (2013).

28. Wang, K. *et al.* Intronless mRNAs transit through nuclear speckles to gain export competence. *J. Cell Biol.* **217**, 3912–3929 (2018).

29. Fan, J. *et al.* mRNAs are sorted for export or degradation before passing through nuclear speckles. *Nucleic Acids Res.* **46**, 8404–8416 (2018).

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

31. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* **14**, 1138–1152 (2017).

32. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28**, 395–404 (2018).

33. Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).

34. Cantara, W. A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195-201 (2011).

35. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).

36. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci USA* **71**, 3971–3975 (1974).

37. Meyer, K. D. & Jaffrey, S. R. Rethinking m6a readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.* **33**, 319–342 (2017).

38. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **3**, 1233–1247 (1997).

39. Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).

40. Wang, P., Doxtader, K. A. & Nam, Y. Structural basis for cooperative function of METTL3 and METTL14 methyltransferases. *Mol. Cell* **63**, 306–317 (2016).

41. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3’ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).

42. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).

43. Wang, X. *et al.* N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388–1399 (2015).

44. Meyer, K. D. *et al.* 5’ UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999–1010 (2015).

45. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).

46. Ke, S. *et al.* m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* **31**, 990–1006 (2017).

47. Du, H. *et al.* YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* **7**, 12626 (2016).

48. Bartosovic, M. *et al.* N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3’-end processing. *Nucleic Acids Res.* **45**, 11356–11370 (2017).

49. Zhang, S. *et al.* m6A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. *Cancer Cell* **31**, 591-606.e6 (2017).

50. Zhang, C. *et al.* Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m^6A-demethylation of NANOG mRNA. *Proc Natl Acad Sci USA* **113**, E2047-56 (2016).

51. Rottman, F., Shatkin, A. J. & Perry, R. P. Sequences containing methylated nucleotides at the 5’ termini of messenger RNAs: possible implications for processing. *Cell* **3**, 197–199 (1974).

52. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).

53. Zhang, Y. & Hamada, M. DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. *BMC Bioinformatics* **19**, 524 (2018).

54. Zhou, Y., Zeng, P., Li, Y.-H., Zhang, Z. & Cui, Q. SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.* **44**, e91 (2016).

55. Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300–304 (2006).

56. Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615–628 (2009).

57. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463 (2004).

58. Montange, R. K. & Batey, R. T. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37**, 117–133 (2008).

59. Babitzke, P. & Romeo, T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.* **10**, 156–163 (2007).

60. Brantl, S. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* **10**, 102–109 (2007).

61. Aiba, H. Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.* **10**, 134–139 (2007).

62. Holmqvist, E., Li, L., Bischler, T., Barquist, L. & Vogel, J. Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends. *Mol. Cell* **70**, 971-982.e6 (2018).

63. Prévost, K. *et al.* The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol. Microbiol.* **64**, 1260–1273 (2007).

64. Dutta, T. & Srivastava, S. Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. *Gene* **656**, 60–72 (2018).

65. Massé, E. & Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci USA* **99**, 4620–4625 (2002).

66. Morita, T., Maki, K. & Aiba, H. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* **19**, 2176–2186 (2005).

67. Darfeuille, F., Unoson, C., Vogel, J. & Wagner, E. G. H. An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell* **26**, 381–392 (2007).

68. Opdyke, J. A., Kang, J.-G. & Storz, G. GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J. Bacteriol.* **186**, 6698–6705 (2004).

69. Giangrossi, M. *et al.* A novel antisense RNA regulates at transcriptional level the virulence gene icsA of Shigella flexneri. *Nucleic Acids Res.* **38**, 3362–3375 (2010).

70. Henry, R. W., Mittal, V., Ma, B., Kobayashi, R. & Hernandez, N. SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. *Genes Dev.* **12**, 2664–2672 (1998).

71. Bohnsack, M. T. & Sloan, K. E. Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol. Chem.* **399**, 1265–1276 (2018).

72. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).

73. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).

74. Liu, J. Control of protein synthesis and mRNA degradation by microRNAs. *Curr. Opin. Cell Biol.* **20**, 214–221 (2008).

75. Baskerville, S. & Bartel, D. P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**, 241–247 (2005).

76. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).

77. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**, 376–385 (2005).

78. Maniataki, E. & Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* **19**, 2979–2990 (2005).

79. Gregory, R. I., Chendrimada, T. P., Cooch, N. & Shiekhattar, R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**, 631–640 (2005).

80. Medley, J. C., Panzade, G. & Zinovyeva, A. Y. microRNA strand selection: Unwinding the rules. *Wiley Interdiscip. Rev. RNA* **12**, e1627 (2021).

81. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).

82. Brodersen, P. *et al.* Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185–1190 (2008).

83. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79**, 351–379 (2010).

84. Schnall-Levin, M. *et al.* Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. *Genome Res.* **21**, 1395–1403 (2011).

85. Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci USA* **95**, 13959–13964 (1998).

86. Mello, C. C. & Conte, D. Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).

87. Golden, D. E., Gerbasi, V. R. & Sontheimer, E. J. An inside job for siRNAs. *Mol. Cell* **31**, 309–312 (2008).

88. Tomari, Y. & Zamore, P. D. MicroRNA biogenesis: drosha can't cut it without a partner. *Curr. Biol.* **15**, R61-4 (2005).

89. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).

90. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M. C. Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes Dev.* **19**, 2837–2848 (2005).

91. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).

92. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).

93. Orban, T. I. & Izaurralde, E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**, 459–469 (2005).

94. Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364–370 (2004).

95. Fisher, J. K. *et al.* Four-dimensional imaging of E. coli nucleoid organization and dynamics in living cells. *Cell* **153**, 882–895 (2013).

96. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol. Microbiol.* **85**, 21–38 (2012).

97. Miller, O. L., Hamkalo, B. A. & Thomas, C. A. Visualization of bacterial genes in action. *Science* **169**, 392–395 (1970).

98. Chai, Q. *et al.* Organization of ribosomes and nucleoids in Escherichia coli cells during growth and in quiescence. *J. Biol. Chem.* **289**, 11342–11352 (2014).

99. Gowrishankar, J. & Harinarayanan, R. Why is transcription coupled to translation in bacteria? *Mol. Microbiol.* **54**, 598–603 (2004).

100. Buskila, A. A., Kannaiah, S. & Amster-Choder, O. RNA localization in bacteria. *RNA Biol.* **11**, 1051–1060 (2014).

101. Libby, E. A., Roggiani, M. & Goulian, M. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proc Natl Acad Sci USA* **109**, 7445–7450 (2012).

102. Fei, J. & Sharma, C. M. RNA localization in bacteria. *Microbiol. Spectr.* **6**, (2018).

103. Weil, T. T., Parton, R. M. & Davis, I. Making the message clear: visualizing mRNA localization. *Trends Cell Biol.* **20**, 380–390 (2010).

104. Sadowski, P. G., Groen, A. J., Dupree, P. & Lilley, K. S. Sub-cellular localization of membrane proteins. *Proteomics* **8**, 3991–4011 (2008).

105. Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490.e26 (2019).

106. Handwerger, K. E. & Gall, J. G. Subnuclear organelles: new insights into form and function. *Trends Cell Biol.* **16**, 19–26 (2006).

107. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).

108. Latonen, L. Phase-to-Phase With Nucleoli - Stress Responses, Protein Aggregation and Novel Roles of RNA. *Front. Cell. Neurosci.* **13**, 151 (2019).

109. Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* **45**, 10350–10368 (2017).

110. Engreitz, J. M. *et al.* RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).

111. Sanford, J. R. *et al.* Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).

112. Miyagawa, R. *et al.* Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* **18**, 738–751 (2012).

113. Dias, A. P., Dufu, K., Lei, H. & Reed, R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.* **1**, 97 (2010).

114. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251-7 (2016).

115. Jung, H., Yoon, B. C. & Holt, C. E. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat. Rev. Neurosci.* **13**, 308–324 (2012).

116. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).

117. Colman, D. R., Kreibich, G., Frey, A. B. & Sabatini, D. D. Synthesis and incorporation of myelin polypeptides into CNS myelin. *J. Cell Biol.* **95**, 598–608 (1982).

118. Smith, R. Moving molecules: mRNA trafficking in Mammalian oligodendrocytes and neurons. *Neuroscientist* **10**, 495–500 (2004).

119. Han, J. W., Park, J. H., Kim, M. & Lee, J. mRNAs for microtubule proteins are specifically colocalized during the sequential formation of basal body, flagella, and cytoskeletal microtubules in the differentiation of Naegleria gruberi. *J. Cell Biol.* **137**, 871–879 (1997).

120. Lécuyer, E. *et al.* Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187 (2007).

121. Paredes, E., Evans, M. & Das, S. R. RNA labeling, conjugation and ligation. *Methods* **54**, 251–259 (2011).

122. Reese, C. B. Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Org. Biomol. Chem.* **3**, 3851–3868 (2005).

123. Rinaldi, A. J., Suddala, K. C. & Walter, N. G. Native purification and labeling of RNA for single molecule fluorescence studies. *Methods Mol. Biol.* **1240**, 63–95 (2015).

124. Moss, B. 5’ end labeling of RNA with capping and methylating enzymes. *Gene Amplif. Anal.* **2**, 253–266 (1981).

125. Anhäuser, L. & Rentmeister, A. Enzyme-mediated tagging of RNA. *Curr. Opin. Biotechnol.* **48**, 69–76 (2017).

126. Motea, E. A. & Berdis, A. J. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim. Biophys. Acta* **1804**, 1151–1166 (2010).

127. Martin, G. & Keller, W. Tailing and 3’-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA* **4**, 226–230 (1998).

128. Merkel, M., Peewasan, K., Arndt, S., Ploschik, D. & Wagenknecht, H.-A. Copper-Free Postsynthetic Labeling of Nucleic Acids by Means of Bioorthogonal Reactions. *Chembiochem* **16**, 1541–1553 (2015).

129. Büttner, L., Javadi-Zarnaghi, F. & Höbartner, C. Site-specific labeling of RNA at internal ribose hydroxyl groups: terbium-assisted deoxyribozymes at work. *J. Am. Chem. Soc.* **136**, 8131–8137 (2014).

130. Carrocci, T. J., Lohe, L., Ashton, M. J., Höbartner, C. & Hoskins, A. A. Debranchase-resistant labeling of RNA using the 10DM24 deoxyribozyme and fluorescent modified nucleotides. *Chem. Commun* **53**, 11992–11995 (2017).

131. Shashkova, S. & Leake, M. C. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Biosci. Rep.* **37**, (2017).

132. König, I. *et al.* Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods* **12**, 773–779 (2015).

133. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008).

134. Malkusch, S. *et al.* Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochem. Cell Biol.* **137**, 1–10 (2012).

135. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).

136. Mueller, F. *et al.* FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).

137. Urbanek, M. O., Galka-Marciniak, P., Olejniczak, M. & Krzyzosiak, W. J. RNA imaging in living cells - methods and applications. *RNA Biol.* **11**, 1083–1095 (2014).

138. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).

139. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475–478 (2011).

140. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).

141. Martin, R. M., Rino, J., Carvalho, C., Kirchhausen, T. & Carmo-Fonseca, M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* **4**, 1144–1155 (2013).

142. Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165**, 976–989 (2016).

143. Wang, C., Han, B., Zhou, R. & Zhuang, X. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell* **165**, 990–1001 (2016).

144. Morisaki, T. *et al.* Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425–1429 (2016).

145. Yang, L.-Z. *et al.* Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Mol. Cell* **76**, 981-997.e7 (2019).

146. Wang, H. *et al.* CRISPR-mediated live imaging of genome editing and transcription. *Science* **365**, 1301–1305 (2019).

147. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).

148. Lee, J. *et al.* Combining SELEX screening and rational design to develop light-up fluorophore-RNA aptamer pairs for RNA tagging. *ACS Chem. Biol.* **5**, 1065–1074 (2010).

149. Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* **136**, 1198–1201 (2014).

150. Song, W. *et al.* Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. *Nat. Chem. Biol.* **13**, 1187–1194 (2017).

151. Lipson, S. G., Lipson, H. & Tannhauser, D. S. *Optical physics*. (Cambridge University Press, 1995).

152. Török, P. & Wilson, T. Rigorous theory for axial resolution in confocal microscopes. *Opt. Commun.* **137**, 127–135 (1997).

153. Klar, T. A. & Hell, S. W. Subdiffraction resolution in far-field fluorescence microscopy. *Opt. Lett.* **24**, 954–956 (1999).

154. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).

155. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. SHORT COMMUNICATION. *J. Microsc.* **198**, 82–87 (2000).

156. Gustafsson, M. G. L. *et al.* Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* **94**, 4957–4970 (2008).

157. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332–1336 (2008).

158. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).

159. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).

160. Balzarotti, F. *et al.* Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355**, 606–612 (2017).

161. Vaughan, J. C., Jia, S. & Zhuang, X. Ultrabright photoactivatable fluorophores created by reductive caging. *Nat. Methods* **9**, 1181–1184 (2012).

162. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc Natl Acad Sci USA* **93**, 6264–6268 (1996).

163. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).

164. Zhang, J. & Ferré-D'Amaré, A. R. Structure and mechanism of the T-box riboswitches. *Wiley Interdiscip. Rev. RNA* **6**, 419–433 (2015).

165. Rollins, S. M., Grundy, F. J. & Henkin, T. M. Analysis of cis-acting sequence and structural elements required for antitermination of the Bacillus subtilis tyrS gene. *Mol. Microbiol.* **25**, 411–421 (1997).

166. Henkin, T. M. The T box riboswitch: A novel regulatory RNA that utilizes tRNA as its ligand. *Biochim. Biophys. Acta* **1839**, 959–963 (2014).

167. Grundy, F. J., Winkler, W. C. & Henkin, T. M. tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proc Natl Acad Sci USA* **99**, 11121–11126 (2002).

168. Grigg, J. C. *et al.* T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proc Natl Acad Sci USA* **110**, 7240–7245 (2013).

169. Lehmann, J., Jossinet, F. & Gautheret, D. A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNAse P and T-box leaders. *Nucleic Acids Res.* **41**, 5494–5502 (2013).

170. Zhang, J. & Ferré-D'Amaré, A. R. Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* **500**, 363–366 (2013).

171. Chan, C. W., Chetnani, B. & Mondragón, A. Structure and function of the T-loop structural motif in noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **4**, 507–522 (2013).

172. Krasilnikov, A. S. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA* **9**, 640–643 (2003).

173. Grigg, J. C. & Ke, A. Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure* **21**, 2025–2032 (2013).

174. Grundy, F. J., Moir, T. R., Haldeman, M. T. & Henkin, T. M. Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Res.* **30**, 1646–1655 (2002).

175. Zhang, J. & Ferré-D'Amaré, A. R. Direct evaluation of tRNA aminoacylation status by the T-box riboswitch using tRNA-mRNA stacking and steric readout. *Mol. Cell* **55**, 148–155 (2014).

176. Chetnani, B. & Mondragón, A. Molecular envelope and atomic model of an anti-terminated glyQS T-box regulator in complex with tRNAGly. *Nucleic Acids Res.* **45**, 8079–8090 (2017).

177. Fang, X., Michnicka, M., Zhang, Y., Wang, Y.-X. & Nikonowicz, E. P. Capture and Release of tRNA by the T-Loop Receptor in the Function of the T-Box Riboswitch. *Biochemistry* **56**, 3549–3558 (2017).

178. Ha, T. *et al.* Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature* **419**, 638–641 (2002).

179. Hohng, S., Joo, C. & Ha, T. Single-molecule three-color FRET. *Biophys. J.* **87**, 1328–1337 (2004).

180. Wang, J. & Nikonowicz, E. P. Solution structure of the K-turn and Specifier Loop domains from the Bacillus subtilis tyrS T-box leader RNA. *J. Mol. Biol.* **408**, 99–117 (2011).

181. Suddala, K. C. *et al.* Hierarchical mechanism of amino acid sensing by the T-box riboswitch. *Nat. Commun.* **9**, 1896 (2018).

182. Avcilar-Kucukgoze, I. *et al.* Discharging tRNAs: a tug of war between translation and detoxification in Escherichia coli. *Nucleic Acids Res.* **44**, 8324–8334 (2016).

183. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.* **260**, 649–663 (1996).

184. Emilsson, V. & Kurland, C. G. Growth rate dependence of transfer RNA abundance in Escherichia coli. *EMBO J.* **9**, 4359–4366 (1990).

185. Winkler, W. C., Grundy, F. J., Murphy, B. A. & Henkin, T. M. The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* **7**, 1165–1172 (2001).

186. Zhang, J. & Landick, R. A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem. Sci.* **41**, 293–310 (2016).

187. Grundy, F. J. & Henkin, T. M. Kinetic analysis of tRNA-directed transcription antitermination of the Bacillus subtilis glyQS gene in vitro. *J. Bacteriol.* **186**, 5392–5399 (2004).

188. Gutiérrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C. & Merino, E. Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol. Mol. Biol. Rev.* **73**, 36–61 (2009).

189. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).

190. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. in *Recombinant DNA Methodology* 587–601 (Elsevier, 1989). doi:10.1016/B978-0-12-765560-4.50038-1.

191. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

192. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

193. Fei, J., Kosuri, P., MacDougall, D. D. & Gonzalez, R. L. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol. Cell* **30**, 348–359 (2008).

194. Blanchard, S. C., Kim, H. D., Gonzalez, R. L., Puglisi, J. D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci USA* **101**, 12893–12898 (2004).

195. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. & Wiggins, C. H. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97**, 3196–3205 (2009).

196. Bartley, L. E., Zhuang, X., Das, R., Chu, S. & Herschlag, D. Exploration of the transition state for tertiary structure formation between an RNA helix and a large structured RNA. *J. Mol. Biol.* **328**, 1011–1026 (2003).

197. Fei, J. *et al.* Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proc Natl Acad Sci USA* **106**, 15702–15707 (2009).

198. Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55 (2007).

199. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).

200. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

201. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41**, D262-7 (2013).

202. Nachtergaele, S. & He, C. Chemical Modifications in the Life of an mRNA Transcript. *Annu. Rev. Genet.* **52**, 349–372 (2018).

203. Maity, A. & Das, B. N6-methyladenosine modification in mRNA: machinery, function and implications for health and diseases. *FEBS J.* **283**, 1607–1630 (2016).

204. Warda, A. S. *et al.* Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep.* **18**, 2004–2014 (2017).

205. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).

206. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).

207. Kasowitz, S. D. *et al.* Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet.* **14**, e1007412 (2018).

208. Shi, H., Wei, J. & He, C. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol. Cell* **74**, 640–650 (2019).

209. Liu, N. *et al.* N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).

210. Chen, K. *et al.* High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. *Angew Chem Int Ed Engl* **54**, 1587–1590 (2015).

211. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).

212. Lei, Z. & Yi, C. A Radiolabeling-Free, qPCR-Based Method for Locus-Specific Pseudouridine Detection. *Angew Chem Int Ed Engl* **56**, 14878–14882 (2017).

213. Dai, Q. *et al.* Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N6-methyladenosine. *Nucleic Acids Res.* **35**, 6322–6329 (2007).

214. Liu, W. *et al.* Identification of a selective DNA ligase for accurate recognition and ultrasensitive quantification of N6-methyladenosine in RNA at one-nucleotide resolution. *Chem. Sci.* **9**, 3354–3359 (2018).

215. Harcourt, E. M., Ehrenschwender, T., Batista, P. J., Chang, H. Y. & Kool, E. T. Identification of a selective polymerase enables detection of N(6)-methyladenosine in RNA. *J. Am. Chem. Soc.* **135**, 19079–19082 (2013).

216. Wang, S. *et al.* N^6 -Methyladenine hinders RNA- and DNA-directed DNA synthesis: application in human rRNA methylation analysis of clinical specimens. *Chem. Sci.* **7**, 1440–1446 (2016).

217. Xiao, Y. *et al.* An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N6 -Methyladenosine Modification. *Angew Chem Int Ed Engl* **57**, 15995–16000 (2018).

218. Potapov, V. *et al.* A single-molecule sequencing assay for the comprehensive profiling of T4 DNA ligase fidelity and bias during DNA end-joining. *Nucleic Acids Res.* **46**, e79 (2018).

219. Harada, K. & Orgel, L. E. Unexpected substrate specificity of T4 DNA ligase revealed by in vitro selection. *Nucleic Acids Res.* **21**, 2287–2291 (1993).

220. Liu, N. *et al.* Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848–1856 (2013).

221. Zhang, Z. *et al.* Single-base mapping of m6A by an antibody-independent method. *Sci. Adv.* **5**, eaax0250 (2019).

222. Garcia-Campos, M. A. *et al.* Deciphering the “m6A Code” via Antibody-Independent Quantitative Profiling. *Cell* **178**, 731-747.e16 (2019).

223. Sednev, M. V. *et al.* N6 -Methyladenosine-Sensitive RNA-Cleaving Deoxyribozymes. *Angew Chem Int Ed Engl* **57**, 15117–15121 (2018).

224. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).

225. Tserovski, L. *et al.* High-throughput sequencing for 1-methyladenosine (m(1)A) mapping in RNA. *Methods* **107**, 110–121 (2016).

226. Fei, J. *et al.* Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.* **130**, 4180–4192 (2017).