

VISUALIZATION OF RNA-PROTEIN INTERACTION  
WITH RNA TRAP BASED FLUORESCENCE THREE-  
HYBRID ASSAY

VORGELEGT VON

NINGJUN DUAN

AUS NANJING, CHINA

MÜNCHEN, 25.11.2019

Erstgutachter: Prof. Dr. Heinrich Leonhardt

Zweitgutachter: Prof. Dr. Wolfgang Frank

Tag der Abgabe: 25. 11. 2019

Tag der mündlichen Prüfung: 19. 02. 2020

# Content

Summary.....	7
1 Introduction.....	9
1.1 Structures of RNAs.....	11
1.2 Functions of RNAs.....	13
1.3 Life Cycle of RNAs.....	17
1.4 RNA-Protein Complexes.....	23
1.5 RNA Binding Domains.....	26
1.6 Evolution of RNAs.....	30
1.7 RNA Tracking.....	32
1.8 RNA-Protein Interaction Study.....	35
1.9 Aims of This Work.....	37
2 Materials and Methods.....	39
2.1 Materials.....	41
2.2 Plasmid Construction.....	44
2.3 Cell Manipulations.....	49
2.4 Imaging.....	51
2.5 Gene Expression Measurement.....	52
3 Results.....	55
3.1 Establishment of the Fluorescence Tri-hybrid Method.....	57
3.2 Optimizations of the Fluorescence Tri-hybrid Method.....	62
3.3 Measurement of RNA-protein Interaction.....	75
3.4 Characterization of the Precise Binding Sites.....	81
3.5 Development of Alternative Anchor Sites.....	86
3.6 Development of Alternative RNA Traps.....	91
3.7 Measurement of RNA-protein binding kinetics.....	99
4 Discussion.....	101

4.1 Fluorescence Tri-hybrid Method for RNA-protein Interaction Assay .....	103
4.2 RNA Expression Cassettes .....	105
4.3 Alternative Anchor Sites.....	106
4.4 Genome Targeting with CRISPR/Cas9 System .....	108
4.5 RNA Tracking Tools.....	110
4.6 Measurement of RNA-protein Binding Kinetics.....	113
4.7 High-throughput Screening in RNA-protein Interaction Assay.....	115
5 Annex.....	117
5.1 Abbreviations .....	119
5.2 Plasmids used in this study .....	123
5.3 References.....	141
5.4 Statutory Declaration and Statement.....	159
5.5 Acknowledgements.....	161

## Summary

RNAs are the multifunctional polymers of nucleotides. From carriers and messengers of genetic information to functional executors and regulators, RNA molecules usually form stable ribonucleoprotein complexes with their binding proteins and function together in these biological processes. Therefore, identification and characterization of the interactions between RNAs and their binding proteins are the foundations for understanding the roles of RNAs.

So far, several methods have been established to measure the interaction between RNAs and proteins. Each method has its advantages and disadvantages, and is suitable for different applications. Some in vitro methods can provide precise binding parameters, but they need a large number of samples and complicated experimental processes. Although the traditional yeast-based two-hybrid can give quick results, the accuracy is still the most significant problem. Therefore, we developed a new animal cell-based fluorescence tri-hybrid method, which can be used to measure the interactions between RNAs and proteins more conveniently and precisely.

Our method was developed on a GFP labeled RNA trap, which can capture the ms2 tagged test RNAs at the *lacO* array of the BHK cells. The RFP labeled test protein, which can bind to the RNA, would also be recruited to the *lacO* site. The red spot of the test protein would appear and co-localize with the green spot of the RNA trap, and the relative brightness reveals the interaction between the RNA and the protein. We examined this method with different RNA-protein pairs, and the results indicated that the RNA trap is able to recruit the test pairs and illustrate the interactions. In order to make the measurement more sensitive and reliable, we tried several ways to enhance the positive signal and reduce the background. Our studies have demonstrated that this method could be used for not only measuring the interactions between RNAs and proteins but also finding out the precise RNA binding site of a protein or the exact protein binding sequence of an RNA molecule. Moreover, to minimize the dependency of the specific *lacO* cell lines, we modified the RNA trap so that it can target the test RNA-protein pairs to different anchor sites in the cell. And we also tried to combine the programmable RNA binding proteins with our RNA trap to make the RNA trap available for any RNA molecules without additional modifications.

In conclusion, the RNA trap based fluorescence tri-hybrid method is a powerful tool for the measurement of RNA-protein interactions, and the application of the RNA trap can also be widely expanded in biological studies.

# 1 Introduction



The technology of whole genome sequencing causes a great revolution in the area of molecular biology because it provides the ability that human beings can fundamentally understand life, including itself. According to the results of the Human Genome Sequencing Project, only less than 30,000 protein-coding transcripts have been found in such a large genome, which is about 3 billion base pairs [Venter, Adams et al. 2001]. And one question comes up: what makes up the rest parts of our genome? After years of study, more and more non-coding RNA genes have been discovered, which produce transcripts that directly have the functions of structure, catalysis, or regulation, rather than encoding proteins, and this may be the possible answer to this question [Eddy 2001].

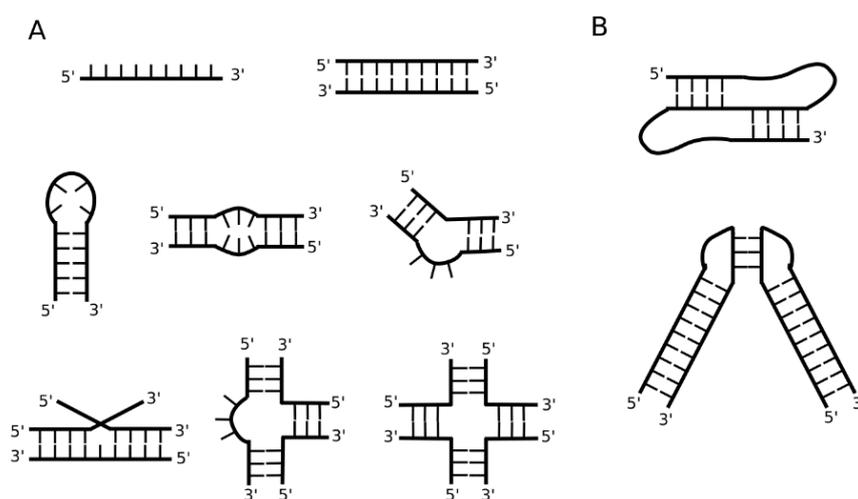
The classical central dogma of molecular biology has explained how genetic information flows between three large molecules of life: DNA passes the genetic information to DNA via replication or to RNA by transcription, while RNA transfers the genetic information to protein by translation, RNA serves as the messenger between DNA and protein [Crick 1970]. However, RNA viruses show that RNA has the ability to store genetic information independently, while the discovery of ribozymes indicates that RNAs could also have catalytic activities, which had been thought to be owned only by proteins before. And the “RNA world” hypothesis illustrates us a potential evolution path of life: RNAs are believed to be the earliest elements of life on earth, which can produce filial generations according to their own sequences. However, after billion years of evolution, the genetic information has been stored in DNAs, which are more stable in both structure and chemical properties, while the catalytic activity had been taken over by proteins, which have more abundant structures and activities [Joyce 1989]. RNA seems to fade in the background, but is that really true?

### **1.1 Structures of RNAs**

Both DNAs and RNAs are the polymers of repeated units called nucleotides, which consist of three parts: a five-carbon sugar, a single base, and up to three phosphates. The five-carbon sugar in RNA is ribose, which holds an -OH group at C2' instead of a -H as deoxyribose in DNA. This change makes vast chemical and structural differences between DNA and RNA: on the one side, the 2' -OH group has a higher polarity than -H, which provides a stronger activity to ribose, and then give RNA molecules more chances to take part in chemical processes; on the

other side, the size of -OH group is larger than -H group, the additional requirement of space twists the shape of ribose and disturbs the formation of the stable double helix that is similar with DNA molecules [Auffinger and Westhof 1997]. In general, DNA and RNA both contain four different bases. Cytosine (C), guanine (G) and adenine (A) exist in both DNA and RNA, uracil (U) is only found in RNA while thymine (T) belongs to DNA. As uracil is able to form two hydrogen bonds with adenine, guanine, or even another uracil, it provides possibilities to form wobble base pairs or complex three or four-strand helices and is also the base of the structural diversities of RNA molecules [Crick 1966].

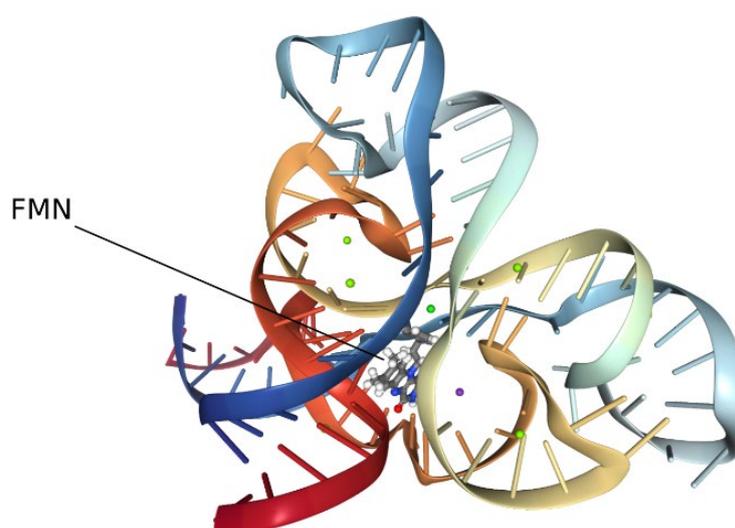
The formation of the double-strand helix structure of DNA molecules depends on the intramolecular base pairs and base-stacking interaction, while the folding of RNA molecules are simultaneously influenced by both intramolecular and intermolecular interactions between base pairs [Yakovchuk, Protozanova et al. 2006]. The higher structures of RNA depend on additional conditions such as positively charged metal ions, especially magnesium ions, which contribute to the folding of the RNA molecule by neutralizing the negative charges on phosphate ribose backbones via direct or indirect interactions [Draper, Grilley et al. 2005]. Until now, besides the primary single and double strands, complex secondary structures such as hairpin loop, bulges, and even two, three, or four stem junction had also been discovered (Figure. 1) [Shen, Cai et al. 1995].



**Figure. 1 RNA second and tertiary structures**

(A) Secondary structures of RNA molecules. From left to right in each row are single-strand, double-strand, hairpin loop, internal loop, bulges, two-stem junction, three-stem junction, and four-stem junction. (B) Two kinds of RNA tertiary motifs: pseudoknot and hairpin-hairpin interaction.

Tertiary structure is defined as the three-dimensional arrangement of RNA building blocks, which includes helical duplexes, multiple strand structures, and other secondary components that are held together through connections [Butcher and Pyle 2011]. Base pairs, ribose 2' -OH associated hydrogen bonds, metal ions mediated charge interactions, and coaxial stacking, all of them take part in the formation of RNA tertiary structures [Shiman and Draper 2000, Butcher and Pyle 2011]. From the simple riboswitches (Figure. 2) in bacterial mRNAs, which can sense small molecules and regulate transcription, to the complex rRNAs in the ribosome, which catalyze the formation of peptide bonds in protein synthesis. In all, the diversity of RNA tertiary structure is the foundation of their functions.



**Figure. 2 Structure of FMN riboswitch**

The flavin mononucleotide (FMN) riboswitch from *Bacillus subtilis* forms a stable complex with its ligand FMN and magnesium ions (PDB 6DN3) [Vicens, Mondragón et al. 2018]. It locates at the 5' UTR of the ribDEAHT operon and controls gene expression by causing premature transcription termination [Winkler, Cohen-Chalamish et al. 2002].

## 1.2 Functions of RNAs

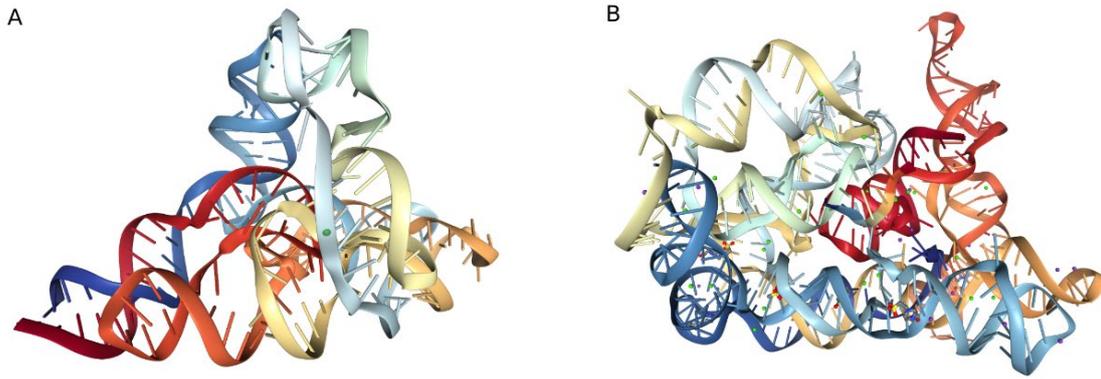
The diversity of amino acids is the foundation that provides proteins with so many different structures and functions. Similarly, although RNAs are made of fewer components, they can still have similar features as proteins.

### 1.2.1 Catalysis

The role of an enzyme is to use the active site to bring substrates together, stabilize the transition state, and prompt substrates to convert into products [Cornish-Bowden 2014]. Protein enzymes use side chains of amino acids with different polarities or electric charges as well as the shapes of active sites, to fill the correct substrates and arrange them into suitable conformation and orientation [Knowles 1991, Benkovic and Hammes-Schiffer 2003]. Similarly, ribozymes are also able to form hydrogen bonds and electrical interactions between substrates and ribose or bases to recognize and arrange substrates [Pyle, Murphy et al. 1992, Niranjankumari, Stams et al. 1998].

During the process of catalytic reaction, chemical bonds in substrates are broken and then re-formed in products. For protein enzymes, side chains of different properties in the catalytic core work together to prompt the process: the charged side chains form ionic bonds with the intermediates to stabilize their transition states, while the polar groups work as proton donors or acceptors to maintain the developing charge of the intermediate, together with the hydrophobic interaction between substrates and the non-polar side chains and the conformational change of enzyme, the correct conformations of substrates and orientations of reactive groups can be kept to increase the rate of reaction [Hammes, Benkovic et al. 2011]. For ribozymes, metal ions can coordinate with water ( $M-OH_2$ ) acting as proton donors or coordinate with hydroxyl groups ( $M-OH$ ) acting as proton acceptors in acid/base catalysis, while the amide groups ( $NH$ ) in bases can function as proton donors or acceptors in particular pH environment, furthermore, the 2' -OH of ribose can also act as the donor of proton [Fedor 2002, Fedor and Williamson 2005]. These properties provide the possibilities that allow ribozymes to take part in chemical reactions as normal protein enzymes.

Two major classes of reactions can be catalyzed by ribozymes. One is the breakage and formation of phosphodiester bonds: hammerhead ribozyme, RNase P and group I/II introns (Figure, 3) are all able to catalyze the strand breakage of substrate RNAs or themselves or even the strand re-joint of the products; another is the formation of peptide bonds: the rRNAs forming the peptidyl transferase centers of prokaryotic and eukaryotic ribosomes are able to catalyze the formation of peptide bonds during the translation process [Doherty and Doudna 2001, Ramakrishnan 2002].



**Figure. 3 Structures of Class I and II introns**

(A) The tertiary structure of class I intron (PDB 4P8Z) [Meyer, Nielsen et al. 2014]. (B) The tertiary of class II intron (PDB 4E8K) [Marcia and Pyle 2012].

### 1.2.2 Recognition and Structure

The properties of forming intramolecular and intermolecular base pairs provide RNAs the possibility to construct complex structures and recognize other polynucleotides. These two properties play essential roles in the interactions between RNAs and other large molecules.

Splicing is an essential step in the maturation process of eukaryotic pre-mRNA, which will remove the introns and generate mature products, and is performed by the spliceosome. The spliceosome is a multi-megadalton complex of several RNAs and proteins. Two unique spliceosomes coexisting in most eukaryotic cells: the U2-dependent spliceosome, which takes part in the removal of U2-type introns, and the less abundant U12-dependent spliceosome, which catalyzes the removal of the rare U12-type introns [Sharp and Burge 1997, Will and Lührmann 2011]. The U2-dependent spliceosome consists of U1, U2, U5, and U4/U6 snRNPs as well as several non-spliceosome proteins, and each snRNP contains a snRNA core, which provides the scaffold for the binding of spliceosome proteins [Hodges and Beggs 1994]. The assembly of the spliceosome and the splicing process both depend on RNA-RNA interactions. At the early stage, U1 snRNA base pairs with the 5' splice site of the intron while U2 snRNA base pairs with the branchpoint sequence within the intron to form a short U2-branchpoint complex, in which the adenosine at the branchpoint is bulged out, specifying its 2' -OH as the electron donor for the first step reaction [Will and Lührmann 2011]. In U4/U6 snRNPs, U4 and

U6 snRNAs are extensively base paired with each other, and U4/U6 snRNP binds U2 to form the pre-splice complex [Will and Lührmann 2011]. During splicing process, U1 leaves the 5' splice site of the intron, U4 snRNP releases from the complex of U2/U4/U6, while U6 snRNP binds the 5' splicing site instead, and this transition prompt the formation of the catalytic core and the execution of the cut-ligation reaction [Will and Lührmann 2011]. In general, snRNAs are responsible for holding the components together, recognizing the substrates and catalyzing the breakage and re-joint of the phosphodiester bonds of substrate RNAs while proteins drive the conformational changes of snRNAs, For example, DEAD (Asp-Glu-Ala-Asp) box helicases use the energy in adenosine triphosphate (ATP) or guanosine triphosphate (GTP) to unwind RNA helices or break RNA-protein interactions to drive the re-assembly of spliceosome in each stage [Soto-Rifo and Ohlmann 2013].

### 1.2.3 Information Carrier and Template

The sequences of nucleotides are the foundations of RNA structure and function and can also be considered as the information that RNA molecules carry. Such kind of information can be transferred to the offspring RNAs or other molecules.

Some kinds of viruses use RNA as their genetic material. The well-studied influenza A virus owns an eight segmented negative chain RNA genome, which encodes ten identified proteins, including cell entry- and exit-associated envelope proteins, components of trimeric RNA dependent RNA polymerase (RDRP) and proteins with other functions [Portela and Digard 2002]. After infection, the 5' capped and 3' polyadenylated mRNAs are synthesized from its negative-strand genome by virus-encoded RDRP and then translated into viral proteins [Bouvier and Palese 2008]. During the replication cycle, the full-length positive chain template RNAs are synthesized for the viral negative chain RNA genome first, and then the new negative RNA genomes are manufactured according to the template RNAs [Krug, Alonso-Caplen et al. 1989]. On the contrary, hepatitis C virus (HCV) contains a positive-strand RNA genome, which includes a single open reading frame (ORF) encoding a polyprotein of about 3000 amino acids, which will be split into ten smaller proteins or peptides by host proteinase or self-catalytic cleavage after translation, and the 5' internal ribosome entry site (IRES) of the genomic RNA will initial the transcription in a cap-independent manner [Takamizawa, Mori et al. 1991]. The replication of genomic RNAs starts from the synthesis of a complementary

negative-strand template RNA and is followed by the synthesis of the positive strand genomic RNAs from the template RNA [Moradpour, Penin et al. 2007].

Besides genomic RNAs, Information can also be carried and transferred to other molecules by some smaller RNA molecules. The ends of eukaryotic linear chromosomes need the protections of telomeres, which are big nucleoprotein complexes containing several kilobases of repetitive sequences and telomere proteins [Blackburn 1991]. Telomerase keeps the length of telomeres to respond to the shortening during each replication cycle [Gomez, Armando et al. 2012]. Human telomerase is a kind of nucleoprotein: its protein component hTERT is an RNA dependent DNA polymerase, which has C-terminal reverse transcriptase activity and N-terminal RNA binding ability and is able to extend telomeric DNA according to an RNA template; and the nucleotide component hTR, which is about 450 nucleotides long and transcribed by RNA polymerase II, contains one and a half times of single telomere repeats at the 5' end (46 to 53 bp) [Masutomi, Evan et al. 2003].

### **1.3 Life Cycle of RNAs**

The life cycles of RNAs are almost the same in both prokaryotic and eukaryotic cells. Besides functioning, the other major stages, including generation, modification, transport, and degradation, are all crucial to cell activities.

#### **1.3.1 Generation of RNAs**

According to the modern "RNA World" hypothesis, RNA molecules are the earliest life elements, and they cannot exist long unless they have the ability to replicate themselves [Cech 2012]. Although there is no known ribozyme that can catalyze template-directed polymerization of nucleotides, molecules with similar functions have been obtained from the test tube evolution [Joyce 2002]. One of these molecules is named magnesium-dependent class I ligase, which is able to perform the template-directed joining of 3' -OH and 5' phosphate group of oligonucleotides [Ekland, Szostak et al. 1995, Bergman, Lau et al. 2004].

Although DNAs have taken the place of RNA that become the major genetic materials, some kinds of viruses like positive/negative-strand RNA viruses and retroviruses still use RNAs as their genetic materials. For positive or negative chain RNA viruses hepatitis C virus and

influenza A virus, RDRPs are needed to synthesize the complementary template RNAs and later viral genomic RNAs [Krug, Alonso-Caplen et al. 1989, Freed 2001, Moradpour, Penin et al. 2007]. Human immunodeficiency virus (HIV), one member of retroviruses, its genomic RNA is reverse transcribed into a DNA molecule by RNA dependent DNA polymerase (RDDP) and then integrated into the DNA genome of host cells after infection, and the new viral genomic RNAs are generated by host's RNA polymerase II as normal transcription process [Krug, Alonso-Caplen et al. 1989, Freed 2001, Moradpour, Penin et al. 2007].

For both unicellular and multicellular organisms, RNAs are typically generated according to the templates in the DNA genomes by DNA dependent RNA polymerases (DDRPs). For prokaryotic cells such as *Escherichia coli* (*E. coli*), all the RNA molecules are synthesized by only one kind of RNA polymerase, which consists of a  $\beta\beta'\alpha_2$  core enzyme that provides the catalytic activity. The recognition of specific promoters was performed by one of seven different species of  $\sigma$  factors, which is also an essential part of the complete RNA polymerase [McClure 1985, Ishihama 2000]. For the more complex eukaryotic cells, different RNAs are generated by different RNA polymerases, since each kind of promoter can only be recognized by specific RNA polymerase, and different RNA products have different processes and transport pathways, which have close relationships with the polymerases. Normally, there are three major RNA polymerases existing in eukaryotic cells: RNA pol I generates rRNA precursors, which will mature into 28S, 18S and 5.8S rRNAs, RNA pol II takes part in the synthesis of mRNAs, most snRNAs, and some microRNAs, and RNA pol III produces 5S rRNA, tRNAs and other small RNAs [Chambon 1975, Sentenac 1985]. In some plant cells, two additional polymerases RNA Pol IV and V have been discovered, which mainly focus on the synthesis of siRNAs [Zheng, Wang et al. 2009].

### 1.3.2 Processing and Modification of RNAs

After RNAs are generated, maturation processes are necessary before they have complete functions. RNA process and modification are universal mechanisms existing from simple viruses to complex cell organisms.

Hepatitis D virus (HDV) is a subviral pathogen that is frequently found in patients infected with the hepatitis B virus [Pascarella and Negro 2011]. It owns a circular RNA genome and undergoes a rolling-circle replication during reproduction, in which RNA editing plays a critical

role: the specific modification at position 1012 from U to C results in the elimination of the stop codon, leading to a transformation of the product HDAg from a shorter form p24 to a 19 or 20 amino acids longer form p27, which will inhibit the replication process and start viral particle package [Casey and Gerin 1995].

In the genome of *E. coli*, 5S, 16S, and 23S rRNA, as well as a few tRNAs, are organized as a co-transcript operon and sequentially synthesized by RNA polymerase as a precursor RNA [Klappenbach, Dunbar et al. 2000]. After transcription, 16S rRNA will be released from the precursor RNA by endonuclease RNase III and modified by exonuclease RNase M16 at 5' end, 5S rRNA will be cleaved by RNase III, RNase P, and RNase E, and 23S rRNA will be released and modified by RNase III and RNase M23, meanwhile, tRNAs will be cleaved by RNase F and RNase P, and then undergo 3' CCA sequence addition by nucleotidyltransferase [Srivastava and Schlessinger 1990]. Nucleotide modifications, including uridine to pseudouridine ( $\psi$ ) and other kinds of base conversions, nucleotide insertions, and deletions, as well as methylations, are also applied during rRNA and tRNA maturation [Brennicke, Marchfelder et al. 1999].

For eukaryotic cells, RNA processing has been developed to a new level. Similar to prokaryotes, eukaryotic ribosomal 18S, 5.8S, and 28S rRNAs are transcribed as a 45S precursor by RNA pol I, and the 5' and 3' external as well as spacer sequences are removed by RNase to generate the mature rRNAs [Eichler and Craig 1994]. Nucleotide modifications also happen during maturation, and the processes like uridine to pseudouridine ( $\psi$ ) and nucleotide methylation are performed by nucleoproteins, which use small nucleolar RNAs as guides to recognize the correct positions instead of by proteins themselves [Henras, Plisson-Chastang et al. 2015]. tRNA precursors are generated by RNA pol III, followed by 5' end removal by RNase P, 3' CCA addition, nucleotide modifications, and even intervening sequence cleavages in some molecules [Deutscher 1984, Björk, Ericson et al. 1987]. The processing of mRNA precursors occurs simultaneously with transcription: 7-methylguanosine ( $^7\text{mG}$ ) cap is added to the 5' end of mRNA precursor, poly(A) tail is added to the 3' end by polyadenylase, which is able to keep mRNA stable and suitable for transport and translation, and intron sequences are removed by spliceosome to generate the correct products [Bentley 1999]. Furthermore, the methylation on N<sup>6</sup> of adenosines in 3' untranslated region (UTR) or near stop codon also demonstrates the regulatory functions during gene expression [Nevins 1983]. Small RNAs such as miRNAs come from precursor RNAs that are synthesized by RNA pol II, and become

functional after the cleavage by the nuclease Drosha/Dicer and base editing [Cai, Yu et al. 2009].

### 1.3.3 Transport and Distribution of RNAs

The lack of nuclei and the coupling between transcription and translation give rise to the theory that the specific RNA localization is not relevant for bacterial cells, but studies have documented that various RNA molecules can be localized to different subcellular regions [Keiler 2011]. For *E.coli*, *lacZ* mRNAs have been observed that distribute throughout the cytoplasm, 5S rRNAs are in the areas devoid from nucleoid and considered relevant to the distributions of ribosomes, and some short non-coding RNAs accumulate at cell poles [Valencia-Burton, McCullough et al. 2007]. Some assumptions have been developed to explain the localized translation, for example, translation of RNAs at the sites where their protein products function, or the localizations may be correlated with RNA processing [Nevo-Dinur, Govindarajan et al. 2012].

Eukaryotic cells are defined by the existence of nuclear membranes as well as other membrane-bound organelles. Compared with prokaryotic cells, an advanced intracellular membrane system separates the cell environment into different areas so that life processes can carry out more orderly and efficiently, but new challenges come out that large molecules like RNAs cannot pass through the membranes freely unless with the help of a complete transport system.

During transcription, precursor mRNAs are surrounded by proteins not only for capping, poly(A) tailing and splicing but also the so-called mRNA export adaptors including transcription/export complexes (TREXs) and SR proteins, which bind mRNAs by the recognition of conserved sequences in all transcripts or are loaded to transcripts during the mRNA maturation processes [Reed and Cheng 2005]. Splicing is critical for mRNA export: on the one hand, the TREX complex and SR proteins are loaded to mRNA during splicing; on the other hand, introns act as the signals that firmly retain the mRNAs in nucleus [Zahler, Lane et al. 1992, Le Hir, Nott et al. 2003, Masuda, Das et al. 2005]. Furthermore, mRNA will stay in the nucleus without the addition of the poly(A) tail, because the export process cannot carry on generally without the coating of poly(A) binding proteins (PABPs) [Chekanova and Belostotsky 2003]. As large molecules, mRNAs cannot pass the nuclear membrane directly

unless via the channels called nuclear pore complexes (NPCs), which are approximately 50 megadaltons assemblies of proteins and able to selectively transport cargos through the nuclear envelope [Alber, Dokudovskaya et al. 2007]. mRNAs reach the NPCs by random diffusion, and they will pass through the NPCs with the help of the additional proteins called nuclear export receptors. mRNA export adaptors TREX and SR protein can be recognized by the nuclear export receptor TAP and associated protein p15, which can promote mRNAs pass through NPCs by using ATP [Huang, Gattoni et al. 2003]. During transport, mRNAs firstly contact the nuclear basket of NPCs and then pass through the pores in a 5' to 3' direction by the hydrophobic interactions between TAP proteins and NPC inner channel protein FG nucleoporins [Stewart 2010].

The subcellular localization of mRNAs connects closely to their functions. During the oogenesis process of *Drosophila*, the specification of the oocyte is closely related to the selective accumulation of RNA molecules, which are generated and provided by neighboring, interconnecting ovarian nurse cells [St Johnston 1995]. Subsequently, the depositions of mRNA transcripts at selected sites within the oocyte can lead to localized translation of proteins, and furthermore, the establishment of the prospective embryonic body axes [Bullock and Ish-Horowicz 2001]. Similarly, during zygotic development, especially the unicellular syncytial blastoderm embryo, several gene products, including pair-rule and wingless segmentation, lie exclusively at the top of the layer of several thousand peripheral nuclei [Davis and Ish-Horowicz 1991, Simmonds, Livne-Bar et al. 2001]. The “post stamps” in mRNAs are considered as several nucleotides in 3' UTR, which are called zipcodes, they have either particular sequences or fold into individual structures so that they can be recognized by protein factors, and the stamped mRNA cargoes can be loaded on the boats of the molecular motors like myosin and kinesin, which can move along the cytoskeleton system, and transported to their destinations [Singer 1993]. For example, the zipcodes in  $\beta$ -actin mRNAs contain several ACACCC repeats, which can be recognized by ZBP1, ZBP2, and EF1 $\alpha$ , and the mRNA-protein factor complexes are transported to the locations such as axon terminals by motor proteins [Ross, Oleynikov et al. 1997, Zhang, Eom et al. 2001].

The assembly of Eukaryotic ribosomal subunits is taking place in both nucleolus and nucleoplasm; moreover, additional modification steps are also needed in the cytoplasm [Kressler, Hurt et al. 2010]. Pre-ribosomal particles are transported into the cytoplasm

through NPCs with the help of export factors like Xpo1, Crm1 and Ran GTPase [Moy and Silver 2002]. Nmd3 functions as a bridge between the export factor and the pre-ribosomal particles: on the one hand, it docks on the particles by its N-terminal domain; on the other hand, it provides a short nuclear export sequence (NES) in its C-terminal sequence, which can be recognized by Xpo1, the export factor that interacts with the NPC inner channel protein FG nucleoporins [Kressler, Hurt et al. 2010]. For tRNA transport, they are recognized by transporter exportin-t and moved out through NPCs with the help of Ran GTPase [Arts, Kuersten et al. 1998].

### 1.3.4 Degradation of RNAs

When the amount of RNA exceeds the demand or the quality of RNA changes, both prokaryotic and eukaryotic cells have a series of mechanisms to remove the unneeded RNA molecules. Three major enzymes are responsible for RNA degradation: endonucleases cleave RNA into small fragments internally, while 5' and 3' exonucleases digest RNA into smaller particles from both ends. During the degradation process, additional protein factors are also needed to recognize the unneeded RNA molecules or recruit nucleases [Houseley and Tollervey 2009].

For *E.coli*, the steady-state concentrations of mRNAs are directly proportional to their half-lives [Belasco 1993]. During the degradation process, mRNA molecules firstly undergo the removal of 5' triphosphates by RppH and internal cleavage by the endonucleases like RNase E, which prefers A/U rich regions in close proximity to stem-loops, followed by degradation from the 3' terminal end, which is carried out by three main exonucleases: PNPase, RNase R and RNase I [Rauhut and Klug 1999]. Among the three, both RNase R and RNase II are hydrolases and digest RNA molecules into nucleoside monophosphates with water, while PNPase is a phosphorolytic nuclease that uses ATP and yields nucleoside diphosphates [Silva, Saramago et al. 2011]. However, the work that looking for the nuclease that takes part in 5' terminal digestion in *E.coli* is still going on [Rauhut and Klug 1999].

For eukaryotic cells, degradations accompany each life stage of RNA [Houseley and Tollervey 2009]. For the products of RNA pol II, the failure of 5' capping will induce 5' degradation by exonuclease Rat5 and termination of transcription, and the unsuccessful packaging and export processes will lead to 5' or 3' degradation [Rougemaille, Villa et al. 2008, West,

Proudfoot et al. 2008]. When translation starts, the appearance of the premature stop codon leads to degradation [Isken and Maquat 2008]. During translation, the poly(A) tail is progressively shortened by the complex containing Ccr4 and Caf1 or PARN deadenylase, which mainly functions in regulated deadenylation [Schwede, Ellis et al. 2008]. After decapped by the Dcp1-Dcp2 complex, the deadenylated mRNAs will be digested from the 5' end by Xrn1 or the 3' end by the exosome [Houseley, LaCava et al. 2006]. Furthermore, degradation will also happen in different surveillance pathways such as no-stop decay (NSD) and no-go decay (NGD) dealing with the stalled ribosomes [Sachs 1993]. Small RNA guided mRNA degradation is also substantial in gene expression regulation: miRNA or siRNA containing RNA Induced Silencing Complexes (RISCs) are able to target to its complementary mRNA and start digestion by recruiting Argonaute nuclease, or inhibit the translation process by aiming to the 3' UTR of mRNA [Pratt and MacRae 2009]. The external and internal space regions of RNA pol I polycistronic products are removed by RNase during the maturation process, and most defective pre-ribosomal rRNAs will be oligo(A) tagged by TRAMP complexes and then degraded by exosomes [Allmang, Mitchell et al. 2000]. Although the mature ribosomes are stable, they can still be engulfed by vacuoles and then degraded via an autophagy pathway under stress conditions [Beau, Esclatine et al. 2008, Mroczek and Kufel 2008]. And tRNAs, the products of RNA pol III, can be cleaved by 5' nuclease Rat1 and endonuclease Xrn1 [Thompson, Lu et al. 2008].

## **1.4 RNA-Protein Complexes**

Although some RNA molecules can work by themselves, most of them still function in complexes with proteins. On the one side, RNAs provide scaffolds and docking sites to bring proteins together; on the other side, proteins can help to keep the RNAs in suitable conformations or protect RNAs from external damages.

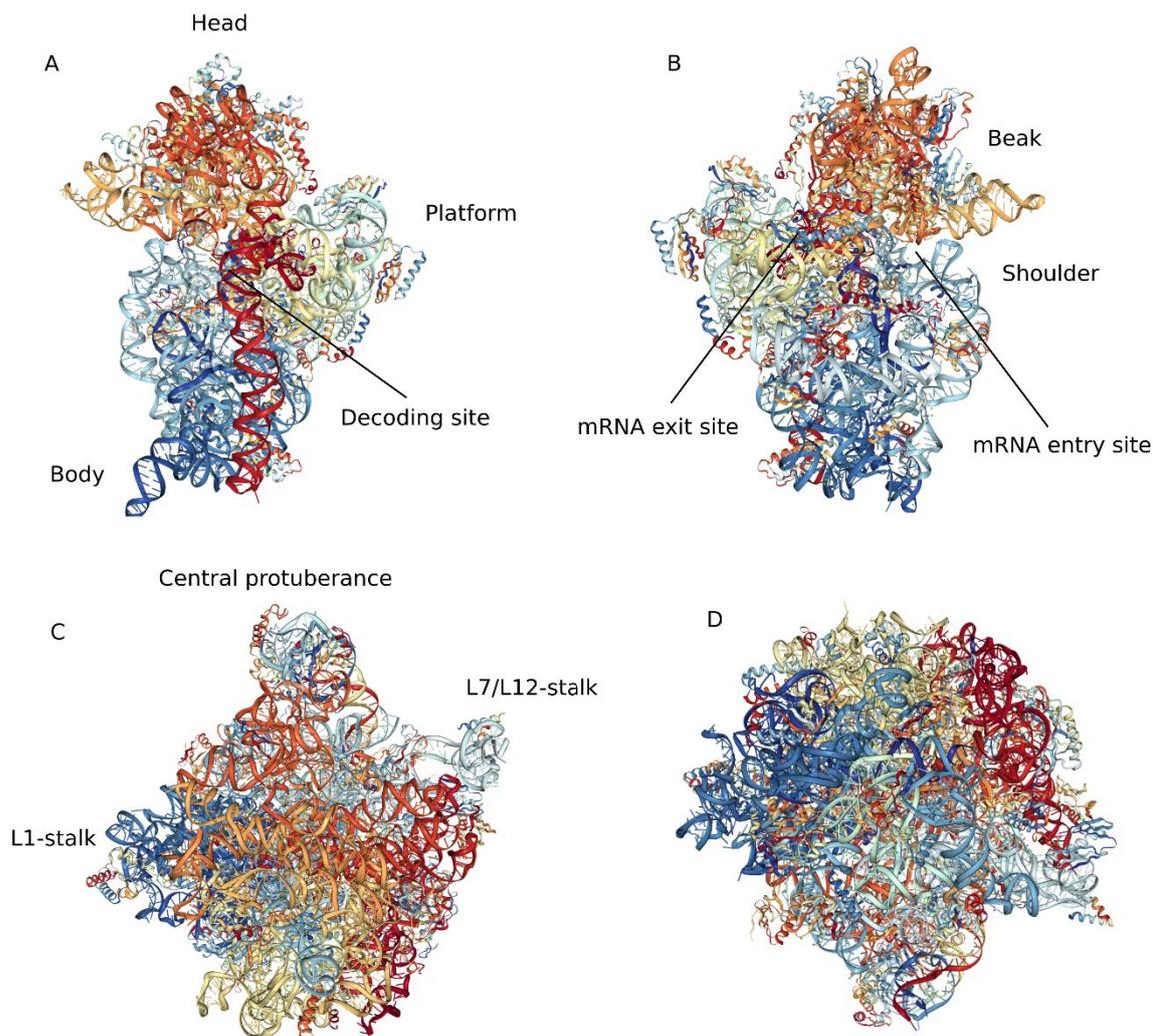
### **1.4.1 RNA Virus Nucleocapsid**

The core structure of RNA viruses is the nucleocapsid, in which the RNA genome is surrounded in a helical or polyhedral structure formed by coat proteins, is one of the most common RNA protein complexes [Lidmar, Mirny et al. 2003, Vernizzi and de la Cruz 2007]. Bacteriophage MS2 has an icosahedral capsid shell, which consists of 180 coat proteins, to protect its

positive-strand RNA genome [Valegård, Liljas et al. 1990]. The 3569 nucleotide-long RNA genome encodes four products: mature protein, coat protein, replicase, and lysis protein [Fiers, Contreras et al. 1976]. Besides, the specific stem-loop structure in the MS2 genome can be recognized by its coat protein to accomplish the repression of translation and genome encapsulation [Koning, van den Worm et al. 2003].

### 1.4.2 Ribosome

Ribosomes are the most common and important nucleoprotein complexes in cell organisms. The prokaryotic 70S ribosome can be divided into 30S and 50S subunits, for example, the *E.coli* 30S small subunit consists of a 16S rRNA and 21 proteins while its 50S large subunit is made of 5S and 23S rRNAs, as well as 31 proteins [Wittmann 1983]. Eukaryotes have 80S ribosomes in the cytoplasm, and they consist of 40S and 60S two parts. The 40S subunit has 18S rRNA and 33 proteins, while the 60S subunit is composed of 5S, 5.8S and 28S rRNAs, as well as 46 proteins [Ben-Shem, de Loubresse et al. 2011]. Although the compositions of prokaryotic and eukaryotic ribosomes are different, similar structures have been found in the subunits of both ribosomes (Figure. 4). The 30S and 40S small subunits share a similar shape, and both of them include landmarks such as 'head', 'body', 'platform', 'beak' and 'shoulder'. Both tRNA and mRNA binding sites are all located on the subunit interface, the mRNA enters through a gap existing between 'head' and 'shoulder' and exits between 'head' and 'platform', and the decoding center, where codon pairs with anticodon, is also located on the interface surface and consists of three domains from 'head', 'shoulder' and penultimate stem [Melnikov, Ben-Shem et al. 2012]. The 50S and 60S large subunits also share a similar overall crown-like shape, which includes the 'central protuberance', 'L1-stalk' and 'L7/L12-stalk', three tRNA binding sites, and the peptidyl transferase center, which catalyzes the formation of peptide bonds [Melnikov, Ben-Shem et al. 2012]. The interactions from several contact sites at the interfaces of subunits promote the assembly of the complete ribosome, and each of these contact sites is formed with the interactions between the ribosomal proteins of different subunits [Ben-Shem, de Loubresse et al. 2011].



**Figure. 4 Structures of bacterial 30s and 50s ribosome subunits**

(A, B) The subunit and solvent interfaces of the 30s subunit of the bacterial ribosome, which includes the landmarks like ‘Head’, ‘Platform’, ‘Body’, ‘Beak’, and ‘Shoulder’ (PDB 4V9D) [Dunkle, Wang et al. 2011]. (C, D) The subunit and solvent interfaces of the 50s subunit of the bacterial ribosome, which includes the landmarks such as ‘Central protuberance’, ‘L1-stalk’, and ‘L7/L12-stalk’ (PDB 1NJI) [Hansen, Moore et al. 2003].

### 1.4.3 Chromatin Remodeling Complex

The importance of histone modifications and chromatin structures in the regulation of eukaryotic gene expression has been widely accepted over the past several years. Chromatin remodeling complexes can be partitioned into two major groups according to their function modes, one group is named ATP dependent complex, which uses ATP as energy to arrange or

alter the associations between histones and genome DNA, another is called histone modification complex, which includes histone acetyltransferase (HAT), histone deacetylase (HDAC), and histone methyltransferase (HMT), they regulate gene expression by attaching the side chains of histones with different groups, such as acetyl, methyl, and ubiquitin [Vignali, Hassan et al. 2000]. Long non-coding RNAs play important roles in gene expression regulation, and one mechanism is that changing the modification situations of histones via chromatin remodeling complexes [Rinn and Chang 2012]. The 2.2 kb lncRNA Hox Transcript Antisense RNA (HOTAIR), which is generated from *HoxC* loci, is able to silence the transcriptions across the 40 kb *HoxD* loci *in trans* by inducing a repressive chromatin condition. The 5' sequence of HOTAIR is able to recruit the polycomb repression complex 2 (PRC2), which catalyzes the trimethylation of the lysine at the 27<sup>th</sup> position of histone H3 and then inhibits the expression of surrounding genes [Margueron and Reinberg 2011, Davidovich and Cech 2015]. And interestingly, the 3' sequence of HOTAIR can recruit lysine-specific demethylase 1 (LSD1), which is able to remove the methyl group of lysine at the 4<sup>th</sup> position of histone H3, shows a transcriptional inhibitory function together with PRC2. [Yu and Li 2015].

### 1.5 RNA Binding Domains

The assembly of an RNA protein complex is based on the interactions between RNAs and proteins, especially the RNA binding domains within the proteins. Several kinds of RNA binding domains have been found so far (Table. 1), which can be divided into two groups based on their binding types: one recognizes specific nucleotide sequence in target RNA, the other recognizes the higher structure that RNA folds into, and in addition, more than one RNA binding domains can exist in one RNA binding protein, which allows protein to recognize different RNA substrates, or different parts of a single RNA molecule [Glisovic, Bachorik et al. 2008].

RNA binding domain	Structure	Binding type	Example protein
<b>RRM (RNA recognition motif)</b>	$\beta\alpha\beta\beta\alpha\beta$ barrel-like	ssRNA, ssDNA	hnRNP A1, RMB3, PABP

<b>KH (K-homology)</b>	Three strands $\beta$ -sheet packed against $\alpha$ -helices	ssRNA, ssDNA	hnRNP K, Nova
<b>CSD (cold-shock domain)</b>	$\beta$ -barrel structure	ssRNA, ssDNA	CspB, YB-1
<b>dsRBD</b>	A $\beta\beta\beta\alpha$ fold	dsRNA	Dicer, RNase III
<b>Zinc finger</b>	B $\beta\alpha$ structure	dsRNA, ssRNA	TFIIIA, WT1
<b>PAZ</b>	B-barrel reminiscent with OB-fold	dsRNA	Dicer
<b>PIWI</b>	RNase H core, five-stranded $\beta$ -sheet surrounded by $\alpha$ helices	dsRNA	Argonaute
<b>PUF</b>	Eight repeats of a three $\alpha$ -helices bundle	ssRNA	Pumilio
<b>Pentatricopeptide repeats</b>	2-30 repeats of two anti-parallel $\alpha$ -helices	ssRNA	PPR10
<b>Homeodomain</b>	Helix-turn-helix	dsRNA	Jerky, bicoid

**Table 1.** Some kinds of RNA binding domains, including their structures, binding substrates, and example proteins [Auweter, Oberstrass et al. 2006, Mackereth and Sattler 2012].

### 1.5.1 Sequence-specific Binding Domains

The most well-studied RNA binding domain is RNA Recognition Motif (RRM), which has been discovered in all life kingdoms, and is also one of the most abundant protein domains in eukaryotic cells [Maris, Dominguez et al. 2005]. RRM domains distribute in a wide range of proteins, especially those participating in post-transcriptional events, such as mRNA processing, alternative splicing, RNA editing, and translational regulation [Maris, Dominguez et al. 2005]. A typical RRM domain contains 80 to 90 amino acids, which fold into a four-strand anti-parallel  $\beta$ -sheet with additional two  $\alpha$ -helices to generate an ordered barrel-like  $\beta\alpha\beta\beta\alpha\beta$  structure, and for different RRM domains, two highly conserved areas called RNP1 and RNP2, which locate in  $\beta$ -sheet 1 and  $\beta$ -sheet 2 respectively, are essential for RNA binding [Nagai, Oubridge et al. 1990]. Normally, a single RRM is able to recognize eight nucleotides. The target ssRNA

sequence lies across the surface of the  $\beta$ -sheets in the RRM, while the amino acids in the RNP1 and RNP2 provide the side chains for base tracking and ionic interactions with the RNA molecule [Daubner, Cléry et al. 2013]. Furthermore, multiple RRMs with flexible linker sequences provide the RNA binding protein additional possibilities in target molecule recognition and binding kinetics [Glisovic, Bachorik et al. 2008].

Single-cell organisms need rapid response abilities to the sudden changes in environments, for example, bacteria can deal with the drastic changes of temperature by generating a series of cold shock proteins or cold-induced proteins (CIPs), and most of which are RNA binding proteins [Ermolenko and Makhatadze 2002]. A highly conserved motif has been found in these CIPs called cold shock domain (CSD), which consists of five anti-parallel  $\beta$  strands to form a two  $\beta$ -sheets structure, and highly conserved aromatic and basic side chains protruding from the solvent face of  $\beta$ -sheets to mediate the specific RNA binding in combination with other structures [Graumann and Marahiel 1998]. In prokaryotic cells, CIPs work as RNA chaperons to prevent the formation of RNA secondary structures in low temperature, which may inhibit the synthesis of proteins by blocking the translation process [Jiang, Hou et al. 1997]. CSDs have also appeared in eukaryotic proteins like Y-box proteins, which are involved in the regulation of mRNA translation in germline and somatic cells [Didier, Schiffenbauer et al. 1988, Bouvet, Matsumoto et al. 1995].

Pumilio and FBF homology (PUF) proteins, which are eukaryotic RNA binding proteins and function as translational regulators by targeting the 3' UTR of mRNA, mainly involve development and differentiation regulations [Glisovic, Bachorik et al. 2008]. Typically, a PUF protein includes a domain that contains eight repeats of three  $\alpha$ -helices structure, and each helix is made of a 36 amino acid core sequence as well as N-/C-terminal flanking regions [Wang, McLachlan et al. 2002]. Each repeat responds the recognition of single-nucleotide, in detail, amino acids at the 12<sup>th</sup> and 16<sup>th</sup> positions contact the base via hydrogen bonds or van der Waals contacts with the Watson-Crick edge, while the 13<sup>th</sup> position amino acid makes a stacking interaction [Wang, McLachlan et al. 2002, Filipovska, Razif et al. 2011].

### 1.5.2 Structure-specific Binding domains

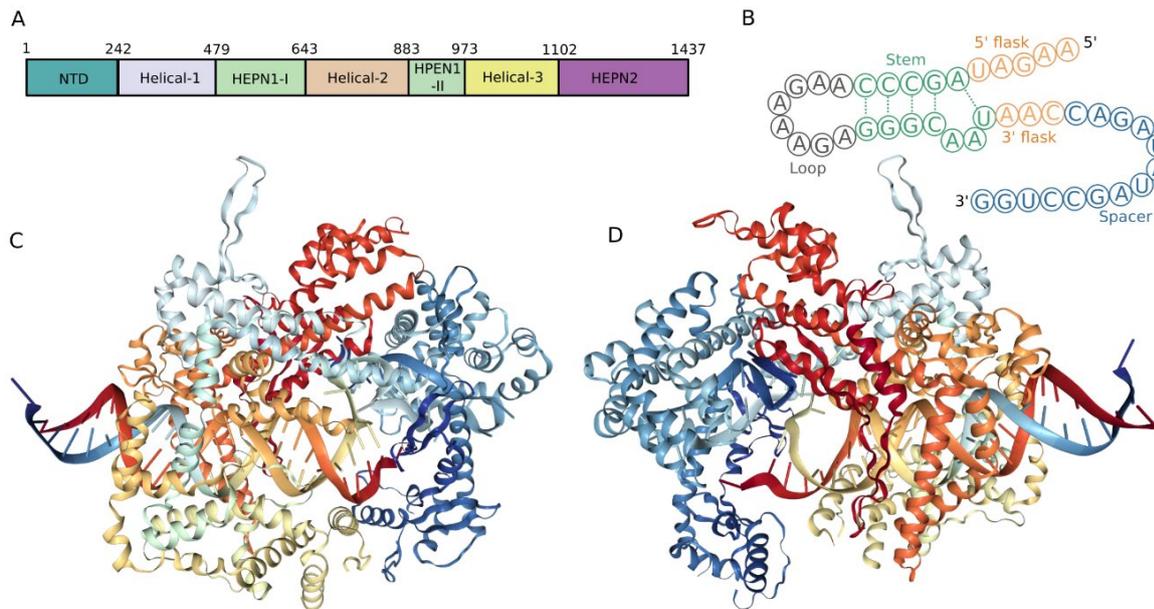
The double-strand RNA binding domain (dsRBD) is one of the most abundant RNA binding structures after RRM, and it has also been found in many proteins, which are mainly involved

in RNA maturation and localization, such as RNase III, Dicer and Drosha [Masliah, Barraud et al. 2013]. A normal dsRBD is approximate 65 to 70 amino acids in length and can fold into a compact  $\alpha\beta\beta\alpha$  structure, in which the two  $\alpha$ -helices are packed against a three-strand anti-parallel  $\beta$ -sheet [Ramos, Grünert et al. 2000]. The N-terminal parts of the 1<sup>st</sup> and 2<sup>nd</sup>  $\alpha$ -helix, together with the loop which connects the 1<sup>st</sup> and 2<sup>nd</sup>  $\beta$ -sheet, take part in the recognition of the A-form RNA helix by forming contacts with the bases and the ribose moieties in the minor groove and the phosphate backbone in the major groove [Doyle and Jantsch 2002].

The zinc finger proteins are generally considered as DNA binding proteins, however, certain classes of zinc finger proteins including the universal Cys2-His2 and Cys4 zinc fingers, which are around 30 amino acids long and consist of two anti-parallel  $\beta$ -stands followed by an  $\alpha$ -helix ( $\beta\beta\alpha$ ) and stabilized by the coordination between zinc ion and Cys/His, can also function as RNA binding proteins [Brown 2005, Hall 2005]. One typical zinc finger protein TFIIIA is able to bind 5S rRNA by recognizing the stem-loop structure within the rRNA [Joho, Darby et al. 1990, Theunissen, Rudt et al. 1992].

### 1.5.3 RNA Guided RNA Binding Protein

The well-known RNA guided Cas9 nuclease, which comes from the Type II clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system, can recognize and digest the double-strand DNAs from infected viruses, has been widely modified and used in genome engineering [Mali, Yang et al. 2013]. Similarly, the RNA guided Cas13a nuclease, which belongs to the class II type IV CRISPR/Cas system, holds the RNA-activated RNase activity and responds for crRNA processing and single-strand RNA degradation in RNA virus defense (Figure. 5) [Tambe, East-Seletsky et al. 2018]. Cas13a recognizes target RNA molecules with the help of a short crRNA that consists of a direct repeat stem-loop structure and an spacer sequence, which is anti-parallel with the target RNA, and undergoes the cleavage function by two higher eukaryotes and prokaryotes nucleotide-binding domains (HEPNs) [Knott, East-Seletsky et al. 2017, O'Connell 2019]. Nowadays, Cas13a has also been modified as a targeting tool for RNA tracking and editing by eliminating the RNase activity and fusing with molecules of different functions, such as fluorescent proteins and RNA modification enzymes [Kim 2018].



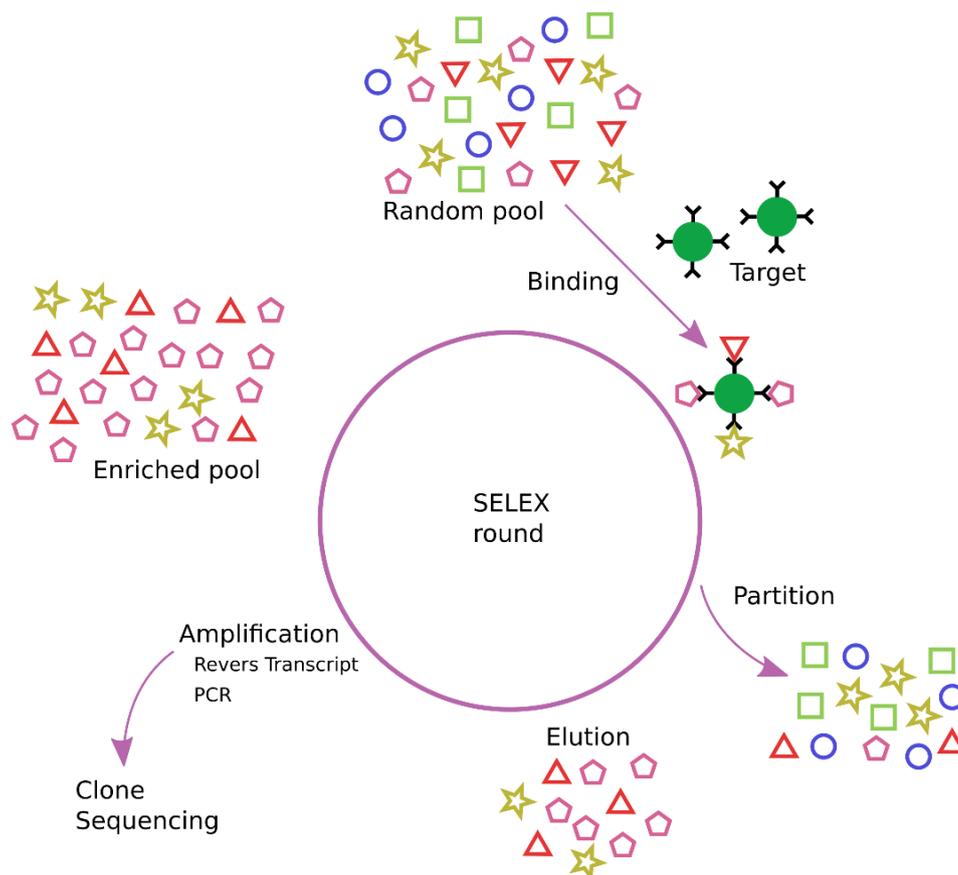
**Figure. 5 Structures of Lbu-Cas13a and crRNA**

(A) The motifs of Lbu-Cas13a. Domains from N-terminal to C-terminal are NTD, Helical-1, HEPN1-I, Helical-2, HEPN1-II, Helical-3, and HEPN2 [Liu, Li et al. 2017]. (B) The structure of crRNA. It contains a stem-loop structure and a spacer sequence, as well as 5' and 3' flasks [Knott, East-Seletsky et al. 2017]. (C, D) Two sides of Lbu-Cas13a, crRNA and target RNA ternary complex structure (PDB 5XWY) [Liu, Li et al. 2017].

## 1.6 Evolution of RNAs

The “RNA World” hypothesis is the theory that RNA should be the earliest life type on earth, and ancient RNA molecules had the ability that generates descendant RNAs by themselves [Joyce 1989]. The most important feature of these molecules is that the genotype and phenotype are integrated into a single unit, so they provide the possibilities that the theory of Darwinian evolution can be visualized at the molecular level: mutational errors and recombination events are the sources of the genetic variations, and these variations can be displayed as a variety of phenotypes. Moreover, the differences of chemical behaviors provide special abilities to some RNA molecules, one of which enables them to replicate faster than other molecules, to make their populations in a dominant position in the RNA world [Joyce 1989].

The requirement of RNA molecules for scientific research is highly growing now. Instead of looking for suitable molecules in nature, a technology called systematic evolution of ligands by exponential enrichment (SELEX) had been developed to speed up the evolution process of RNAs, which allows us to obtain RNA molecules with specific ligand binding abilities in a short time (Figure. 6) [Ellington and Szostak 1990]. The generation of a ligand-binding RNA molecule or aptamer is shown as follows. The original aptamers come from a complex random oligonucleotides library consisting of as much as  $10^{15}$  different particles that can be produced by chemical synthesis [James 2007]. In the beginning, the random RNA pool is incubated with the target molecule, the binding complexes are subsequently partitioned from the unbound as well as the weakly bound oligonucleotides. This is the critical step of the whole selection process and will strongly affect the binding feature of the aptamers [Stoltenburg, Reinemann et al. 2007]. After that, the target-binding oligonucleotides are eluted from their ligands, followed by reverse transcription, PCR amplification, and *in vitro* transcription, to generate a smaller but more condensed oligonucleotide pool, which has fewer RNA motifs but their affinity and specificity are much stronger than those in the original pool, and will be used for the next round selection [Stoltenburg, Reinemann et al. 2007]. Moreover, additional steps can also be added into each round to increase the selection efficiency according to certain conditions. Negative selection and subtraction are recommended to minimize the enrichment of unspecific binding or to direct the selection to a specific epitope of the target [Stoltenburg, Reinemann et al. 2007]. As the affinity of aptamers to their ligands can be influenced by the selection conditions, the partition stringency can be increased progressively during the whole SELEX process (usually the binding and wash conditions, such as buffer compositions, value and time) [Marshall and Ellington 2000]. The cycle number of SELEX depends on several parameters such as the characteristics and concentration of the target, the design of the beginning nucleotide pool, and the partition methods, typically, 5 to 20 SELEX rounds are essential to obtain the aptamers with high affinity and specificity [Joyce 1989]. After SELEX, sequencing and detailed characterizing of binding features are necessary for individual aptamer clones. Phylogenesis can also be applied to optimize their binding performances [Gopinath 2007].



**Figure. 6 General stages of *in vitro* target binding aptamers selection using SELEX**

Each process starts from a synthetic random oligonucleotide pool. The whole process consists of selection (binding, partition, and elution), amplification, and conditioning. In each round, unbinding oligonucleotides will be removed by several washing steps, and the binding oligonucleotides will be amplified for next round SELEX. The final oligonucleotides will be cloned for sequencing.

## 1.7 RNA Tracking

In protein studies, test proteins are usually tagged with specific labels, so that their activities can be tracked both *in vitro* and *in vivo*. Several labeling technologies have been established so far, from the classical antibody-based immunofluorescence to the widely used labeling with fluorescent proteins. Similarly, unique labels are also necessary for RNA studies, which allow RNAs of interest to be tracked *in vitro* and *in vivo* as well.

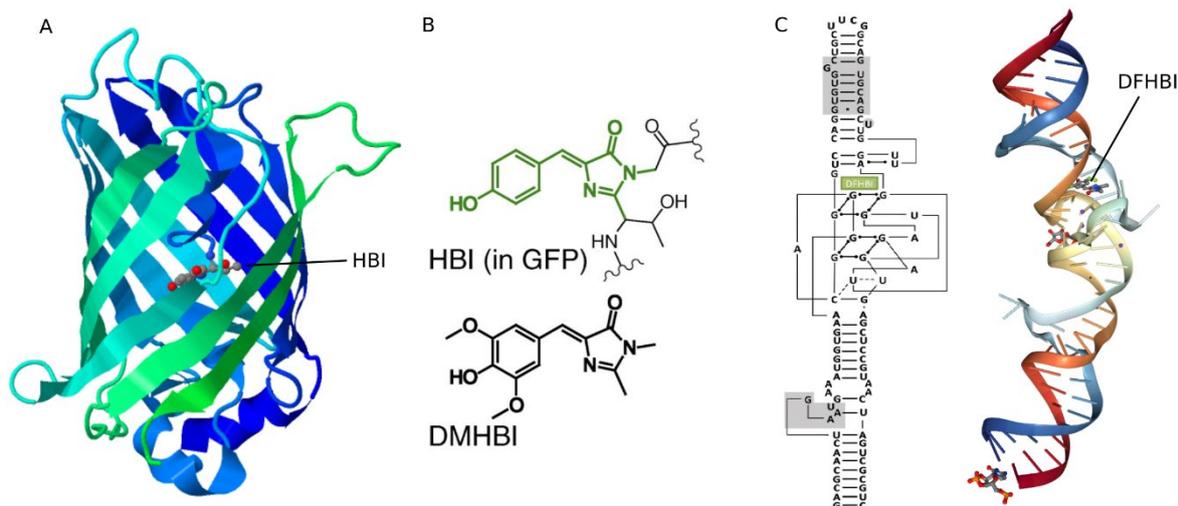
The most widely used technology for RNA detection is *in situ* hybridization by using oligonucleotide probes, which are generated from chemical synthesis or *in vitro* transcription. At first,  $^{32}\text{P}$  was used in the production of the radiolabeled single-strand DNA or RNA probes,

which allows target RNA molecules to be detected by autoradiography [Cox, DeLeon et al. 1984]. Later, biotin-labeled probes came out, together with fluorescence-labeled anti-biotin antibodies, RNAs can be observed by using a fluorescence scanner or microscope [Leary, Brigati et al. 1983]. So far, fluorescence dyes labeled probes have been widely used in RNA studies, and the more advanced fluorescence *in situ* hybridization (FISH) technologies make it possible to study the kinetics of a single RNA molecule *in vivo* [Swiger and Tucker 1996, Nath and Johnson 2000, Cabili, Dunagin et al. 2015].

The discovery and development of fluorescent proteins result in a great revolution in protein studies, with simple genetic operations, we can obtain labeled proteins and understand their activities in a convenient way. RNAs do not have intrinsic fluorescence, but fluorescence can also be transferred to them through non-covalent interactions [Armitage 2011]. One technology, which is based on the discovery of RNA-protein interaction from bacteriophage MS2, has been developed. Multiple copies of a specific stem-loop from MS2 RNA genome are inserted in the RNA molecule of interest, and the co-expression of the RNA and a GFP fused MS2 coat protein (MCP) in the same cell will result in the delivery of GFP to RNA and successful visualization [Bertrand, Chartrand et al. 1998]. Furthermore, in order to minimize the background, two parts of a split-GFP are fused to two different RNA binding proteins separately, only when the two RNA binding proteins recognize their target RNA motifs which are next to each other on target RNA molecules, the complete GFP can be assembled and generate fluorescence [Tyagi 2009].

In order to develop a genetically encodable RNA reporter technology, which does not rely on fluorescent proteins, a series of work has been done on artificial RNA motifs (aptamers), which can stably bind small, cell-permeant, non-fluorescent dyes, whereupon the bound dyes become fluorescent [Tyagi 2009]. The principle of aptamer induced fluorescence generation is based on the electronic configurations of these dyes [Tyagi 2009]. It is difficult for these dyes to form the correct conformations for fluorescence emission in the free condition unless they are captured and fixed in the cages of aptamers. [Babendure, Adams et al. 2003]. For example, a series of dyes share a common core structure that is similar to the 4-hydroxybenzylidene imidazolinone (HBI) chromophore of green fluorescent protein (GFP), which is constructed by the autocatalytic intramolecular cyclization of Ser<sup>65</sup>-Tyr<sup>66</sup>-Gly<sup>67</sup> during the maturation process (Figure. 7 A) [Zimmer 2002]. One of these dyes is named 3, 5-

dimethoxy-4-hydroxybenzylidene imidazolinone (DMHBI) and has a similar structure as HBI (Figure. 7 B). Its weak green fluorescence will be strongly enhanced when it binds the aptamers such as “Spinach” or “Broccoli” (Figure. 7 C) [Han, Leslie et al. 2013, Filonov, Moon et al. 2014]. What’s more, by modifying the side chains of DMHBI, several dyes with stronger fluorescence or different excitation and emission wavelengths are generated and can be widely used in RNA studies [Paige, Wu et al. 2011].



**Figure. 7 Structures of GFP, HBI, DMHBI, and spinach aptamer**

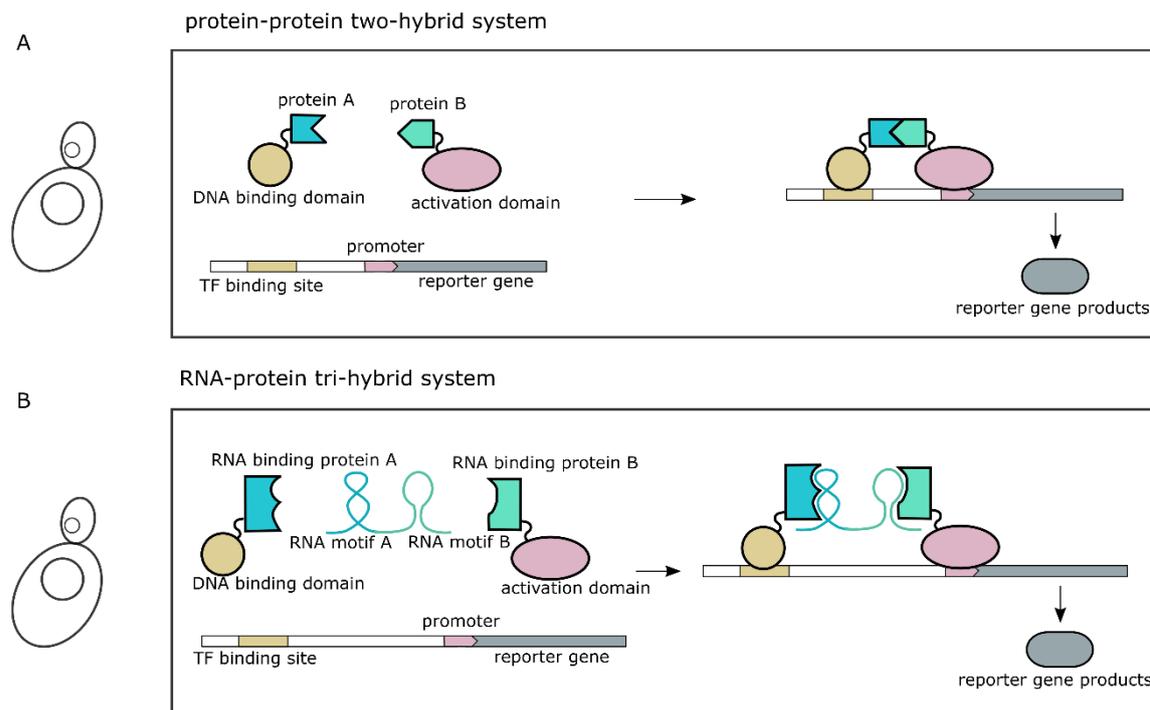
(A) The structure of the green fluorescent protein and its chromophore BHI (PDB 2AWK) [Wood, Barondeau et al. 2005]. (B) The structure of HBI and its synthetic analog DMHBI [Paige, Wu et al. 2011]. (C) The secondary and tertiary structures of spinach aptamer with DFHBI chromophore (PDB 4TS2) [Warner, Chen et al. 2014].

Some indirect methods are also used to deliver chemical fluorophores to target RNAs, which have specific aptamers. These kinds of dyes consist of one fluorescent group, one quenching group, and a flexible linker to hold them together [Johansson, Fidler et al. 2002]. At first, the activity of the fluorescent group is inhibited by the quenching group so that little fluorescence can be detected. When the fluorescent or quenching group binds to the RNA aptamer with high affinity, the inhibition within the dye will be eliminated, and the strong fluorescence can be detected [Johansson, Fidler et al. 2002]. Based on this principle, a series of “fluorescence turn-on” probes, which contain a fluorophore such as sulforhodamine B, and a quenching group such as nitrobenzylamine, trinitroaniline, dinitroaniline, and carbazole, have been developed (Figure. 8) [Sparano and Koide 2005, Murata, Sato et al. 2011]. These dyes are able



stable covalent bonds between RNAs and the proteins close to them, the complexes can be digested by RNase and then separated and partitioned by gel electrophoresis or other methods [Ule, Jensen et al. 2005]. By using the antibodies against the proteins of interest, we can pull down and obtain the RNA sequences that interact with these proteins, and this method is well known as crosslinking and immunoprecipitation (CLIP) [Jensen and Darnell 2008]. If the RNA molecules are labeled with specific tags such as ms2 or pp7, the MCP or PCP coupled beads can also be applied to isolate these RNAs and obtain their binding proteins [Yoon, Srikantan et al. 2012].

Although the *in vitro* RNA protein interaction studies can provide straightforward information, *in vivo* methods are still needed to observe how RNAs and proteins work in the cellular environment. The yeast tri-hybrid system, which is based on the protein-protein two-hybrid system, had been established and widely used in the detection and analysis of the interaction between RNA and protein [Putz, Skehel et al. 1996]. A typical protein-protein two-hybrid system consists of three elements: DNA binding domain fused test protein A, transcription activation domain combined test protein B, and a reporter gene, which is under the control of the transcription factor binding site (Figure. 9 A) [Fields and Song 1989]. If protein A and B interact, the DNA binding domain, and the transcription activation domain will be held together and enhance the expression of the reporter gene, by measuring the products of the reporter gene, we can deduce the binding strength between the two test proteins [Coates and Hall 2003]. In order to meet the RNA-protein interaction study, the yeast hybrid system has been upgraded to a three components version: potential RNA binding proteins are fused with DNA binding and transcription activation domains separately, a test RNA which is combined with a tag sequence that can be recognized by the known RNA binding protein is also added to the system (Figure. 9 B), as the result, if the test RNA can be identified by the potential RNA binding protein, the DNA binding and transcription activation domain will be held together and enhance the reporter gene expression [Putz, Skehel et al. 1996]. Similarly, by measuring the products of the reporter gene, we can deduce the binding strength between test RNA and protein [Venables, Ruggiu et al. 2001, Jaeger, Eriani et al. 2004].



**Figure. 9 Principles of the protein-protein and the RNA-protein hybrid system**

(A) Protein-protein two-hybrid system. Test proteins A and B are fused with DNA binding domain and activation domain separately, and if two test proteins interact, the DNA binding domain and the activation domain will be held together to activate the transcription of the reporter gene. (B) RNA-protein tri-hybrid system. The known and potential RNA binding protein A and B are fused with DNA binding and activation domain separately, while the test RNA motif B is combined with the motif A that interacts with protein A. So, if RNA motif B can interact with protein B, the DNA binding domain and activation domain will be held together to activate the expression of the report gene.

## 1.9 Aims of This Work

RNAs play essential roles in life, and most of them function as RNPs with different proteins. In order to understand how they work, it is critical and necessary to study the contact between the RNAs and proteins. Although various technologies have been established such as EMSA and CLIP to measure the RNA-protein interactions, both methods need the preparations of a large number of samples and are multi-steps operations. In general, they are material- and time-consuming. In order to overcome these weaknesses, convenient but reliable methods are still needed urgently.

Since the establishment of the yeast hybrid system, it has been widely used in protein-protein interaction studies, and the meaningful results obtained by this technology have propelled dramatic progression in life science. However, problems such as the deviation caused by the dependence of reporter gene products make the result not so accurate. In order to overcome these shortages, a novel tri-hybrid system, which depends on fluorescence instead of the reporter gene, had been established to detect the interaction between two proteins. With the help of a fluorescence microscope, protein-protein interaction can be measured according to the relative strength of two different color fluorescences in a small area in fixed or live cells in a convenient way.

The aim of this study is to establish a new RNA-protein interaction detector tool that is based on the fluorescent hybrid system, which allows studying the interaction between RNAs and proteins in an imaging-based way. In this system, a green fluorescent protein labeled MCP RNA trap is developed to capture the test RNA molecules which are marked with ms2 tags to the multiple Lac operon (*lacO*) array, which have been integrated into the genome of a BHK cell line, and the red fluorescent protein labeled potential RNA binding protein may also be recruited to the *lacO* array if it interacts with the test RNA. By measuring the relative intensity of two co-localized fluorophores at the same spot, we can deduce the interaction between test RNA and protein. In order to make this technology more reliable, several optimizations are applied to enhance precision. And to extend the range of application, different anchoring positions are tried to make the method usable in different cell lines, and various RNA traps with editable RNA binding proteins such as PUF and dCas13a are also constructed to extend the technique to any RNA molecule without additional modifications.

## 2 Materials and Methods



## 2.1 Materials

### 2.1.1 Primers for Cloning and Mutation

Name	Sequence
ms2-F BcuI	ggACTAGTtatactttaacgtcaaggagaaaaaac
ms2-R-NotI	agtcGCGCCGCgaatgtaagcgtgacataactaattacatg
2xPP7-F BamHI	acgtggatccAGAAAGGAGCAGACGATATG
2xPP7-R XhoI	acgtctcgagTTTCTAGGCAATTAGGTACCTTAG
HOTAIR-F BglII	agctAGATCTGTGCTCTGGAGCTTGATCC
HOTAIR-R BamHI	agctGGATCCccttaaaaaataaagacgccctc
H100-R MfeI	atcgcaattggggtgttgctgtgg
H200-R MfeI	atcgcaattgttctaaatccgttccattccac
H300-F XhoI	actgCTCGAGggagttccacagaccaaac
link-MCP-F BsrGI	gagctgtacaagGGCGGTGGCGGATCTGGCTACCCCTACGACG
MCP-R NotI	agtcgcgccgcTTAGTAGATGCCGGAGTTTGCTG
EGFP-U	GACGCCACCATGGGCGATCGCCAAAGAAGAAGAGAAAGGTCATGGTGAGCAA GGGCGAG
EGFP-D	GGGGGAATTCGTTAACTGCTTACTTGTACAGCTCGTCCATGC
lacI NotI R	acg tGC GGC CGC TTA CAG GCT GCT TCG GGA AAC
LaminB1-F AclI	acgtAACGTTGATGGGCAGCCGCGCTG
LaminB1-R BamHI	agctGGATCCTTACATAATTGCACAGCTTCTATTGGATGCTCTTG
dCas9-F AclI	CAGTAACGTTAATGGACAAGAAGTACTCCATTG
dCas9-R MunI	acgtCAATTGTTACACCTTCTTCTTCTTGG
Coilin-F Gib	tagcgctaccggactcagatcatggcagcttccgagac
Coilin-R Gib	CCGCCTAAGCTTGAAGCAGCggcaggttctgtacttgatgtg

2MCP-F Gib	GCTTAATGGTCAAACCAGTAATGCCAAAAAAGAAAAGAAAAGTTGGC
2MCP-R Gib	GATCAGTTATCTAGATCCGGTGTcAAGCAGCGGCCGC
dCas13a-F XhoI	agatCTCGAGATGCCAAAAAAGAAAAGAAAAGTTatgaaagtgcgaaggttaggag
dCas13a-R NotI	CACCGCCTAAGCTTGAAGCGGCCGCgttttcagactttttctcttc
mScarlet-F NotI	CTCCGGCATCTACGCGGCCGCTatggtgagcaagggc
mScarlet-R BamHI	CAGTTATCTAGATCCGGTGGATCCCTTAcagctcgtccatgcc
Ezh2-F XhoI	gactcagatcTCGAGctATGGGCCAGACTGGGAAGAAA
Ezh2-R KpnI	GGGCCCCGCGTtacctcAGGGATTTCATTTCTCG
Ezh2N-R BamHI	agctGGATCCCGGATGGTGGGGGTGCTG
Ezh2 345A-F	GAGCGTATAAAGACACCACCTAAAC
Ezh2 345A-R	GTTTAGGTGGTGCCTTTATACGCTC
Ezh2 345D-F	GAGCGTATAAAGACACCACCTAAAC
Ezh2 345D-R	GTTTAGGTGGTGTCTTTATACGCTC
PABPC1-F NheI	atccgCTAGCatgCCAAAAAAGAAAAGAAAAGTTaaccctcagtgccccag
PABPC1-R BamHI	GCCTGGATCCCGaacagttggaacaccggtggc
PUM2-F gib	CTACCGGACTCAGATCGATGCCAAAAAAGAAAAGAAAAGTTaatcatgattttcaagc tcttgc
PUM2-R gib	GAGGCCATGTCGACCGGTGGcagcattccatttggtgg
PCP V83Y F	GGACAGCTTCAAGTAGTCGGGGATGTC
PCP V83Y R	GACATCCCCGACTACTTGAAGCTGTCC
U6-F	gagagtgcacccatattgggtacAGGTCCGGCAGGAAGAG
oligo(A)-R	ACGTCTCGAGTTTCTAGGCAATTAGGTACCTTAG
U6 ms2-F	GGATCGGATGGATCCTAAGGTACCTAATTGC
U6 ms2-R	tcacacaggAAAAAACCGGTcGATCC
MCP S47R-F	ACAAAGTAACCTGTAGAGTTCGTCAGAGCTC

MCP S47R-R	GAGCTCTGACGAACTCTACAGGTTACTTTGT
U6-MSgRNA-F	GGCCTTTTGCTCAGAGGGCCTATTTCCCAT
U6-MSgRNA-R	AGCCTTATTTTAACTTGCTATTTCTAGCTC
U6-ms2-crRNA-F	gatccacatgtAGGTCCGGCAGGAAGAG
U6-ms2-crRNA-R	gatccacatgtgAAAAAAgATTCTAGAACTAGTGGATCCTAAGGTAGTTTTAG

### 2.1.2 Fragment Sources

Fragment	Plasmid	Catalog	Source/Reference
EGFP	pEGFP-C1	pc0592	Leonhardt lab
LacI	pGFPbinder-lacI	pc1398	Leonhardt lab
MCP	GEX2TMS2	pc1827	Leonhardt lab
ms2	p12xMS2-loop	pc1841	Leonhardt lab
dCas9	pCAG-dCas9	pc2946	Leonhardt lab Anton et al. Nucleus 2014
Lamin B1	pEGFP-LaminB1	pc1084	EMBL-Heidelberg Daigle et al. 2001
Ezh2	pCMV_RFP-mEzh2	pc2664	Leonhardt lab
pp7	pcDNA3.1(-)-PP7		Leonhardt lab
PCP	pHAGE-UBC NLS-HA-2XPCP-GFP		Leonhardt lab
U6 promoter	pEX-A-U6-tracrRNA	pc2979	Leonhardt lab
Coilin	pEhGFP-Coilin	pc0458	Leonhardt lab
PABPC1	pCI-MS2V5-PABPC1	65807	addgene Fatscher T et al. RNA 2014
Pum2	pFRT/FLAG/HA-DEST PUM2	40292	addgene

Hafner et al. Cell 2010			
NORAD	pcDNA3.1-NORAD	120383	addgene
Tichon et al. Nat Commun 2016			
Lbu-dCas13a	pDuBir-Lbu-dCas13a-avitag	100817	addgene
Tambe et al. Cell Rep 2018			
HOTAIR	LZRS-HOTAIR	26110	addgene
Gupta et al. Nature 2010			

## 2.2 Plasmid Construction

### 2.2.1 Clone PCR

DNA fragments for cloning are amplified using Phusion High Fidelity DNA Polymerase (Thermo Fisher). A standard 20  $\mu$ l reaction mixture is prepared as follows:

Component	Volume
Nuclease-free water	add to 20 $\mu$ l
5x Phusion HF Buffer	0.2 $\mu$ l
10 mM dNTPs	1 $\mu$ l
10 mM forward primer	1 $\mu$ l
10 mM reverse primer	1 $\mu$ l
Template	0.2 to 1 $\mu$ M
Phusion DNA Polymerase	0.2 $\mu$ l

After all the components are added, mix each sample gently and spin down. A PCR Mastercycler Pro thermocycler (Eppendorf) is applied for the amplification process. And the three-step program is set as follows:

Cycle step	Temperature	Time
Initial Denaturation	98°C	30 s
Denaturation	98°C	15 s

Annealing	According to primers	15 s
Extension	72°C	30 s per kb
Final Extension	72°C	10 min

### 2.2.2 Enzyme digest

Single or double digestion is applied with FastDigest endonuclease (Thermo Fisher). The standard fast digestion mixture is prepared as follows:

Component	Volume
Nuclease free water	up to 20 $\mu$ l
10x FastDigest Buffer	2 $\mu$ l
DNA	1 to 5 $\mu$ g
FastDigest enzyme	1 $\mu$ l each

Mix each sample gently and incubate at 37°C for 2 h. And in order to prevent the self-ligation of vector, 1  $\mu$ l of FastAP Thermosensitive alkaline phosphatase (Thermo Fisher), which is able to remove the 5' end phosphate, can be added to each reaction.

### 2.2.3 PCR products purification

The products of PCR reaction and enzyme digestion are purified by agarose electrophoresis. According to the size of products, agarose concentration can be chosen from 0.8% to 2%, and the process is generally applied with 100 or 120 V voltage for 20 to 30 min. After electrophoresis, cut target bands from the gel under UV. And in order to prevent damage, the UV exposure time should be as short as possible.

NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL) is used to extract nucleotides fragment from agarose gel. Transfer the Gel into Eppendorf tube, add NTI solution as 200  $\mu$ l per 100 mg gel, incubate at 50°C for 10 min, and vortex every 2 min until the gel dissolves completely. Load the sample to a binding column, place the column on a collection tube, centrifuge for 30 s at 11000 x g and discard the flow-through. Wash the binding column twice with 600  $\mu$ l Buffer NT3 followed by centrifuge for 30 s at 11000 x g, and dry the binding column

by centrifugation for 1 min at 11000 x g. To elute the DNA, place the binding column on a new Eppendorf tube, add 20  $\mu$ l of ddH<sub>2</sub>O and incubate at room temperature for 2 min, and then centrifuge for 2 min at 11000 x g. The eluted DNA can be used for the next experiments.

### 2.2.4 Ligation

New plasmids are constructed with T4 DNA ligase (Thermo Fisher). The standard reaction mixture is prepared as follows:

Component	Volume
Nuclease free water	up to 20 $\mu$ l
10x T4 Buffer	2 $\mu$ l
Vector DNA	100 ng
Insert DNA	3:1 molar ratio to vector
T4 DNA ligase	1 $\mu$ l

The reaction can be applied at room temperature for 1 h or 16°C for a longer time. The ligation products can be used for competent cell transformation directly.

### 2.2.5 Gibson assembly

Gibson assembly (NEB Gibson Assembly Master Mix) is a useful tool for cloning DNA fragments which lack suitable endonuclease site into vectors or constructing large fragments from small parts. For a typical 2-3 fragments assembly, each reaction contains 50-100 ng vector and inserts of a 3:1 molar ratio first, and then add the same volume Gibson Assembly Master Mix (2x). Mix the samples gently and incubate in a thermocycler at 50°C for 15 to 30 min and then keep at 4°C. After the reaction, samples can be used for competent cell transformation or stored at -20°C.

### 2.2.6 Competent cell preparation and transfection

Inoculate 5 ml of LB medium with *E. coli* JM109 strain and incubate overnight at 37°C. Use the overnight culture to inoculate 100 ml of LB medium at 30°C and let the cell grow until the absorbance at 600 nm is between 0.4-0.6. After chilling the culture on ice for 10 min, spin down the cell at 4000 rpm for 10 min, resuspend the cell in 100 ml of ice-cold CC buffer (10

mM Hepes, 15 mM CaCl<sub>2</sub>, 55 mM MnCl<sub>2</sub>·4H<sub>2</sub>O, 250 mM KCl), and then keep on ice for 10 min. Spin the cell again at 4000 rpm for 10 min at 4°C. Gently resuspend the cell in 18.4 ml of ice-cold CC buffer and add 1.4 ml of DMSO, incubate the cell on ice for 10 min, and distribute the cell in 50 µl of aliquots in Eppendorf tubes. Flash-free the cell in liquid nitrogen and store the cell in -80°C condition.

50 µl of chemically competent cells are used for each plasmid transformation. Take the competent cells from -80°C stocks and keep on ice, add 10 µl of ligation or Gibson product in each tube, mix gently by pipetting up and down and keep on ice for 15 min. After heat shock at 42°C for 60 s and keep on ice for 2 min, add 400 µl of LB medium to each tube and incubate in a shaker at 37°C, 180 rpm for 45 min. Then spread the cells on the LB agar plate with antibiotics and incubate overnight at 37°C.

### **2.2.7 Plasmid mini preparation**

Inoculate each clone into a sterile tube containing 2 ml of LB medium with antibiotics and incubate in a shaker overnight at 180 rpm and 37°C. Spin cells at 5000 x g for 5 min and discard the supernatant. Resuspend the cells with 100 µl of Solution I (50 mM Glucose, 25 mM Tris, 10 mM EDTA with 100 µg/ml RNase A in ddH<sub>2</sub>O, pH 8.0), then add 200 µl of fresh prepared Solution II (0.2 M NaOH, 1% SDS in ddH<sub>2</sub>O), mix well and incubate at room temperature for 5 min, and then add 150 µl of Solution III (5 M KAc in ddH<sub>2</sub>O, pH 4.8), mix well and incubate on ice for 10 min. Centrifuge the lysates at 10000 x g for 10 min and transfer the supernatant in a new Eppendorf tube. For plasmid sedimentation, add 1 ml of ice-cold 100% ethanol into the supernatant, mix well and spin down for 10 min at 10000 x g, 4°C, remove the supernatant, resuspend the pellet carefully with 1 ml of ice-cold 70% ethanol and then centrifuge for 10 min at 10000 x g, 4°C. Finally, remove the supernatant and dry the pellet at 50°C. The pellet can be resolved with ddH<sub>2</sub>O for future analysis.

### **2.2.8 Screening PCR**

Colony PCR is a convenient method for large scale clone screening. In order to minimize the false positive frequency, normally one primer on insert fragment and one primer on the backbone are needed to amplify the corresponding sequence. Pick each colony from the agar plate with tip and dissolve in 10 µl of ddH<sub>2</sub>O and use 1 µl as the template for PCR reaction. PCR reaction uses MyTaq Red Mix (Bioline). Each 20 µl reaction is prepared as follows:

Component	Volume
Nuclease-free water	add to 20 $\mu$ l
10x PCR Red Buffer	2 $\mu$ l
10 mM forward primer	1 $\mu$ l
10 mM reverse primer	1 $\mu$ l
Template	1 $\mu$ l
MyTaq DNA Polymerase	1 $\mu$ l

PCR is also performed with a PCR Mastercycler Pro (Eppendorf) thermocycler, and the program is set as follows:

Cycle step	Temperature	Time
Initial Denaturation	94°C	5 min
Denaturation	94°C	30 s
Annealing	According to primers	30 s
Extension	72°C	1 min per kb
Final Extension	72°C	10 min

PCR products are checked with agarose gel electrophoresis.

### 2.2.9 Plasmid midi preparation

Inoculate the correct clones in 200 ml of LB medium supplied with the specific antibiotic, and incubate in a shaker overnight at 37°C, 180 rpm, and then harvest the cell by centrifugation at 5000 x g, 4°C for 5 min.

NucleoBond Xtra Midi kit (MACHEREY-NAGEL) is used for plasmid midi preparation. Resuspend the pellet in 8 ml of Resuspension Buffer supplied with RNase A by pipetting the cells up and down or vortexing. Add 8 ml of Buffer LYS and mix gently by inverting the tube 5 times and incubate at room temperature for 5 min. At the same time, equilibrate the binding column filter with 12 ml of Buffer EQU. Next, add 8 ml of neutralization buffer NEU to the suspensions and immediately mix the lysates by inverting the tubes until the blue color

disappears completely, and add the mixture to the binding column filter. After the supernatant is empty from the column, wash the filter and binding column with 5 ml Buffer EQU. After that, remove the filter and wash the binding column again with 8 ml of Buffer WASH, elute the plasmid from binding column into a new 15 ml of Eppendorf tube with 5 ml of Buffer ELU, and then add 3.5 ml of isopropanol, mix well, and precipitate the plasmids at 15000 x g, 4°C for 30 min. Finally, wash the pellet with 70% ethanol and dry it at 50°C temperature. The plasmid can be resolved in ddH<sub>2</sub>O for later experiments.

## 2.3 Cell Manipulations

### 2.3.1 Cell culture

Hamster BHK cells are cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM) supplied with 10% fetal bovine serum and 50 µg/mL gentamycin at a 37°C and 5% CO<sub>2</sub> environment.

Human cervix epithelial Hela cells are cultured in high glucose DMEM supplied with 10% fetal bovine serum and 50 µg/mL gentamycin at a 37°C and 5% CO<sub>2</sub> environment.

Mouse muscle myoblast C<sub>2</sub>C<sub>12</sub> cells are cultured in high glucose DMEM supplied with 10% fetal bovine serum and 50 µg/mL gentamycin at a 37°C and 5% CO<sub>2</sub> environment.

Mouse embryo stem J1 cells are cultured in high glucose DMEM supplied with 16% fetal bovine serum, 10 U/ml penicillin/streptomycin, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol, 1% NEAA (v/v), 1 µM PD0325901, 3 µM CHIR99021 and 1000 U/ml LIF at a 37°C and 5% CO<sub>2</sub> environment.

### 2.3.2 Cell splitting

When the cells are approximately 80% confluent, remove the culture medium, wash the cells with room temperature DPBS, add enough warmed trypsin, and place the dish at 37°C for several minutes until the cells are detached from plate. Add 5 ml of culture medium into the plate, pipette up and down gently to break up the clumps of cells. Finally, dilute the cells into a new plate at a regular ratio.

### 2.3.3 Cell freezing

When the cells are approximately 80% confluent, remove the culture medium, wash the cells with room temperature DPBS, add enough warmed trypsin, and place the dish in 37°C for several minutes until the cells are detached from plate. Add 5 ml of culture medium into the plate, pipette up and down gently to break up the clumps of cells. Transfer all the cells into a 15 ml falcon tube, and pellet cells at 1200 rpm for 10 min. After that, remove the supernatant, gently resuspend the cells in freezing medium (1x culture medium with 10% DMSO (v/v)), divide cells into freezing tubes and store in -80°C.

### **2.3.4 Cell transfection**

Cell transient transfection is performed with Lipofectamine 3000 (Thermo Fisher). Cells need to grow until approximately 80% confluent to get the best transfection efficiency. Maximum 2.5 µg DNA can be used for the cells in each well of a 6-well plate. Dilute 5 µl of Lipofectamine 3000 Reagent in 125 µl of Opti-MEM medium and mix well, dilute DNA and 5 µl of P3000 Reagent in another 125 µl of Opti-MEM medium and mix well. Mix the two dilutions in a 1:1 ratio and incubate for 15 min at room temperature. Finally, add DNA-lipid mixture to cells, mix the medium gently, incubate overnight at 37°C, and then harvest the cells

### **2.3.5 Cell fixation and DAPI staining**

Cells are firstly cultured on coverslips and transfected with plasmids. When harvesting, remove the medium, wash the cells with PBS for 3 times, fix cells with 3.7% formaldehyde for 10 min at room temperature, and wash the cells again with PBS for 3 times. Stain the Cells with 1 µg/ml DAPI for 15 min at room temperature in a dark box. Finally, coverslips are mounted with Vectashield (Vector Laboratory) and sealed with nail polish.

### **2.3.6 Immunofluorescence**

Cells are firstly cultured on coverslips and transfected with plasmids. When harvesting, remove the medium, wash the cells with PBS for 3 times, fix cells with 3.7% formaldehyde for 10 min at room temperature, and wash the cells again with PBS for 3 times. Cells were permeabilized with 0.25% Triton X-100 in PBS for 10 min, then wash the cells with PBS 3 times for 5 min each. Next, block the cells with 1% BSA in PBST (PBS with 0.1% Tween 20) for 30 min. Dilute the primary antibody with 1% BSA, incubate the cell with the primary antibody in

a humidified chamber for 1 h at room temperature or 4 °C overnight, and then wash the cells with PBS 3 times for 5 min each. Dilute the fluorescence-labeled secondary antibody in 1% BSA, incubate the cells with secondary antibody for 1 h at room temperature in a humidified dark chamber, and then wash the cells 3 times with PBS for 5 min each. Cells can be counterstained with 1 µg/ml DAPI, and the coverslips are mounted with Vectashield and sealed with nail polish.

## 2.4 Imaging

### 2.4.1 Fixed cell imaging

Fixed cell imaging is applied with a confocal microscope (Leica SP5 and SP8). For SP5, a 405 nm diode laser is used for DAPI excitation, a 488 nm Argon laser is used for GFP excitation, a 561 nm diode-pumped solid-state (DPSS) laser is used for mCherry excitation, and a 594 nm HeNe laser is used for Alexa Fluor 594 excitation, and all the four emissions are detected by photomultiplier tube (PMT) sensors. For SP8, a 405 nm diode laser is used for DAPI excitation while the 488, 561 and 594 nm beams all come from a 470-670 nm White laser for the excitations of EGFP, mCherry and Alexa Fluor 594, the emission of GFP is detected by a PMT sensor while other three emissions are all detected by HyD sensors. The detection wavelength for each fluorescence is set as follows: DAPI (410 – 480 nm), EGFP (500 – 550 nm), mCherry/mScarlet (580 – 650 nm), and Alexa Fluor 594 (610 – 750 nm). A 63x objective is chosen for imaging. The linear sequencing method is performed as the default scan method, in which the acquisition of each fluorescence is independent and arranged sequentially. The mean fluorescence intensities of RNA trap and test protein in the same area of both anchor site and nucleoplasm are measured with ImageJ and the values of the relative fluorescence are calculated as the following formulation.

$$\frac{Red_{lacO} - Red_{nucleoplasm}}{Green_{lacO} - Green_{nucleoplasm}}$$

### 2.4.2 Live cell imaging

Live-cell imaging is carried out with an UltraVIEW VOX spinning disk confocal system (PerkinElmer) with a closed imaging chamber, which provides a 37°C, 5% CO<sub>2</sub>, and 60% air

humidity environment. A 63x oil objective is chosen for image acquisition. EGFP signal is activated by a 488 nm excitation laser and detected via a 521 nm emission filter, mCherry signal is excited by a 561 nm laser and detected via a 567 nm filter. FRAP program is set as follows: before bleaching, fluorescence is recorded every second for 3 seconds, after bleaching, fluorescence is recorded every 2 seconds for 20 seconds, then every 3 seconds for 30 seconds, and finally every 5 seconds for 75 seconds. Relative fluorescence is calculated in the same way as the fix cell imaging.

## 2.5 Gene Expression Measurement

### 2.5.1 Total RNA isolation

Total RNA isolation is performed with the NucleoSpin RNA kit (MACHEREY-NAGEL). Harvest and wash the cells with cold PBS twice. Add 350  $\mu$ l of Buffer RA1 and 3.5  $\mu$ l of  $\beta$ -mercaptoethanol to cell pellets and vortex vigorously. Place the filter on a collection tube, add the mixture, and centrifuge for 1 min at 11000 x g. Discard the filter and add 350  $\mu$ l of 70% ethanol to the homogenized lysate and mix by pipetting up and down a few times. Load the lysate to an RNA binding column, centrifuge for 30 s at 11000 x g, and place the binding column on a new collection tube. Add 350  $\mu$ l of MDB and centrifuge at 11000 x g for 1 min to desalt, add 95  $\mu$ l of DNase reaction mixture (10  $\mu$ l of rDNase in 90  $\mu$ l of rDNase Reaction Buffer), and then incubate at room temperature for 15 min to digest the remaining DNA. After that, wash the binding column with 200  $\mu$ l of Buffer RAW2 at 11000 x g for 30 s, 600  $\mu$ l of Buffer RA3 at 11000 x g for 1 min and 250  $\mu$ l of Buffer RA3 at 11000 x g for 2 min. Finally, elute RNA with 60  $\mu$ l of RNase free ddH<sub>2</sub>O and centrifuge at 11000 x g for 1 min.

### 2.5.2 Nuclear and cytoplasmic fraction separation

Harvest the cells and wash twice with cold PBS, remove the supernatant, and keep the pellet. Gently resuspend the cells in 500  $\mu$ l of 1x Hypotonic Buffer (20 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>) by pipetting up and down several times and incubate on ice for 15 min. Add 20  $\mu$ l of detergent (10% NP40) and vortex for 10 s. Centrifuge the homogenate for 10 min at 3000 rpm, 4 °C. Transfer and save the supernatant which contains the cytoplasmic fraction, and the pellet is the nuclear fraction.

### 2.5.3 Nuclear and cytoplasmic RNA isolation

Nuclear and cytoplasmic RNA isolation is also performed with the NucleoSpin RNA kit. For nuclear RNA isolation, the nucleus pellets can be applied directly with the total RNA isolation protocol. Cytoplasmic RNA isolation is carried on as follows. Full up the supernatant that contains the cytoplasmic fraction to 100  $\mu$ l with RNase-free H<sub>2</sub>O. Prepare a Buffer RA1 – ethanol premix with ration 1:1, to 100  $\mu$ l of sample add 600  $\mu$ l of Buffer RA1 – ethanol premix, and mix sample with vortexing. Load the mixture to an RNA binding column and continue with the total RNA isolation protocol.

### 2.5.4 cDNA synthesis

After checking the quality and concentration, the RNA needs to be reverse transcribed into cDNA. High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) is used to synthesize cDNA from RNA. First, prepare the 2x Reverse Transcription Master Mix as follows.

Component	Volume
10x RT Buffer	2 $\mu$ l
25x dNTP Mix (100 mM)	0.8 $\mu$ l
10x RT Random Primers	2 $\mu$ l
MultiScribe Reverse Transcriptase	1 $\mu$ l
RNase-free H <sub>2</sub> O	4.2 $\mu$ l
Total per reaction	10 $\mu$ l

Then prepare the 10  $\mu$ l of RNA solution, which contains maximal 2  $\mu$ g RNA. Mix the RNA solution and the 2x Reverse Transcription Master Mix to a 20  $\mu$ l mixture by pipetting up and down and then load the thermocycler. The program is set as follows.

Cycle step	Temperature	Time
Step 1	25°C	10 min
Step 2	37°C	120 min
Step 3	85°C	5 min

Step 4                      4°C                      ∞

The products can be used for RT-PCR or clone PCR.

### **2.5.5 RT-PCR**

RT-PCR is applied with a LightCycler 480 thermocycler (Roche) and LightCycler SYBR Green I Master Kit (Roche). 20  $\mu$ l of reaction mix contains 10  $\mu$ l of 2x SYBR Green I Master Mix and 10  $\mu$ l of diluted cDNA. The program consists of four stages includes a pre-incubation at 95°C for 5 min, 45 amplification cycles which contain two steps of 95°C for 10 s and 72°C for 20 s, 1 melting curve stage, and 1 cooling stage. And the  $-\Delta\Delta C_t$  method is used for data analysis.

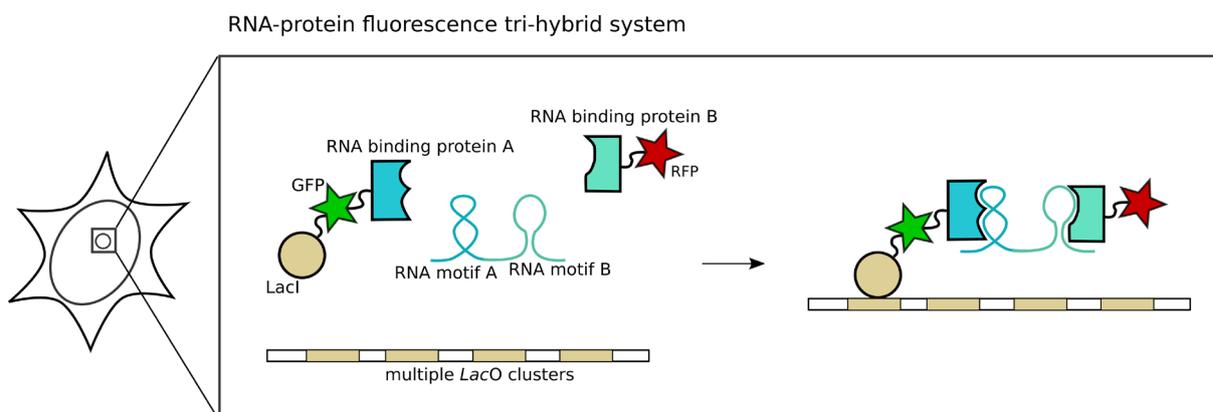
## 3 Results



### 3.1 Establishment of the Fluorescence Tri-hybrid Method

#### 3.1.1 Basic principle

Based on the fluorescence tri-hybrid method for protein-protein interaction assay and the yeast tri-hybrid method for RNA-protein interaction assay, we developed a convenient and reliable RNA trap based fluorescence tri-hybrid (F3H) system for RNA protein interaction assay, which depends on the capture of the target RNA to a *lac* operon (*lacO*) array integrated in the genome of a particular BHK cell line by a well-designed RNA trap (Figure. 10).



**Figure. 10 Basic principle of the RNA trap based RNA-protein tri-hybrid assay**

The system includes an RNA trap that consists of a Lac inhibitor, a GFP, and a known RNA binding protein A, a known RNA tag labeled test RNA and an RFP tagged test protein B, and all three components are in a *lacO* array integrated cell line. If the test RNA can interact with the test protein B, the test RNA-protein complex will be recruited at the *lacO* assay, and the co-localization of the green and red spots will be visualized.

The RNA trap consists of three parts: an MS2 coat protein (MCP), which can recognize the ms2 tags fused with the test RNA, a Lac inhibitor (LacI), which takes part in binding to the *lacO* array, and an enhanced green fluorescent protein (EGFP) between them (Figure. 11). The RNA trap, together with the target RNAs, can be visualized via the fused EGFP as a fluorescence spot at the *lacO* array in the nucleus under a fluorescence microscope. And the potential RNA binding protein, which is tagged with a red fluorescent protein, like mCherry, can also be recruited to the *lacO* array by the trapped RNA. The interaction between the test RNA and the test protein can be identified by measuring the fluorescence intensities of the co-localized green and red spots at the nuclear *lacO* array.

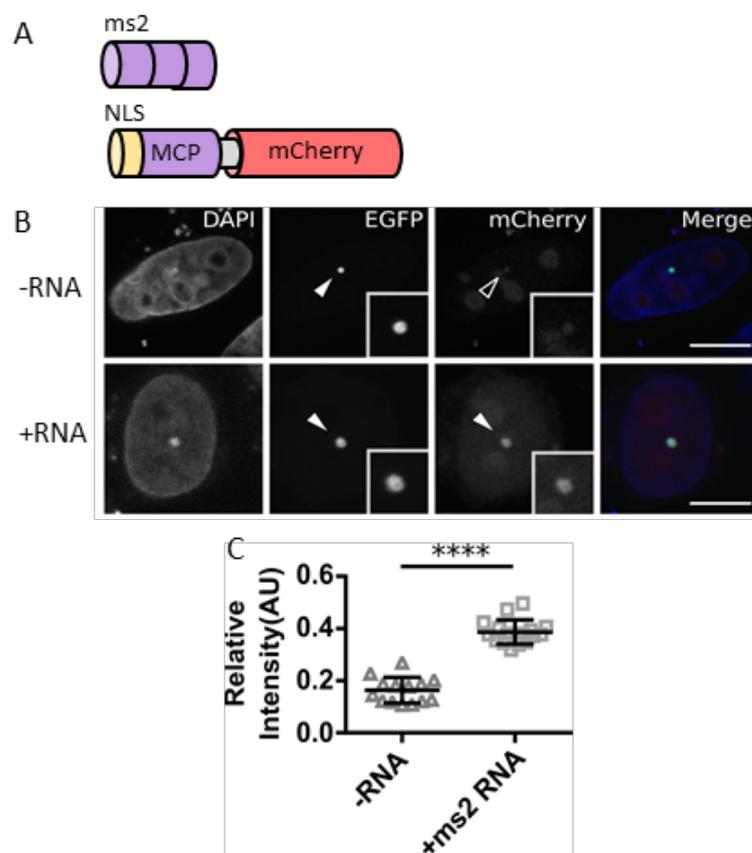


**Figure. 11** The structure of the MCP RNA trap

From the N terminus to the C terminus, the MCP RNA trap consists of four parts: NLS, MCP, EGFP, and LacI.

### 3.1.2 Measure the interaction between MCP and ms2 RNA

As a proof-of-principle, we decided to use a normal RNA-protein pair to test if the system we established works. We constructed an RNA of multiple ms2 stem-loops as the test RNA and cloned it into the CMV cassette of an expression vector, which is under the control of a cytomegalovirus (CMV) promoter and a Simian vacuolating virus 40 (SV40) poly(A) signal that allows the RNA to be transcribed in mammalian cells. Meanwhile, we chose the MCP as test protein, fused a mCherry label to its C terminus, and also cloned it into a mammalian expression vector with CMV cassette. Together with the RNA trap, they were transfected into the *lacO* array containing BHK cells to observe the co-localization. The results showed that in the cell without the ms2 RNAs, only green spots can be visualized in the nucleus, while in the cells co-transfected with the ms2 RNAs, the red spots appeared and co-localized with the green spots (Figure. 12).

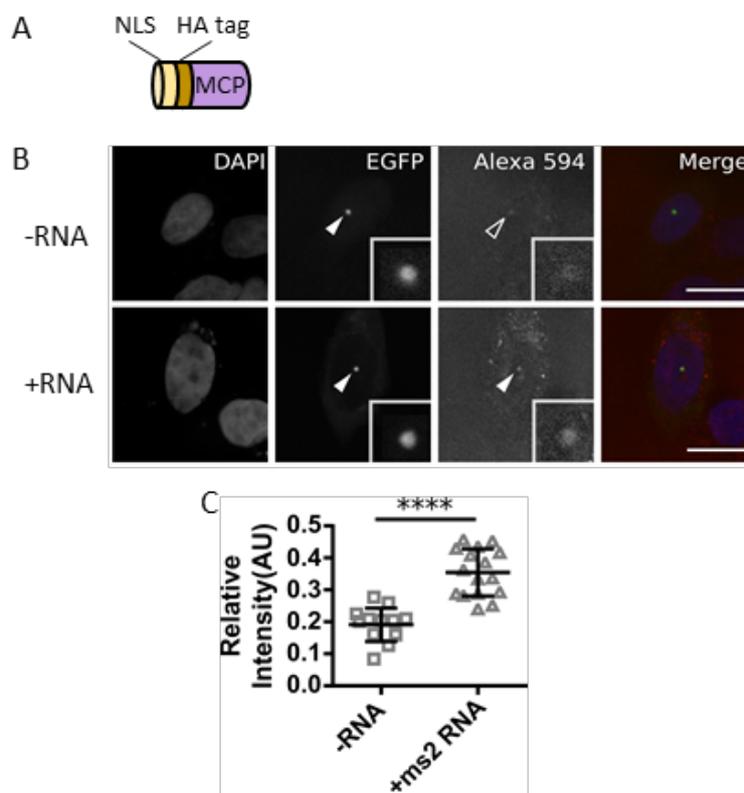


**Figure. 12 MCP proteins accumulate at *lacO* array**

(A) The structures of the ms2 RNA and the mCherry labeled MCP protein. The ms2 RNA contains 6 ms2 stem-loops, and the MCP protein is fused with an N-terminal NLS and a C-terminal mCherry tag. (B) Images of the ms2-MCP interaction test. With the ms2 RNA, the enhanced accumulation of the MCP protein can be visualized as the red spot co-localized with the green spot of the RNA trap at the multiple *lacO* loci (Bar: 10  $\mu$ m). (C) Quantification shows the dramatic enrichment of the MCP at the *lacO* array with the presence of the ms2 RNA (\*\*\*\*  $p < 0.0001$ ).

In order to verify the result, we also checked the enrichment of MCP with immunofluorescence (IF). A Human influenza hemagglutinin (HA) tag was added to the N terminus of the MCP, which allowed the protein to be detected by the anti-HA antibody. This HA labeled MCP was cloned in the expression vector and triple-expressed with the RNA trap and the ms2 RNA. Using immunofluorescence technology with anti-HA primary antibody and Alexa Fluor 594 labeled secondary antibody, we also observed the strong co-localization of the two fluorescence at the *lacO* spot, with an about two times higher MCP accumulation in

comparison to the cells that without the RNAs (Figure. 13), which again confirmed the interaction between the ms2 RNA and the MCP protein.



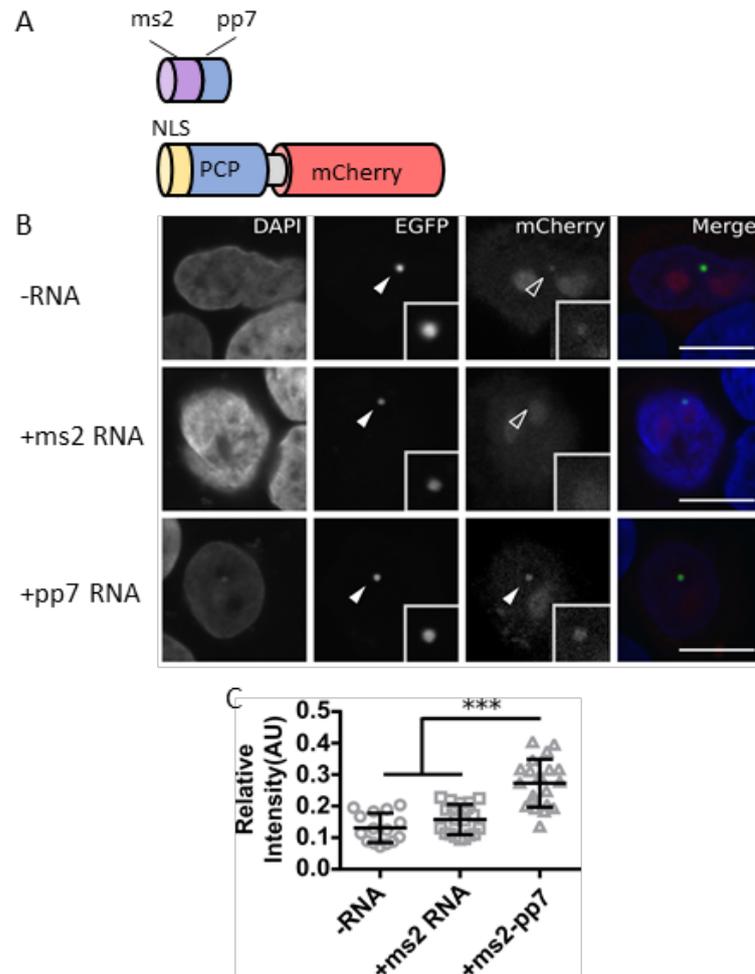
**Figure. 13 Immunofluorescence verifies MCP protein accumulation**

(A) The structure of the HA labeled MCP protein. An NLS and an HA tag are added to the N terminus of the MCP to construct the HA-MCP protein. (B) Images of the ms2-MCP interaction test with the HA-MCP. With the ms2 RNA, the significant aggregation of the MCP protein at the *lacO* array can be visualized by the anti-HA antibody (Bar: 10  $\mu$ m). (C) The quantification result demonstrates the ms2 RNA caused significantly improved recruitment of the MCP protein at the *lacO* array (\*\*\*\*  $p < 0.0001$ ).

### 3.1.3 Measure the interaction between PCP and pp7 RNA

Furthermore, we tested this strategy with another well-characterized RNA-protein pair. The stem-loop structure in the RNA genome could be recognized by the coat protein of bacteriophage PP7 for the precise assembly of virus particles, and this interaction pair has also been used in RNA labeling and tracking in a similar way as the ms2-MCP system. Here we tagged the pp7 RNA with ms2 stem-loops and labeled the PP7 coat protein (PCP) with a C-terminal mCherry. Both of them were cloned into the expression vector and under the control

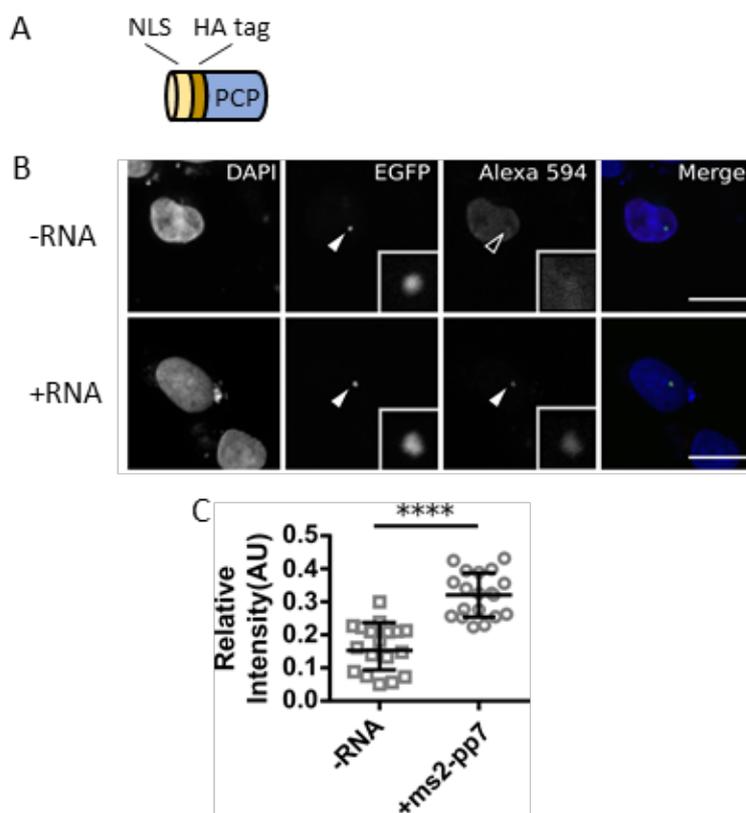
of the CMV cassette. When ms2-pp7 RNA exists, the red spots of the PCP protein appeared and co-localized with the green spots of the RNA trap at the *lacO* loci. However, without RNA, or with non-specific ms2 RNA, only green spots could be observed in the nucleus (Figure. 14).



**Figure. 14 PCP proteins accumulate at *lacO* array**

(A) The structures of the ms2-pp7 RNA and the mCherry labeled PCP protein. The pp7 RNA contains two ms2 stem-loops and two pp7 stem-loops, and the PCP protein is attached with an N-terminal NLS and a C-terminal mCherry tag. (B) Images of the pp7-PCP interaction test. The appearance of the red spot co-localized with the RNA trap revealed that the aggregate of the mCherry labeled PCP proteins at the *lacO* array can be significantly improved by the ms2-pp7 RNA, while non-RNA or irrelevant ms2 RNA cannot cause this phenomenon (Bar: 10  $\mu$ m). (C) Quantification of the relative fluorescence indicated that the ms2-pp7 can lead to a significant enhancement of PCP protein compared with the non-RNA or ms2 RNA (\*\*\*) ( $p < 0.001$ ).

Similar to the ms2-MCP test, we also constructed an HA tag labeled PCP to verify the results. Detected by the fluorescence-labeled antibodies, the co-localization of the RNA trap and the PCP could be visualized when the hybrid ms2-pp7 RNA exits (Figure. 15). Therefore, both the recombinant FP fusion and immunofluorescence results proved a successful detection of RNA-protein interaction with this RNA trapping strategy.



**Figure. 15 Immunofluorescence verifies PCP protein accumulation**

(A) The structure of HA labeled PCP protein. An NLS and an HA tag are added to the N terminus of the PCP to construct the HA-PCP protein. (B) Images of the pp7-PCP interaction test with the HA-PCP. When the ms2-pp7 RNA exists, the significant aggregation of the PCP protein at the *lacO* site can be visualized by the anti-HA antibody (Bar: 10  $\mu$ m). (C) The quantification result indicates the significantly enhanced recruitment of the PCP protein at the *lacO* array that is caused by the ms2-pp7 RNA (\*\*\*\*  $p < 0.0001$ ).

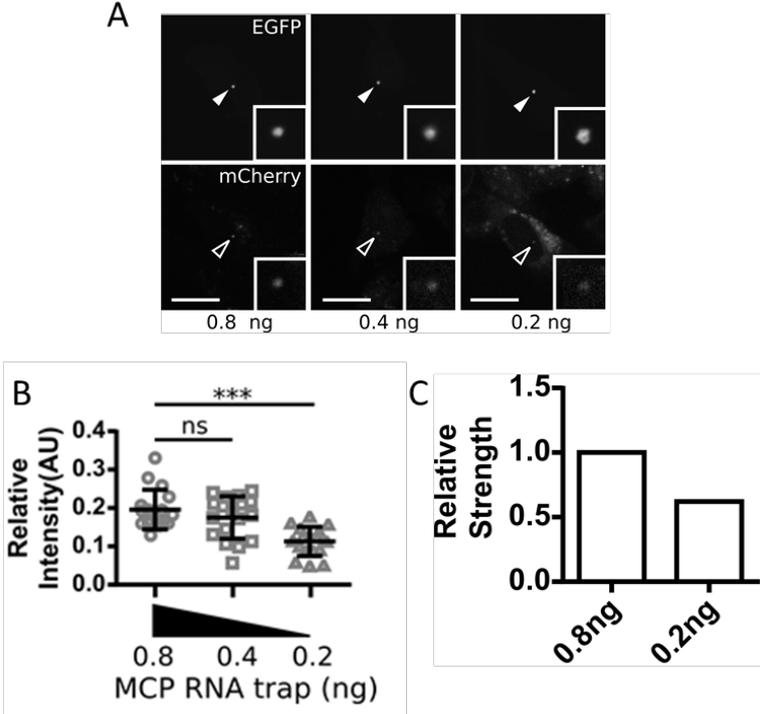
### 3.2 Optimizations of the Fluorescence Tri-hybrid Method

In order to establish a useful and reliable technology, we need to make sure that it can be used to detect the RNA-protein interaction sensitively and precisely. Therefore, a series of

attempts have been applied to optimize the performance of the method. In order to find out the best conditions for different targets, we performed all the optimization processes on both ms2-MCP and pp7-PCP pairs.

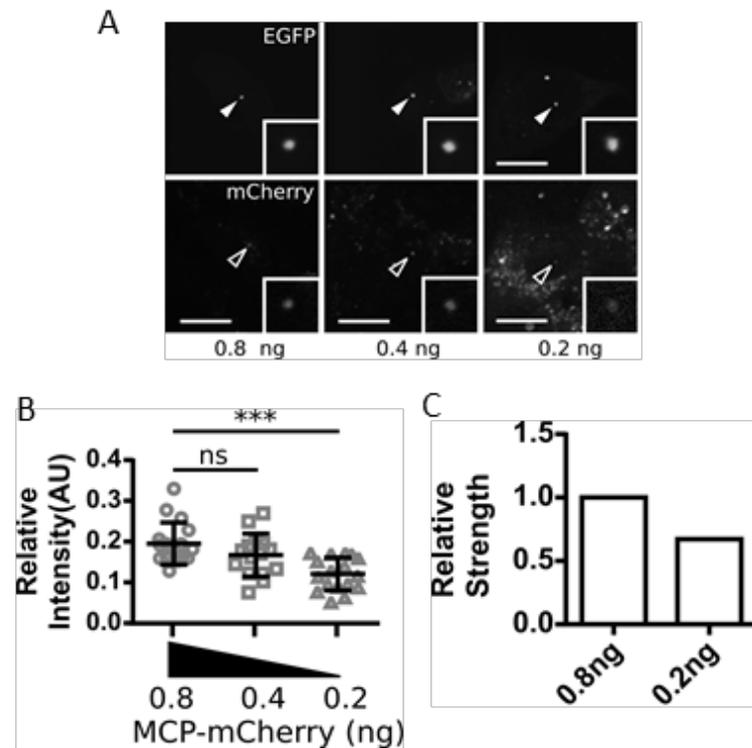
### 3.2.1 Optimization of the component ratio

As there are three parts (RNA trap, test RNA, and test protein) co-existing in this F3H system at the same time, different ratios among the three components will affect the recruitment of the test RNA and protein to the anchor site, and then influence the relative intensity of the two co-localized fluorescence and the interaction measurements. Therefore, the establishment of an appropriate ratio should be helpful to improve the sensitivity and precision of the method. With the constant amount of plasmid for RNA transcription, we adjusted the amounts of both RNA trap plasmids and test protein plasmids, and the background signals all showed positive correlations with the amount of the trap and the test protein. For the ms2-MCP test, the background signal decreased as the reduction of either RNA trap or MCP-mCherry plasmid from 0.8 ng to 0.2 ng, and compared with the original value, each of them led to a 40% and a 35% decrease (Figure. 16, 17). For the pp7-PCP test, the decrease of RNA trap plasmid caused slight reductions of both background and positive signals, to keep the positive/background ratio almost the same (Figure. 18, 19, 20). However, by decreasing the amount of PCP expression plasmid from 0.8 ng to 0.4 ng, the value of the non-RNA control had a 30% decrease while the positive signal had an 18% increase, to improve the positive/background ratio from 2.06 to 3.32 (Figure. 21, 22, 23).



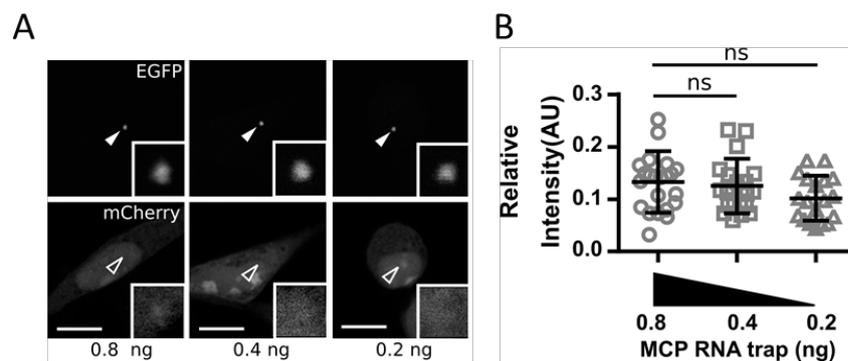
**Figure. 16** The amount of RNA trap affects the background of the ms2-MCP test

(A) Images of the backgrounds of the ms2-MCP interaction test with different amounts of the RNA trap. (B) Quantification indicates that the backgrounds have no significant changes when the amount of the RNA trap reduces from 0.8 to 0.4 ng but decreased when it continues decreasing to 0.2 ng (Bar: 10  $\mu$ m). (C) When the amount of the RNA trap changes from 0.8 ng to 0.2 ng, the final background is almost 60% of the origin (\*\*\*)  $p < 0.001$ ).



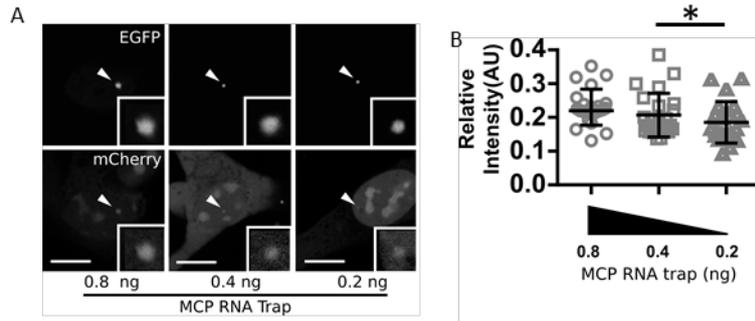
**Figure. 17** The amount of MCP protein affects the background of the ms2-MCP test

(A) Images of the backgrounds of the ms2-MCP interaction test with different amounts of MCP protein. (B) Quantification reveals that the backgrounds have no significant changes when the amount of the MCP protein decrease from 0.8 ng to 0.4 ng but decreased when it continues decreasing to 0.2 ng (Bar: 10  $\mu$ m). (C) As the amount of the MCP protein changed from 0.8 ng to 0.2 ng, the final background lefts nearly 65% of the origin (\*\**p* < 0.001).



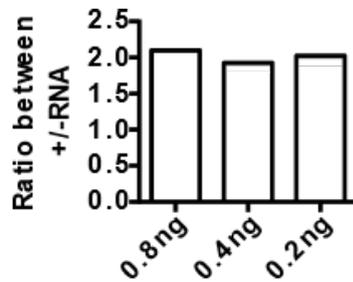
**Figure. 18** The amount of RNA trap affects the background of the pp7-PCP test

(A) Images of the backgrounds of the pp7-PCP interaction test with different amounts of RNA trap (Bar: 10  $\mu$ m). (B) Quantification demonstrates that although the amount of the RNA trap decreases from 0.8 ng to 0.2 ng, the background values do not show significant changes.



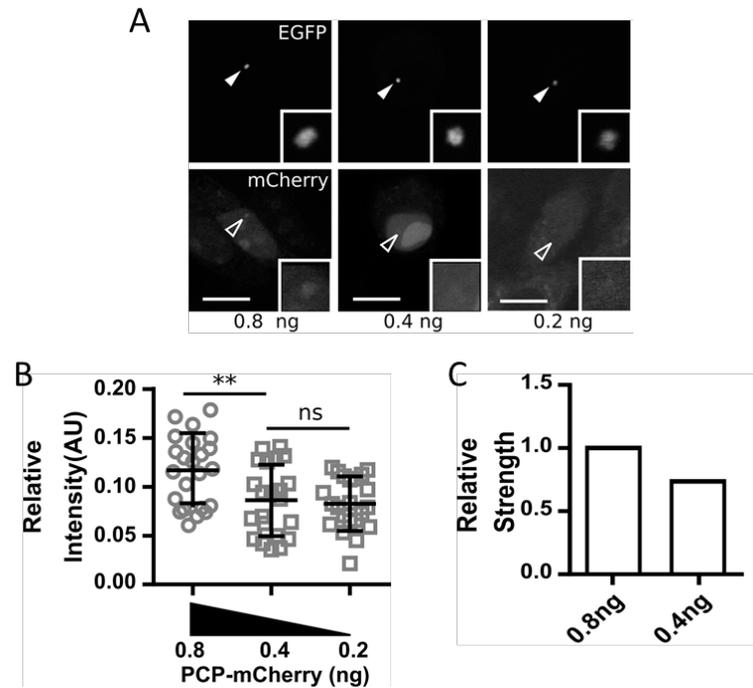
**Figure. 19 The amount of RNA trap affects the positive signal of the pp7-PCP test**

(A) Images of the positive signals of the pp7-PCP interaction test with different amounts of RNA trap (Bar: 10  $\mu$ m). (B) Quantification reveals that the positive signals have slight decreases as the reduction of the RNA trap from 0.8 ng to 0.2 ng (\*  $p < 0.05$ ).



**Figure. 20 The positive/background ratios of the pp7-PCP test with different amounts of RNA trap**

In the pp7-PCP test, both of the background and positive signals decrease along with the reduction of the RNA trap amount. With 0.8 ng, 0.4 ng, and 0.2 ng MCP RNA trap, the positive/background ratios are 2.1, 1.8, and 1.9.

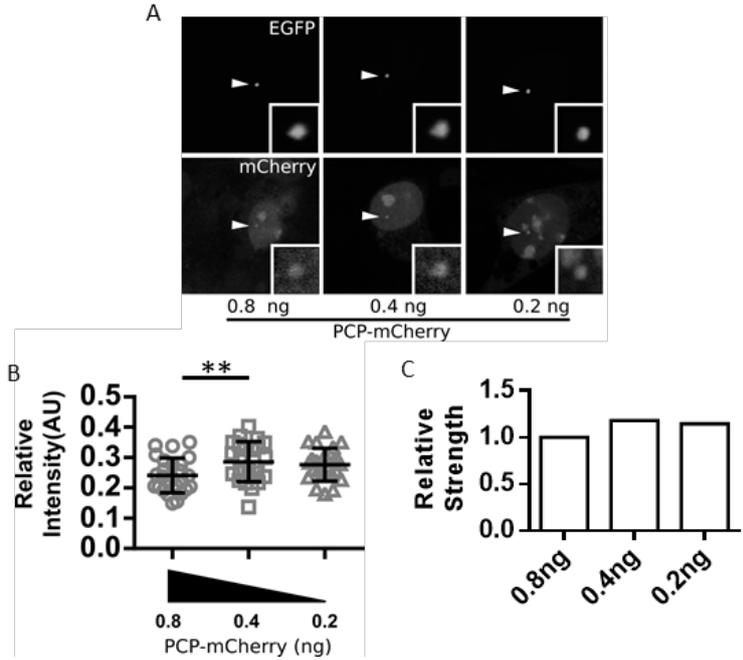


**Figure. 21** The amount of PCP protein affects the background of the pp7-PCP test

(A) Images of the backgrounds of the pp7-PCP test with different amounts of PCP protein (Bar: 10  $\mu$ m).

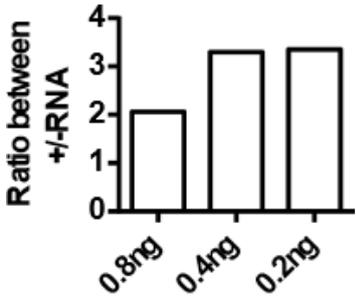
(B) Quantification indicates that when the amount of the PCP protein changes from 0.8 ng to 0.4 ng, but when the protein continues decreasing to 0.2 ng, the background almost keeps the same (\*\*  $p < 0.01$ ).

(C) Compared with 0.9 ng PCP protein, 0.4 ng PCP protein leads to a 30% background decrease.



**Figure. 22** The amount of PCP protein affects the positive signal of the pp7-PCP test

(A) Images of the positive signals of the pp7-PCP interaction test with different amounts of PCP protein (Bar: 10  $\mu$ m). (B) Quantification indicates that as the amount of the PCP protein decreases from 0.8 ng to 0.4 ng, the positive signal has an 18% increase. And when the PCP protein continued decreasing to 0.2 ng, the positive signal was almost the same as 0.4 ng PCP protein (\*\*  $p < 0.01$ ). (C) The 0.4 ng PCP protein causes an 18% higher positive signal than the 0.8 ng PCP protein.

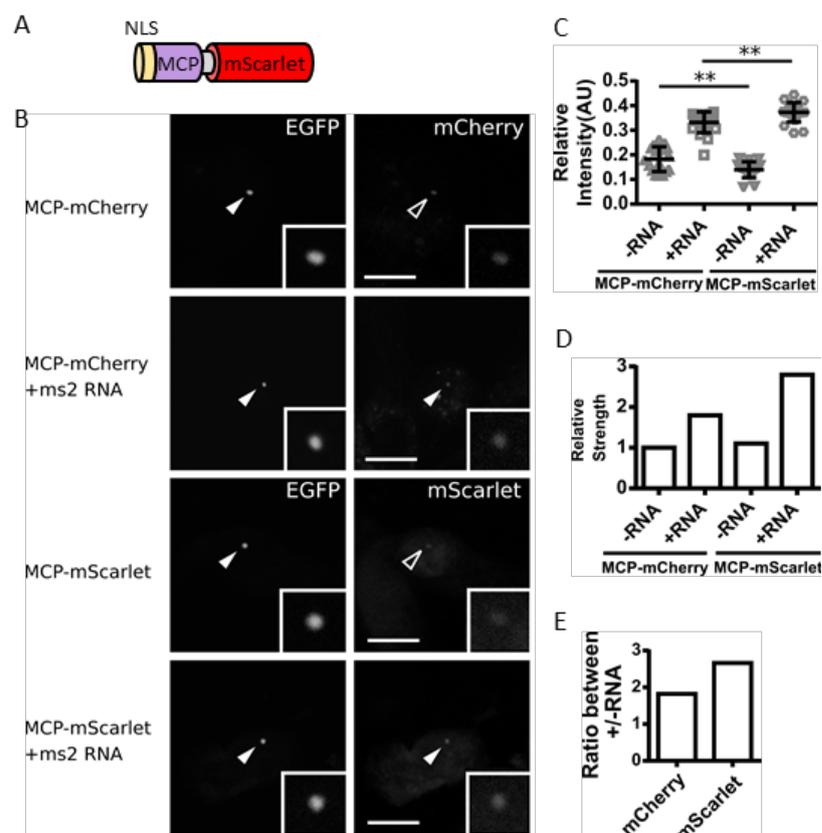


**Figure. 23** The positive/background ratios of the pp7-PCP test with different amounts of PCP protein

In the pp7-PCP interaction test, the background decreases while the positive signal increases along with the reduction of the PCP protein, which finally enhances the positive/background ratio. With 0.8 ng, 0.4 ng, and 0.2 ng PCP protein, the positive/background ratios are 2.0, 3.3, and 3.5.

### 3.2.2 Different fluorescent tags

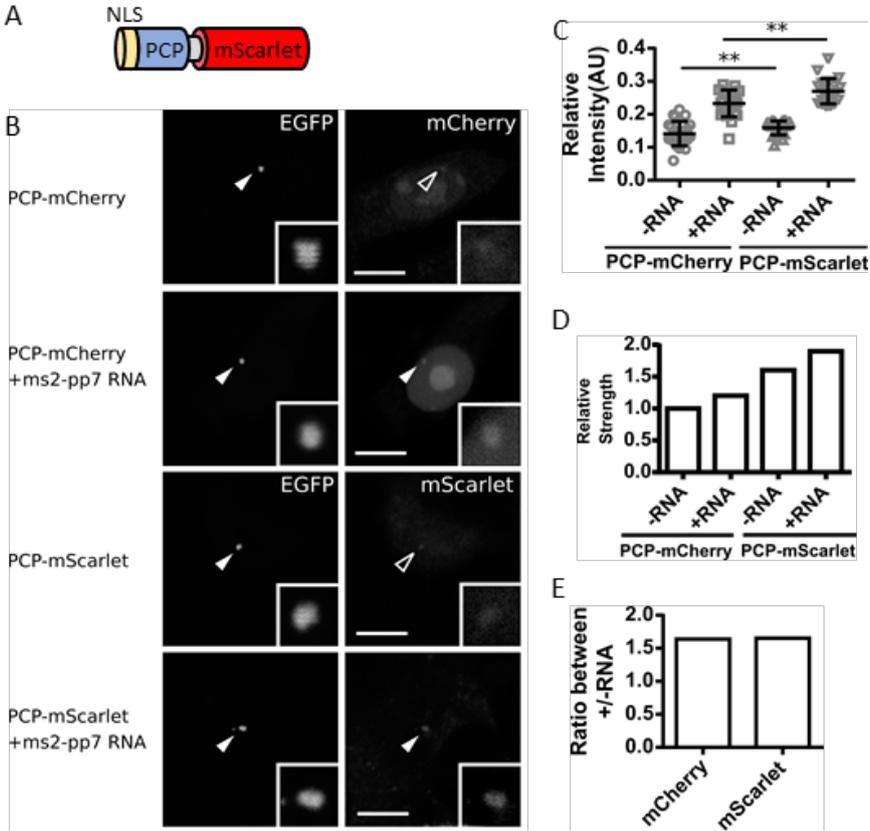
Based on previous observations, we tried to further enhance the sensitivity of this method. As we considered that the brightness of the fluorescent protein tag might affect both the background and positive signals, we constructed MCP or PCP proteins fused with mScarlet tag, which is another kind of red fluorescent protein but has a stronger brightness than mCherry. For ms2-MCP test, the mScarlet tagged MCP showed a 25% weaker background in the absence of the ms2 RNA and a 12% stronger positive signal when the ms2 RNA exists at the *lacO* spots than the mCherry tagged MCP, and finally resulted in the ratio of the positive/background increasing from 1.8 to 2.6 (Figure. 24). Meanwhile, for the pp7-PCP test, the mScarlet tagged PCP could lead to a 16% higher accumulation of red signal at the *lacO* array when the pp7 RNA exists, but the background also showed an 18% increase when there was no RNA, which makes the positive/background ratio almost the same as the mCherry tagged PCP (Figure. 25).



**Figure. 24** Different fluorescent tags affect the ms2-MCP test

(A) The structure of mScarlet labeled MCP protein. MCP-mScarlet protein is constructed by replacing the mCherry of the MCP-mCherry with mScarlet. (B) Images of the ms2-MCP interaction test with both

mCherry and mScarlet labeled MCP protein (Bar: 10  $\mu\text{m}$ ). (C, D) Quantification demonstrates that the mScarlet tag would decrease 25% background. Meanwhile, it can also increase 16% positive signal (\*\*  $p < 0.01$ ). (E) Compared with the mCherry tag, the mScarlet can raise the final positive/background from 1.8 to 2.6 in the ms2-MCP test.



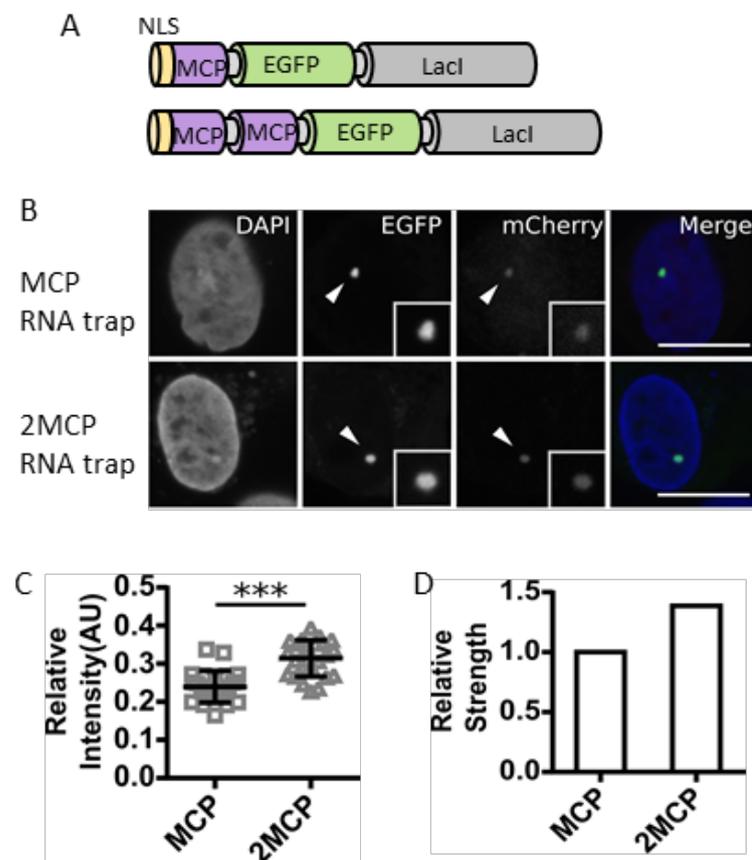
**Figure. 25 Different fluorescent tags affect the pp7-PCP test**

(A) The structure of mScarlet labeled PCP protein. PCP-mScarlet protein is constructed by replacing the mCherry of the PCP-mCherry with mScarlet. (B) Images of the pp7-PCP interaction test with both mCherry and mScarlet labeled PCP protein (Bar: 10  $\mu\text{m}$ ). (C, D) Quantification demonstrates that the mScarlet tag would increase 16% positive signal, however, it can also increase 18% background (\*\*  $p < 0.01$ ). (E) Both the mCherry and the mScarlet tags have similar positive/background ratios in the pp7-PCP test.

### 3.2.3 RNA capturing ability enhancement

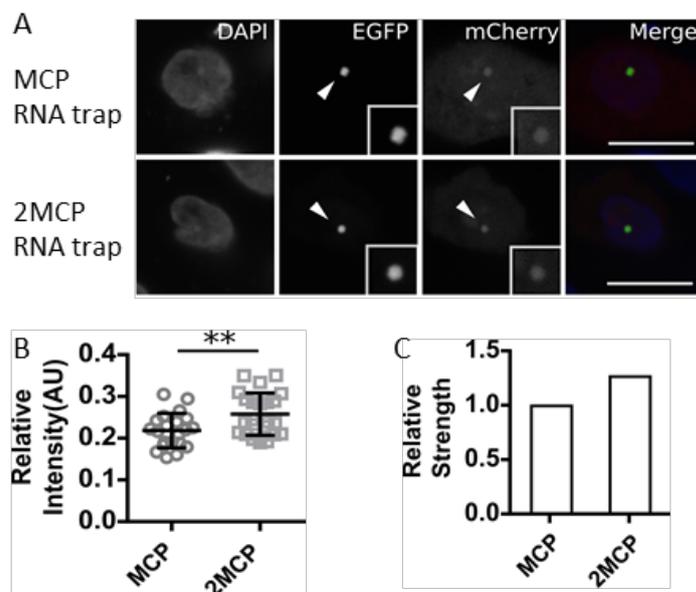
A simple but efficient way to enhance the positive signal is to enhance the RNA capturing ability of our RNA trap. As more test RNAs are anchored at the *lacO* site, more test proteins

will be recruited, and then the positive signal can be enhanced. In order to reach this goal, we doubled the ms2 binding unit in the RNA trap by adding an additional MCP in the RNA trap. Compared with the single MCP RNA trap, the double MCP RNA trap was able to cause a 38% stronger positive signal at the *lacO* array of the BHK cells in the ms2-MCP test, meanwhile, the positive signal of the pp7-PCP test also had a 27% increase (Figure. 17, 18), which indicates that the additional MCP can catch more RNA molecules and test proteins to the *lacO* array in order to enhance the positive signal.



**Figure. 26 Single or double MCP RNA trap affects the ms2-MCP test**

(A) The structures of single and double MCP RNA trap. The double MCP RNA trap is constructed by adding additional to the single MCP RNA trap. (B) Images of the ms2-MCP interaction test with RNA traps containing single or double MCP (Bar: 10  $\mu$ m). (C, D) Quantification indicates that the double MCP RNA trap can lead to a 38% stronger positive signal compared with the single MCP RNA trap (\*\*\*) ( $p < 0.001$ ).



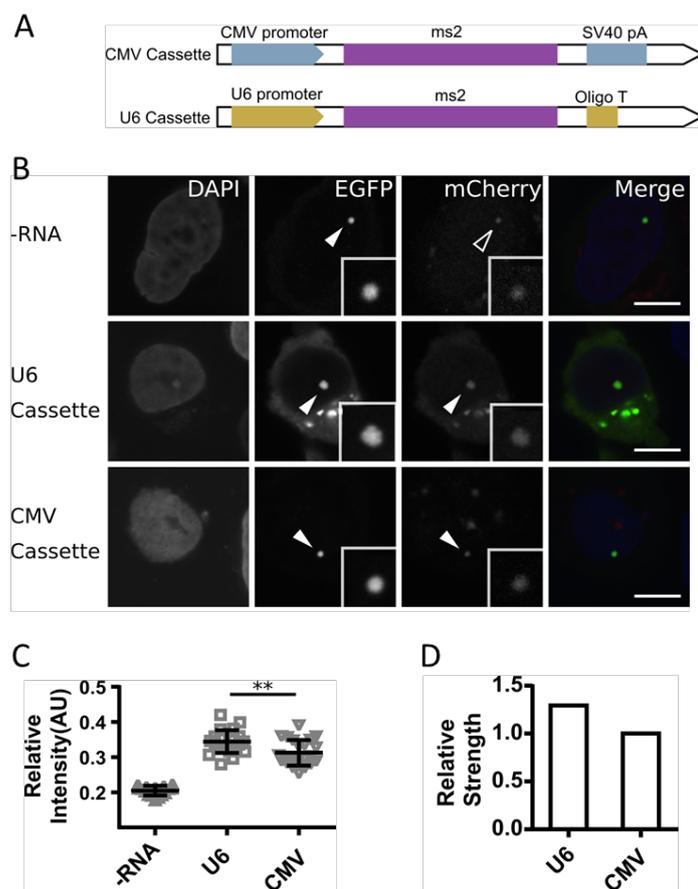
**Figure. 27** Single or double MCP RNA trap affects the pp7-PCP test

(A) Images of the pp7-PCP interaction test with RNA traps containing single and double MCP (Bar: 10  $\mu$ m). (B, C) Quantification indicates that the double MCP RNA trap leads to a 27% stronger positive signal compared with the single MCP one (\*\*\*) ( $p < 0.001$ ).

### 3.2.4 Different RNA expression cassettes

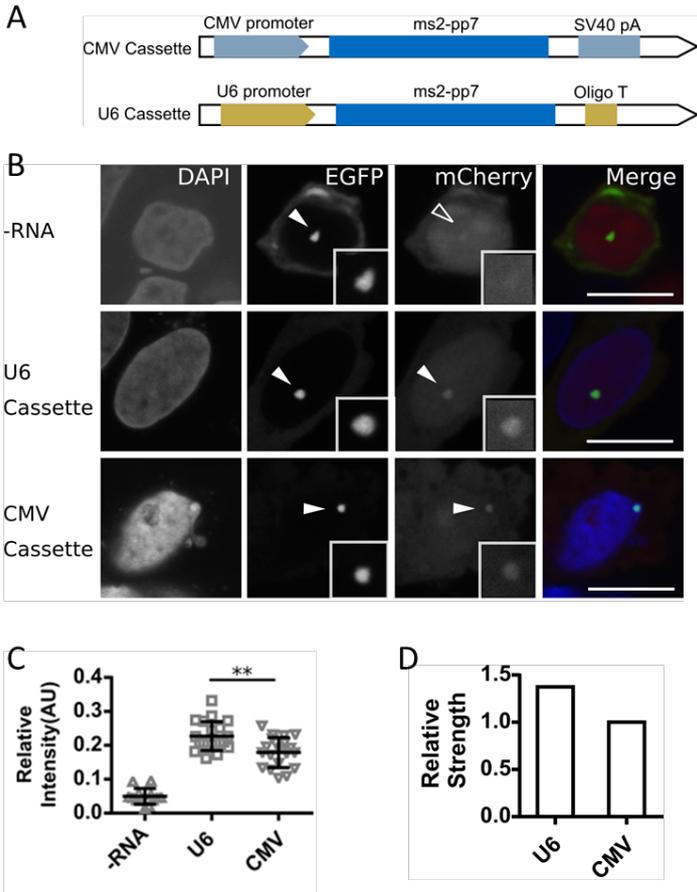
At the cellular level, localization and post-transcriptional processes will affect the concentration and distribution of the RNA, and then influence the performance of RNA-protein interaction assay. Hence, we compared two different RNA expression cassettes which have different transcription properties. One is the widely used CMV cassette, containing a CMV promoter and SV40 poly(A) signal, which can perform an RNA pol II-dependent transcription and is suitable for the expression of large RNA molecules. The other is the U6 cassette, including a U6 promoter, which is able to recruit RNA pol III, and an oligo(T) sequence, which is a stop signal for RNA pol III, and is usually for short RNA expression. Here we constructed the new RNA expression plasmids by cloning the ms2 RNA and ms2-pp7 RNA into the U6 cassette of the pEX-A vector. Image analysis illustrated that the RNA transcribed by the U6 cassette leads to better recruitment of the test protein at the *lacO* spot in both ms2-MCP and pp7-PCP tests. Compared with the CMV cassette, the U6 cassette led to a 29% higher positive signal in the ms2-MCP test and a 37% higher signal in the pp7-PCP test (Figure. 28, 29). Furthermore, in order to understand if the distributions of the test RNA will affect the

testing effect, we measured the concentrations of ms2-pp7 RNA in the total cell and nucleus. The results demonstrated that the RNA yields of the two expression cassettes were almost the same, but the concentration of the U6 cassette products in the nucleus was nearly twice as that of the CMV cassette products, which indicates that the U6 cassette products have a higher trend to stay in the nucleus (Figure. 30).



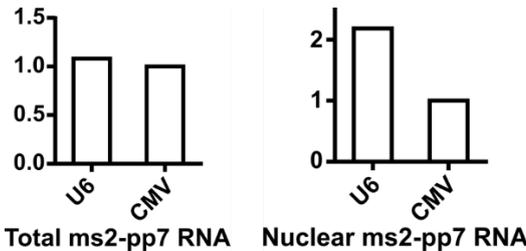
**Figure. 28 Different promoters affect the ms2-MCP test**

(A) The structures of CMV and U6 RNA expression cassettes. In the CMV cassette. The production of the ms2 RNA is controlled by the upstream CMV promoter and the downstream SV40 poly(A) signal, and in the U6 cassette, ms2 RNA synthesis is controlled by the upstream U6 promoter and the downstream Oligo(T) sequence. (B) Images of the ms2-MCP interaction test with CMV and U6 RNA expression cassettes (Bar: 10  $\mu$ m). (C, D) The quantification results reveal that the U6 cassette can lead to a 29% stronger positive signal compared with the CMV cassette in the ms2-MCP test (\*\*  $p < 0.01$ ).



**Figure. 29 Different promoters affect the pp7-PCP test**

(A) The structures of CMV and U6 RNA expression cassettes. In the CMV cassette, the production of the ms2-pp7 RNA is controlled by the upstream CMV promoter and the downstream SV40 poly(A) signal, and in the U6 cassette, RNA synthesis is performed by the upstream U6 promoter and the downstream Oligo(T) sequence. (B) Images of the pp7-PCP interaction test with CMV and U6 RNA expression cassettes (Bar: 10 μm). (C, D) The quantification results demonstrate that the U6 cassette can induce a 37% stronger positive signal compared with the CMV cassette in the pp7-PCP test (\*\* p < 0.01).



**Figure. 20 Different RNA expression cassettes affect RNA distribution.**

The ms2-pp7 RNA produced by U6 and CMV cassettes have a similar amount in the whole cell, but the U6 products have a one-fold higher concentration in the nucleus than the CMV products.

### 3.3 Measurement of RNA-protein Interaction

According to the previous results, the RNA trap works as we expect that can successfully indicate the binding between test RNA and test protein. Followed by several optimization steps, the specification and precision of the method have been raised. In the next stage, we will try to test it with different RNA samples to see if it works well as we expected.

#### 3.3.1 Characterization of the interaction between poly(A) sequence and PABPC1

The polyadenylation process is closely linked to RNA polymerase II. During the last stage of transcription, the polyadenylase, which cuts and polyadenylates at the poly (A) site in the 3' UTR of pre-mRNA, is recruited by the phosphorylated C-terminal domain (CTD) of RNA polymerase II. And the poly (A) tails are coated by the poly(A) binding protein C1 (PABPC1), which not only binds to poly(A) tail, but also is involved in RNA processes such as pre-RNA splicing and helps to keep the stability of RNAs (Figure. 31).

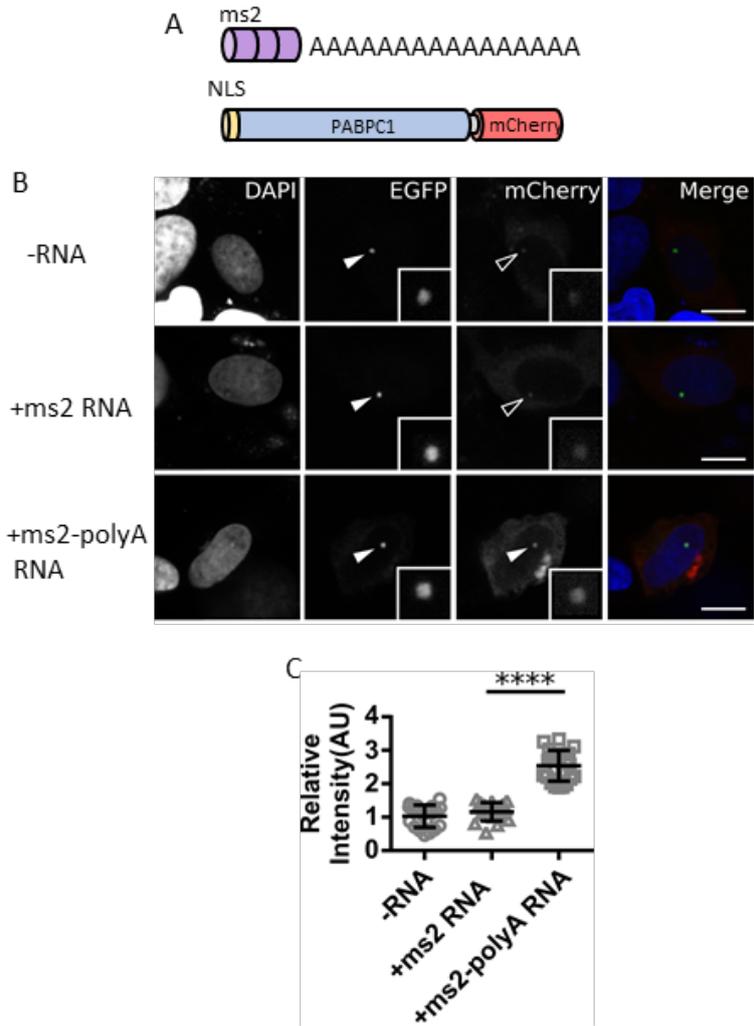


**Figure. 31 The binding between PABP and the poly(A) tail of mRNA**

The eukaryotic mRNA ends with a poly(A) sequence. The poly(A) sequence is normally covered with the PABP proteins, which not only help to keep the stability of the mRNA but also take part in mRNA processes such as transport and translation.

As each poly(A) sequence is covered by several PABPC1 molecules, the interaction between PABPC1 and poly(A) sequence can be used to test the function of our RNA trap. As the SV40 poly(A) signal in the CMV expression cassette is able to induce the poly (A) process, we simply cloned a short ms2 RNA into the cassette to construct the expression plasmid. At the same time, we cloned the PABPC1 protein, fused a mCherry label at C terminus, and inserted it into a CMV expression vector. We also used an ms2 RNA expression plasmid as control, in which

the ms2 RNA is under the control of the U6 cassette and will not be polyadenylated. All the plasmids were used for BHK cell transfection with the RNA trap. When using the ms2 RNA that is under the control of CMV, the enrichment of the red fluorescence of PABPC1 protein co-localized with the green fluorescence of the RNA trap could be visualized at the *lacO* array (Figure. 32), which indicated the accumulations of the PABPC1 proteins to the poly (A) RNAs at the *lacO* array and also verified the function of the RNA trap.

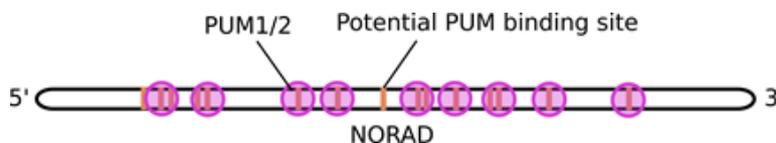


**Figure. 32 Poly(A) and PABPC1 interaction assay**

(A) The structures of polyadenylated ms2 RNA and labeled PABPC1 protein. A Poly(A) tail is added to the 3' end of the ms2 RNA by the eukaryotic polyadenylation mechanism caused by the SV40 poly(A) signal in the CMV cassettes. The PABPC1 protein is also fused with an N-terminal NLS and C-terminal mCherry. (B) Images of the interaction test of the polyadenylated or non-polyadenylated ms2 RNA and the PABPC2 protein at the *lacO* array (Bar: 10  $\mu$ m). (C) Quantification demonstrates that the aggregation of the PABPC1 protein can be significantly enhanced by the polyadenylated ms2 RNA (\*\*\*\*  $p < 0.0001$ ).

### 3.3.2 Characterization of the interaction between NORAD and PUM2

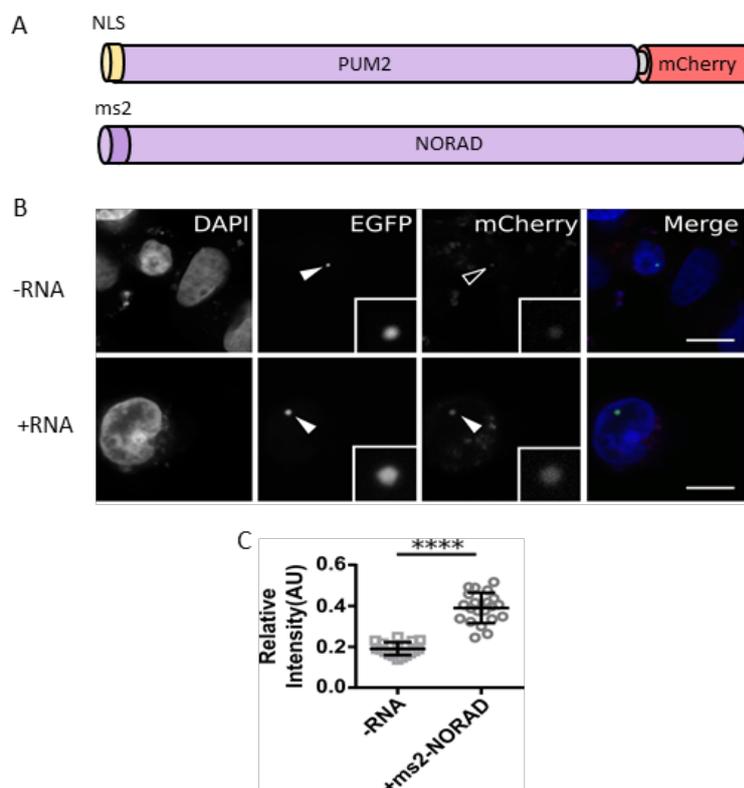
Non-coding RNAs (ncRNAs) play essential roles in cell processes, while most of them function together with their binding proteins. The NORAD (non-coding RNA activated by DNA damage) is one conserved lncRNA that exists in different species and is essential for the maintenance of genome stability (Figure. 33).



**Figure. 33 The binding between PUM1/2 and the NORAD RNA**

The NORAD ncRNA contains more than 10 potential PUM protein binding sites. It can attract free PUM proteins to control the cellular PUM concentration so that to regulate PUM associated cell processes.

Sequence analysis predicted that multiple PUM2 (pumilio RNA binding family member) protein binding sites exist in the NORAD. Therefore, we can use the new method to test the interaction between NORAD and PUM2. We obtained the full-length NORAD, added ms2 tags at the 5' end, and cloned it into a CMV expression vector. The PUM2 protein was labeled with mCherry at the C terminus and also cloned into a CMV expression vector. Without NORAD RNA, only the green signal can be detected at the *lacO* array in the BHK cells, and when the NORAD RNA was added, a strong enrichment of the red signal could be observed, which indicates the interaction between PUM2 protein and NORAD RNA (Figure. 34). And the result also suggested that the RNA trap can be used for interaction assay.



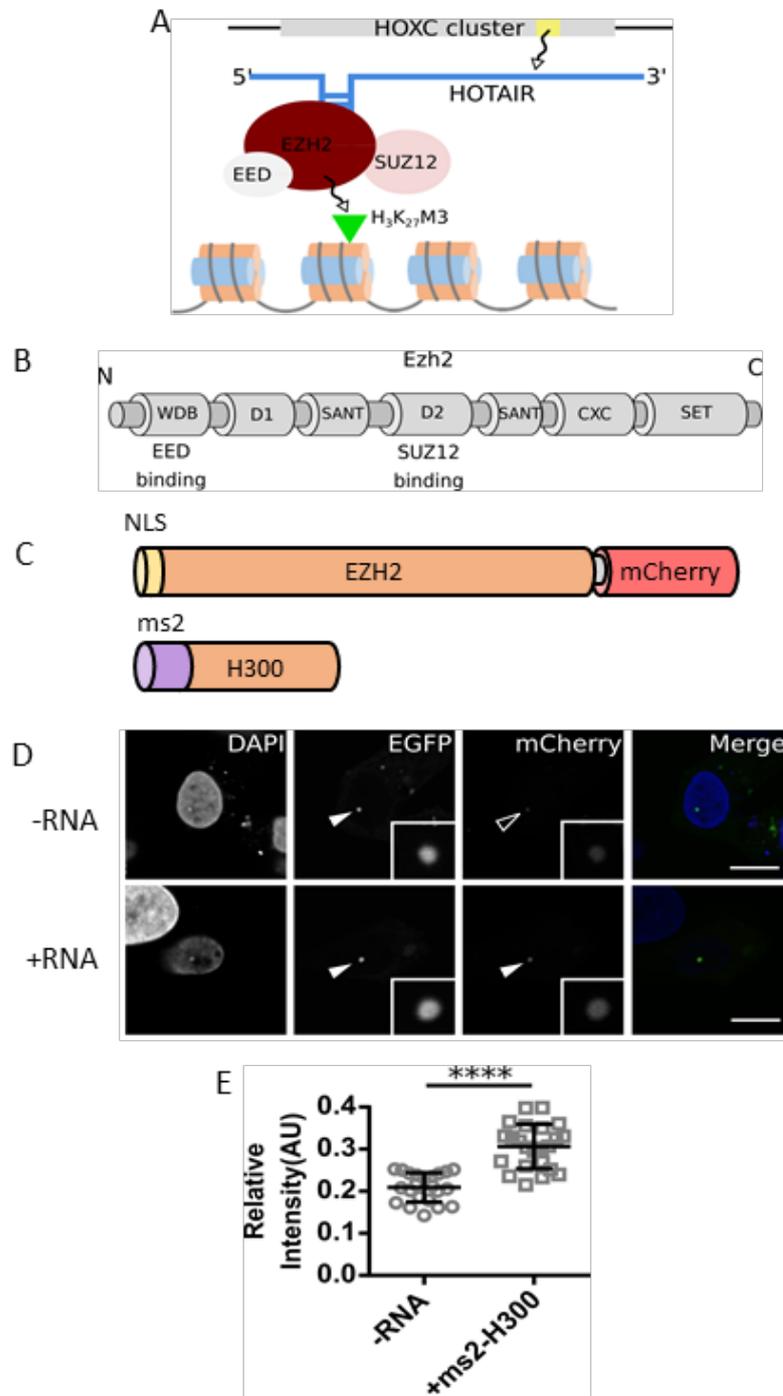
**Figure. 34 NORAD and PUM2 protein interaction assay**

(A) The structures of labeled NORAD RNA and PUM2 protein. NORAD RNA is labeled with 5' ms2 tags, and the PUM2 protein is fused with an N-terminal NLS and a C-terminal mCherry. (B) Images of the interaction assay of the NORAD RNA and the PUM2 protein (Bar: 10  $\mu$ m). (C) Quantification reveals that the recruitment of the PUM2 protein at the multiple lacO loci can be enhanced by the RNA (\*\*\*\*  $p < 0.0001$ ).

### 3.3.3 Characterization of the interaction between HOTAIR and EZH2

Non-coding RNA HOTAIR is the transcript from the HoxC loci and functions as a repressor that inhibits the expression of the Hox associated genes via the recruitment of the Polycomb Repression Complex 2 (PRC2), which is able to methylate the lysine at the 27<sup>th</sup> position of histone H3 followed by inducing the formation of silent chromatin. Previous studies have indicated that EZH2 (Enhancer of zeste homolog 2), one component of PRC2, can bind to the first 300-nucleotide sequence at the 5' terminus of HOTAIR. Hence, we cloned the 1-300 nucleotide sequence of HOTAIR 5' end (H300) and tagged it with ms2 stem-loops. We also cloned the EZH2 protein, fused with a mCherry label at its C-terminus. And both the RNA and the protein were cloned into the CMV expression vector. In the BHK cells, the stronger red

spots co-localized with the green spots showed that the accumulation of EZH2 protein at the *lacO* array is mediated by the H300 RNA (Figure. 35), and also indicated the interaction between the RNA and the protein.

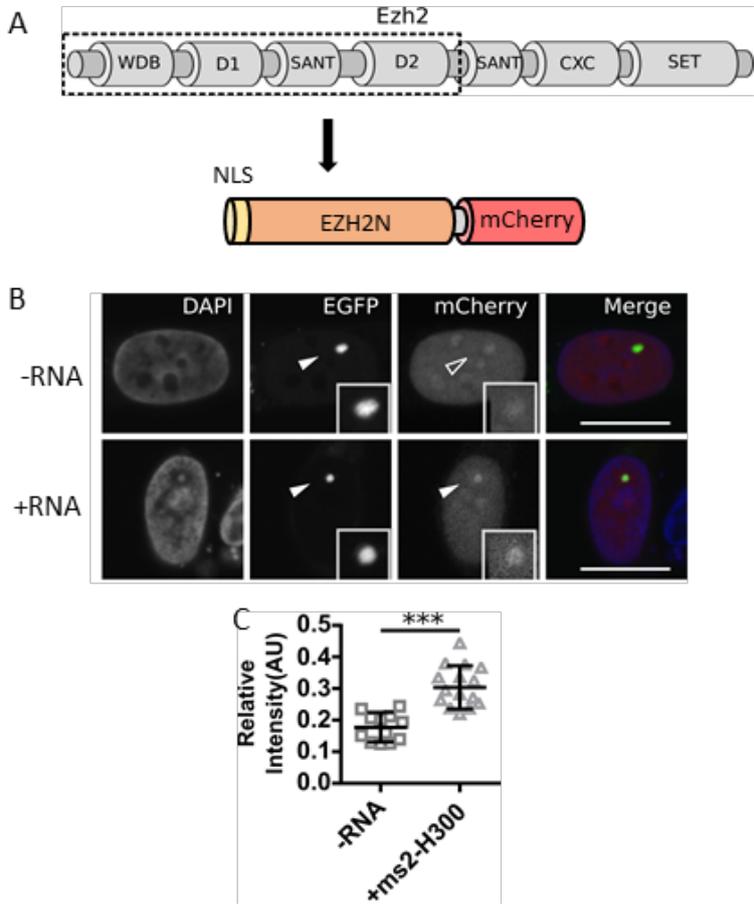


**Figure. 35 HOTAIR and EZH2 interaction assay**

(A) The model of HOTAIR generation and how it takes part in histone modification. The HOTAIR ncRNA is transcribed from the HoxC cluster. It is able to recruit the PRC2 complex, which can tri-methylate the lysine at the 27<sup>th</sup> position of histone H3 and then regulates gene expression. (B) The structure of

Ezh2 protein. From the N-terminus to the C-terminus, EZH2 protein contains the WDB domain, D1 domain, SANT domain, D2 domain, another SANT domain, CXC domain, and SET domain. (C) The structures of ms2 labeled H300 RNA and mCherry labeled EZH2 protein. ms2 tags are added to the 5' of H300 RNA, and the EZH2 protein is fused with an NLS at the N terminus and a mCherry at the C-terminus. (D) Images of the interaction assay of the H300 RNA and the EZH2 protein at the *lacO* array (Bar: 10  $\mu$ m). (E) Quantification indicates that the aggregation of the EZH2 protein can be significantly enhanced by the H300 RNA (\*\*\*\*  $p < 0.0001$ ).

Furthermore, several important residues for RNA binding have been suggested in the N-terminal helix of EZH2 [Long, Bolanos et al. 2017], so we constructed a shortened EZH2 (EZH2N) that only keeps the N terminus of EZH2 (1-370 AA). The EZH2N was also fused with a C-terminal mCherry and cloned into the expression vector. In BHK cells, the EZH2N protein acted similarly as the full-length EZH2 protein that can be recruited to the *lacO* loci by the H300 RNA (Figure. 36), which also confirms the proposition that the N-terminal part of the EZH2 takes part in HOTAIR binding.



**Figure. 36 HOTAIR and EZH2N interaction assay**

(A) The construction of mCherry labeled EZH2N protein. The N-terminal part of the full-length EZH2, which contains 1-370 AA, including the WDB domain, D1 domain, SANT domain and D2 domain, is cloned and fused with an NLS and a mCherry at the N and the C terminus separately. (B) Images of the interaction assay of the H300 RNA and the EZH2N protein (Bar: 10  $\mu$ m). (C) The quantification demonstrates that the aggregation of the EZH2N protein is significantly enhanced by the H300 RNA, and the EZH2N protein has a similar binding ability as the full-length EZH2 protein (\*\*\*)  $p < 0.001$ .

**3.4 Characterization of the Precise Binding Sites**

The measurement of whether the interaction exists between RNA and protein is only part of the functions provided by this RNA trap based method. Discovery of the exact protein binding sequence of the RNA and the RNA binding site of the protein are the two most important applications of this method.

**3.4.1 Detection of pp7 Binding site in PCP**

In order to make sure whether this method can be used for precise binding site detection, we first tested if the method can function with the standard pp7-PCP pair. We constructed a mutant PCP protein by replacing the valine (V) with tyrosine (Y) at the 83<sup>rd</sup> position of the wild-type PCP, tagged it with a mCherry at C terminus, and cloned it into the expression vector. According to the results of the interaction assay, both wild-type and mutant PCP illustrated the accumulations at the RNA trap spot, confirming the binding activities of both PCPs. The quantification results of the relative fluorescence at the *lacO* sites illustrated that the V83Y mutation leads to a 35% decreasing binding affinity compared with the wild-type (Figure. 37). The result not only confirmed the theory that the 83<sup>rd</sup> valine takes part in pp7 RNA binding [Lim, Downey et al. 2001, Chao, Patskovsky et al. 2008], but also demonstrated that this method could be used to find out the precise RNA binding sites of a protein.

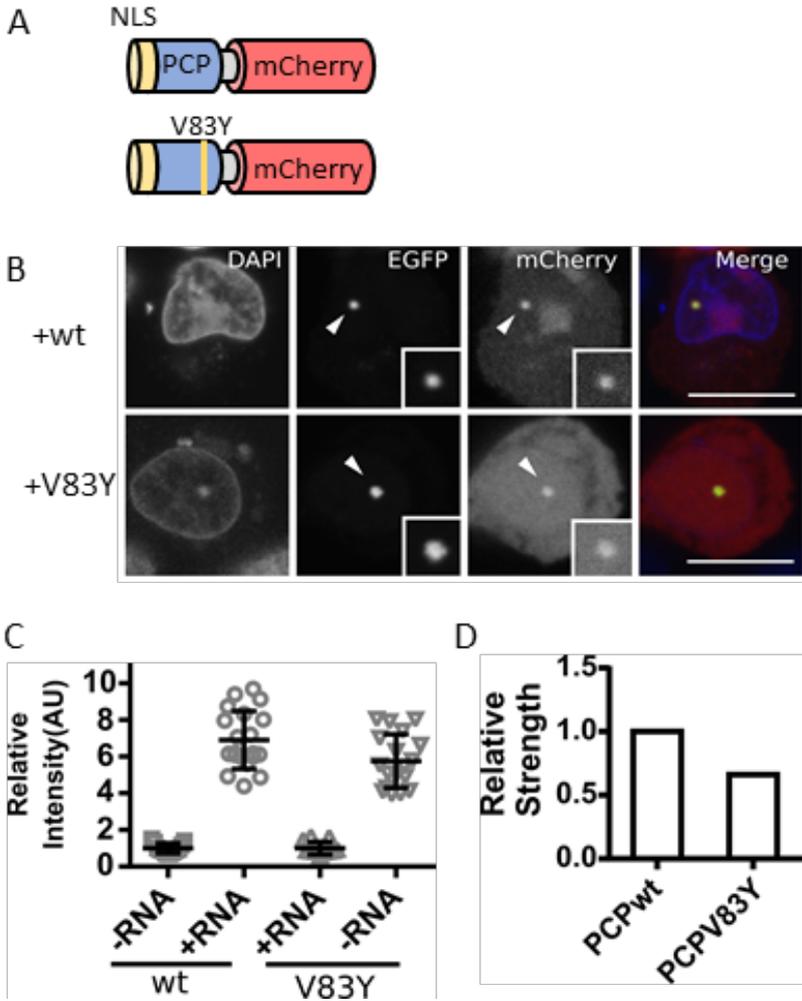


Figure. 37 Measurement of the pp7 RNA binding site in PCP protein

(A) The structures of mCherry labeled wild-type and mutant PCP proteins. The mutant PCP protein was constructed by replacing the valine with tyrosine at the 83<sup>rd</sup> position. (B) Images of the interaction assays of wild-type/mutant PCP protein and pp7 RNA at the *lacO* array (Bar: 10  $\mu$ m). (C, D) Quantification indicates that the recruitment of the mutant PCP protein by pp7 RNA is much weaker than the wildtype, and the value is only about 65% of the wildtype.

### 3.4.2 Detection of HOTAIR Binding site in EZH2

Some amino acids in the N-terminal part of EZH2 play critical roles in the interaction between EZH2 protein and its target RNA HOTAIR. As the phosphorylation of the threonine at the 345<sup>th</sup> position is considered to influence the interaction between EZH2 and HOTAIR, we mutated the threonine to alanine (T345A) or aspartic acid (T345D) to simulate the unphosphorylated and phosphorylated statuses respectively. Two mutant proteins were both fused with a

mCherry tag at the C-terminus and cloned into the expression vector separately, and co-transfected with the ms2-tagged H300 RNA and the RNA trap into the BHK cells. Both of the mutant proteins could be recruited by the RNA at the *lacO* array, but the quantification results of the relative fluorescence indicated that the unphosphorylated EZH2 T345A protein has a similar binding affinity as the wild-type protein to the H300 RNA, while the brightness of the EZH2 T345Y protein illustrates a significant enhancement (Figure. 38), which indicates a similar result as a previous study [Kaneko, Li et al. 2010].

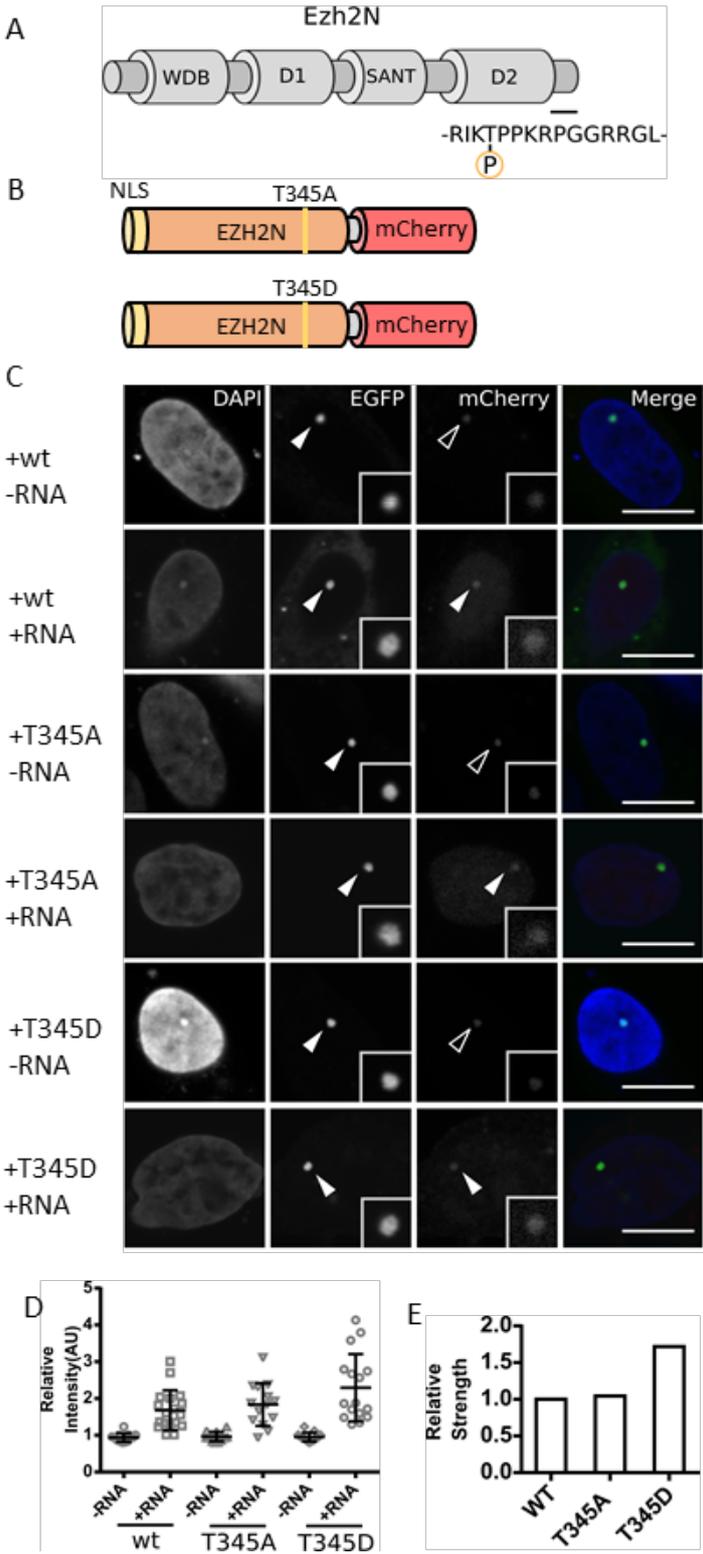


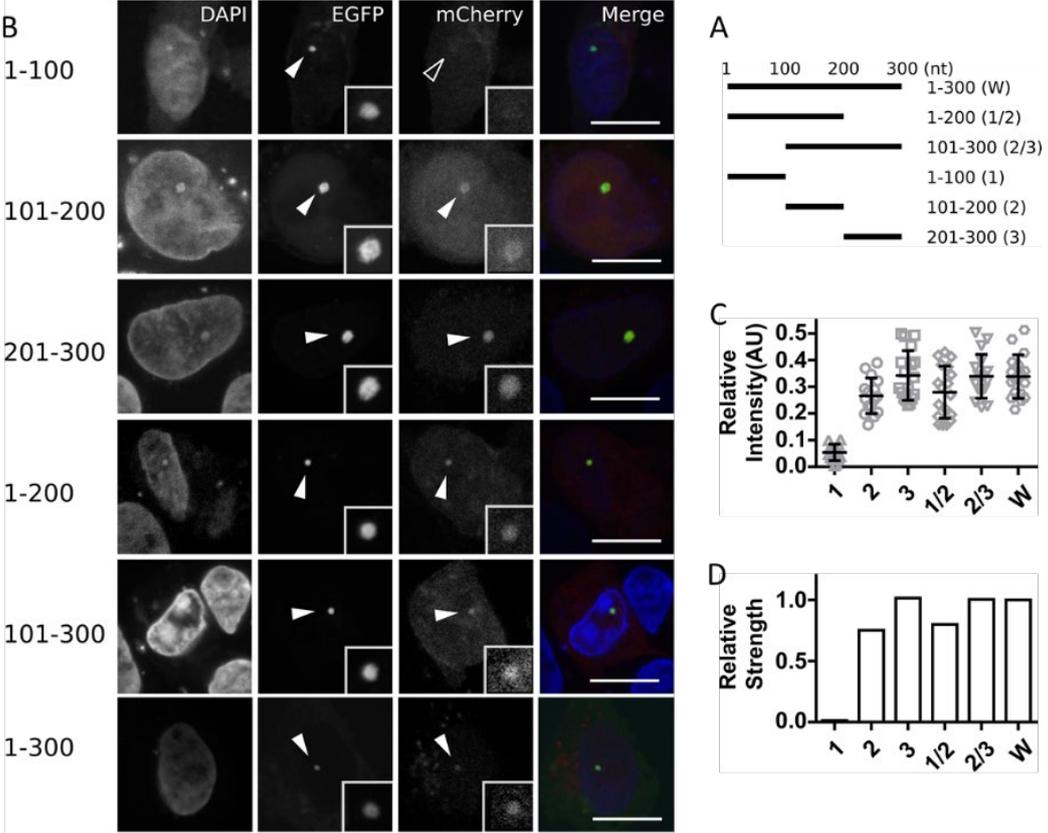
Figure. 38 Measurement of the HOTAIR binding site in the EZH2N protein

(A) The phosphorylation status of the threonine at position 345 of EZH2N is considered to affect the HOTAIR binding ability. (B) The constructions of two mutant EZH2N proteins. The threonine at the

345<sup>th</sup> position of EZH2N is replaced with alanine or aspartic acid. (C) Images of the wild-type/T345A/T345D EZH2N and H300 RNA interaction assays at the *lacO* array (Bar: 10  $\mu$ m). (D, E) Quantification demonstrates that the T345A mutant shows a similar binding ability as the wild-type protein, while the T345D mutant leads to a 70% higher value than the wild-type.

### 3.4.3 Protein binding sequence of HOTAIR

The 1-300 nucleotide sequence of the HOTAIR RNA has been demonstrated to be the binding area of the EZH2 protein [Tsai, Manor et al. 2010]. However, it is still a long sequence for protein recognition. To further narrow down and find out the accurate sequence that takes part in EZH2 binding, we decided to divide the HOTAIR 1-300 into five smaller fragments: first 100 nucleotides (H1), middle 100 nucleotides (H2), end 100 nucleotides (H3), first 200 nucleotides (H1/2), and end 200 nucleotides (H2/3), and examine their interactions with the EZH2N protein separately. We cloned these fragments into the expression vectors with 5' additional ms2-tags and then tested their binding abilities with the EZH2N-mCherry protein in the BHK cells. As shown in Figure. 39, all the fragments except H1 indicated the aggregations of the red fluorescence at the *lacO* array, which means H1 hardly interacts with the EZH2N while other fragments still keep the binding ability. Further quantification results demonstrate that the brightnesses of H3 and H2/3 fragments are similar and the closest to the value of the full-length H300 fragments, which means the 201-300 nucleotide part of the HOTAIR is the major binding region for the EZH2N protein. And the results also match the previous studies [Wu, Murat et al. 2013].



**Figure. 39 Approof the EZH2 binding sequence in the HOTAIR RNA**

(A) Structures of the H300 fragments. Five smaller HOTAIR fragments, which include 1-100 nucleotides, 101-200 nucleotides, 201-300 nucleotides, 1 -200 nucleotides, and 101-300 nucleotides, are cloned and fused with 5' ms2 tags. (B) Images of the interaction assays of five H300 fragments and EZH2N protein at the *lacO* array (Bar: 10  $\mu$ m). (C, D) The quantification results indicate that the H1 fragment loses almost all EZH2N binding ability, the H2 fragment keeps a 75% binding ability of the full-length H300, and the H3 fragment has a similar ability as the full-length H300. Meanwhile, the H1/2 fragment has an 80% binding ability of the full-length H300, and the H2/3 fragment also reveals a similar ability as the full-length H300 RNA.

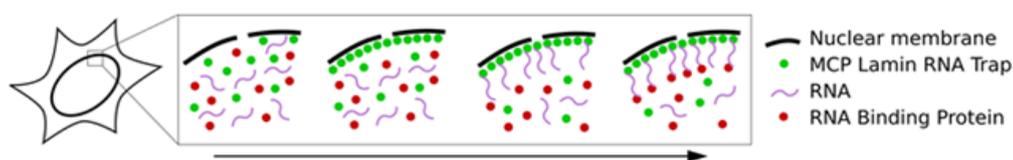
### 3.5 Development of Alternative Anchor Sites

The RNA trap based F3H system was designed to bring the test components to a small area within the cell to make it easier for fluorescence measurement. Until now, all the tests are performed in the specific BHK cell line with multiple *lacO* clusters. Although it is an excellent platform for the measurement, it also restricts the applications of the method. So it is

necessary to look for the new anchor sites existing in general cell lines that allow the method could be used widely.

### 3.5.1 Nuclear envelope

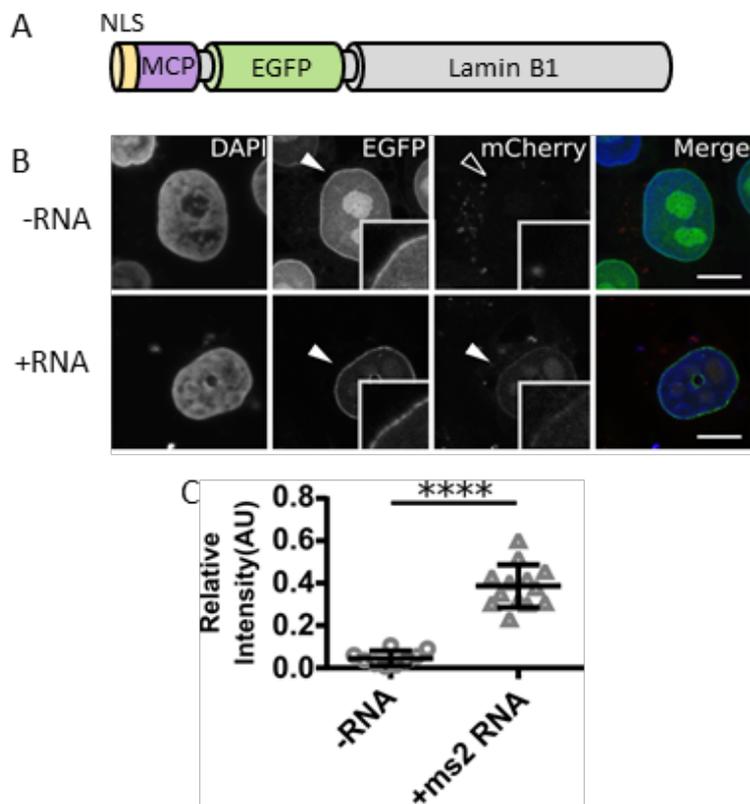
The eukaryotic nucleus is surrounded by the nuclear membrane, which is a complex two-layer structure constructed of lipids and protein molecules. The nuclear envelope lamina is a supramolecular protein complex associated with the nucleoplasmic surface of the nuclear membrane. As it is a dense polymer of lamin proteins, it may be a suitable location for our method for RNA accumulation. We approached this goal by applying a Lamin B containing RNA trap, which is constructed by replacing the LacI molecule of the RNA trap with a Lamin B1 protein (Figure. 40).



**Figure. 40 RNA-protein interaction at the nuclear membrane**

The principle of RNA-protein interaction assay at the nuclear membrane. The RNA trap, which contains the nuclear lamina localization molecule such as Lamin B1, has the ability to recruit the test RNA and the test protein to the inner side of the nuclear membrane for their interaction measurement.

We tested this Lamin B1 RNA trap with the ms2-MCP pair. In the HeLa cell, the RNA trap showed a successful localization at the nuclear membrane that the green RNA trap molecules aggregate and arrange around the nucleus. Furthermore, the red MCP proteins could also be detected that spreads at the same place as the RNA trap when the ms2 RNA exists (Figure. 41).

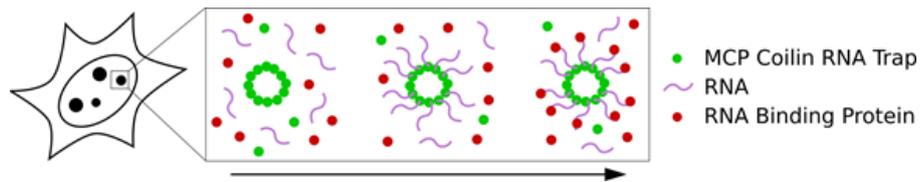


**Figure. 41 Detection of the ms2-MCP interaction at the nuclear envelope**

(A) The construction of the Lamin B1 RNA trap. The LacI in the MCP RNA trap is replaced with the Lamin B1. (B) Images of the ms2-MCP interaction assay with the Lamin B1 RNA trap at the nuclear envelope (Bar: 10  $\mu$ m). (C) Quantification indicates that the ms2 RNA can significantly enhance the accumulation of the MCP protein at the nuclear envelope (\*\*\*\*  $p < 0.0001$ ).

### 3.5.2 Nuclear body

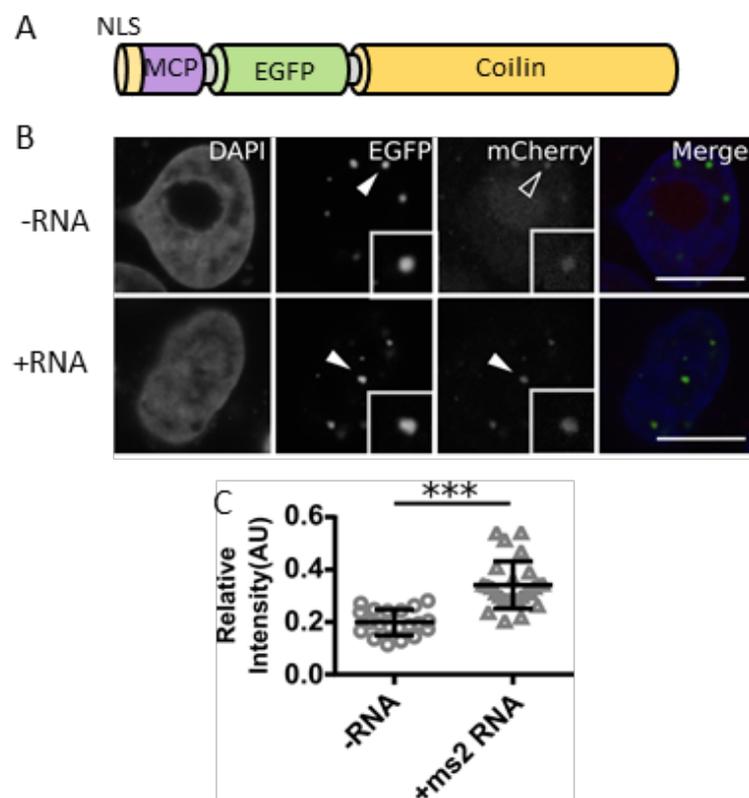
The nuclear bodies (NBs) are the granular structures of proteins and RNAs in the nucleus that have multiple functions. The Cajal nuclear body (CB) is one of them that takes part in cell processes such as transcriptional regulation, genome stability, and apoptosis. As Coilin protein is one of the main components of the CB, we developed a Coilin based RNA trap that can enrich the test RNA to the CB for binding measurement (Figure. 42).



**Figure. 42 RNA-protein interaction at the nuclear body**

The principle of RNA-protein interaction assay at the nuclear body. The RNA trap, which contains the nuclear body localization molecule, such as Coilin, the marker of the Cajal body, has the ability to accumulate the test RNA and the test protein to the nuclear body like Cajal body for the interaction measurement.

Similar to the construction of the Lamin B1 RNA trap, we replaced the LacI of the MCP RNA trap with the Coilin to get the Coilin RNA trap. We also tested its RNA trapping ability with multiple ms2 RNA and MCP-mCherry in HeLa cells. The green spots in the nucleus indicated that the RNA trap can aggregate together at the NBs with the help of the integrated Coilin, and when the RNA exists, the red spots with stronger fluorescence appeared and co-localized with the green spots (Figure. 43) demonstrated the ms2 RNA, as well as the MCP proteins, are also anchored to the CBs.

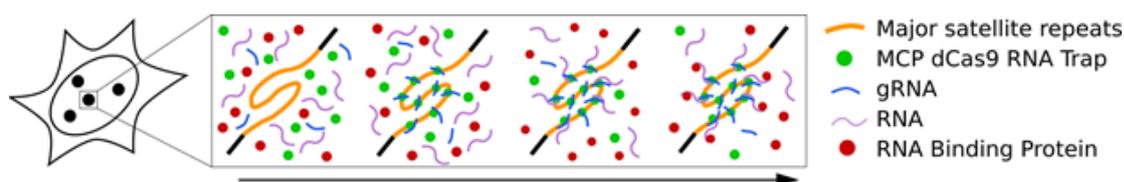


**Figure. 43 Detection of ms2-MCP interaction at the Cajal nuclear bodies**

(A) The construction of the Coilin RNA trap. The LacI in the MCP RNA trap is replaced with the Coilin. (B) Images of the ms2-MCP interaction assay with the Coilin RNA trap at the Cajal nuclear bodies (Bar: 10  $\mu$ m). (C) Quantification demonstrates that the ms2 RNA can largely enhance the accumulation of the MCP protein at the nuclear bodies. (\*\*\*)  $p < 0.001$ ).

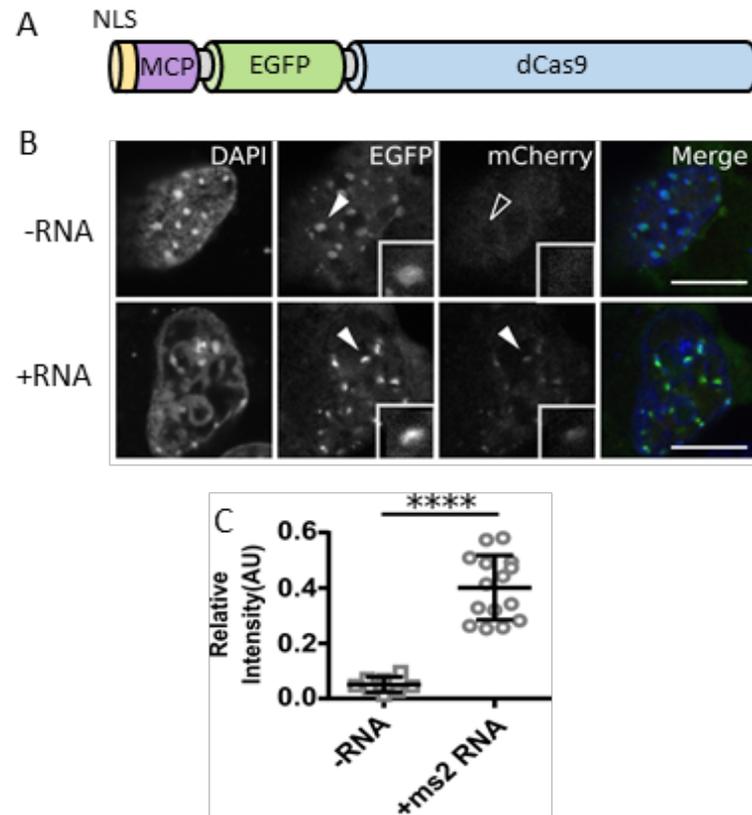
**3.5.3 Chromocenter of major satellites**

With the development of CRISPR-Cas genome editing technology, the deactivated Cas9 (dCas9) has been widely used in genome targeting and visualization. With the advantage of this technology, we designed an RNA trap that is able to target a particular genomic structure by replacing the LacI protein with the dCas9 protein. With the help of the guide RNA that targets the major satellite repeats, the dCas9 RNA trap could be anchored to the centromere areas in the nucleus (Figure. 44).

**Figure. 44 RNA-protein interaction at the chromocenter of major satellites**

The principle of RNA-protein interaction assay at the major satellites. The RNA trap, which contains the deactivated Cas9 protein, has the ability to anchor the test RNA and the test protein to the chromocenter of major satellite repeats with the guide RNA for their interaction assay.

We tested this RNA trap in the J1 embryonic stem cells together with the ms2-MCP pair, as well as the guide RNA, which targets the major satellite repeats that distribute around the centromeres and usually form the chromocenters. The targeting of the dCas9 RNA traps to the chromocenter was visualized as the co-localization of the green spots and the blue DAPI spots. And with the ms2 RNAs, the mCherry tagged MCP proteins could also be recruited to the chromocenters (Figure. 45).



**Figure. 45** Detection of the ms2-MCP interaction at the major satellites

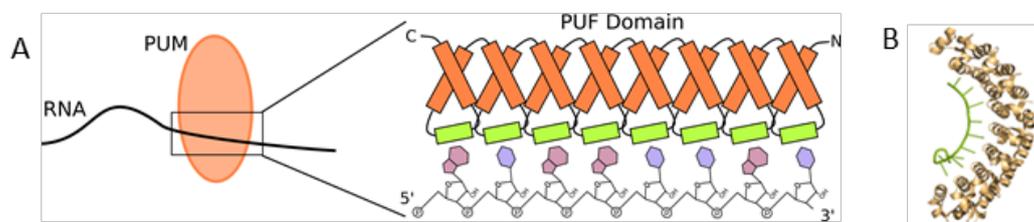
(A) The construction of the dCas9 RNA trap. LacI in the MCP RNA trap is replaced with the deactivated Cas9 protein. (B) Images of the ms2-MCP interaction with the dCas9 RNA trap and major satellites guide RNA (Bar: 10  $\mu$ m). (C) Quantification reveals that the aggregation of the MCP protein at the chromocenter of major satellites can be significantly enhanced by the ms2 RNA (\*\*\*\*  $p < 0.0001$ ).

### 3.6 Development of Alternative RNA Traps

Another limitation of this RNA trap is that it can only be applied to the ms2 tagged RNA molecules, which have to be generated by the construction and co-expression of additional plasmids. If the target RNA in the cell can be recognized and recruited directly to the anchor point by the RNA trap, the method can be more convenient. In order to achieve this goal, we focused on those programmable RNA binding proteins, which can recognize different target RNA molecules with some easy modifications.

### 3.6.1 PUM based RNA trap

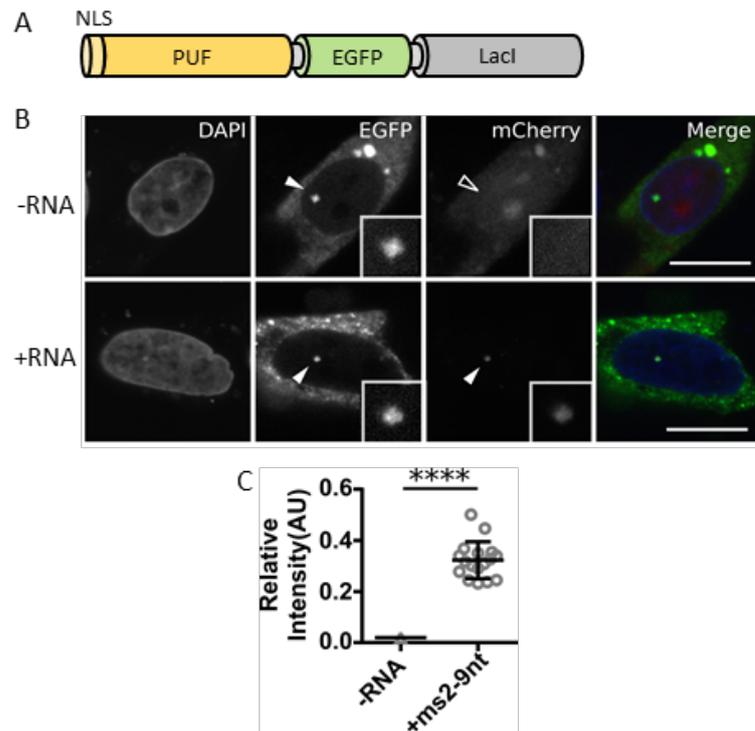
Pumilio and FBF homology (PUF) domain is the RNA binding motif of the PUM proteins, which typically consists of eight tri-helix subunits, and each subunit can recognize and binds to a single nucleotide (Figure. 46).



**Figure. 46** The RNA binding of the PUF domain

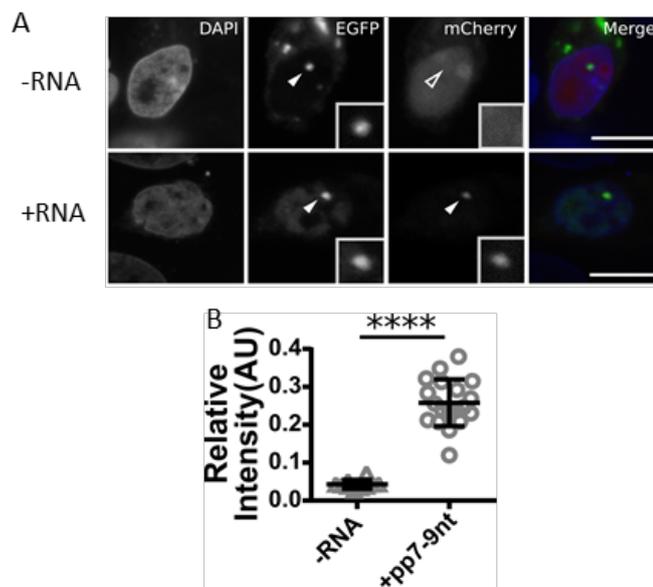
(A) Schematic representation of RNA binding of PUF domain. PUF domain usually contains eight tri-helix repeats, and each repeat takes part in single nucleotide recognition. (B) PUF domain forms a curved anti-parallel structure together with the target RNA, in which the direction of the RNA is 5' to 3' while the protein is C terminus to N terminus.

The modular structure of the PUF domain makes it a suitable candidate for RNA binding engineering. Here we chose an artificial PUF domain, which contains nine subunits, that can recognize a nine nucleotide sequence UGUUGUAUA, which is fused at the 5' end of *ms2*, *pp7*, and *H300* RNAs. Based on this, we developed a new PUF RNA trap by replacing the MCP with the artificial PUF and cloned it into the expression vector. Together with the three test RNAs and their mCherry labeled binding proteins MCP, PCP, and EZH2N, we tested the RNA trapping performance of the new PUF RNA trap in the BHK cells. All the results showed that the red fusions of all three test proteins can be recruited to the *lacO* array and have the co-localizations with the green spots of the PUF RNA trap protein (Figure. 47, 48, 49). Furthermore, we tested if the PUF RNA trap can be used for binding site measurement with the mutant test proteins MCP (S47R), PCP (V89Y), and EZH2N (T345D), the quantification results demonstrated that all the binding ability changes caused by mutations can be visualized in the test cells with the PUF RNA trap (Figure. 50, 51, 52).



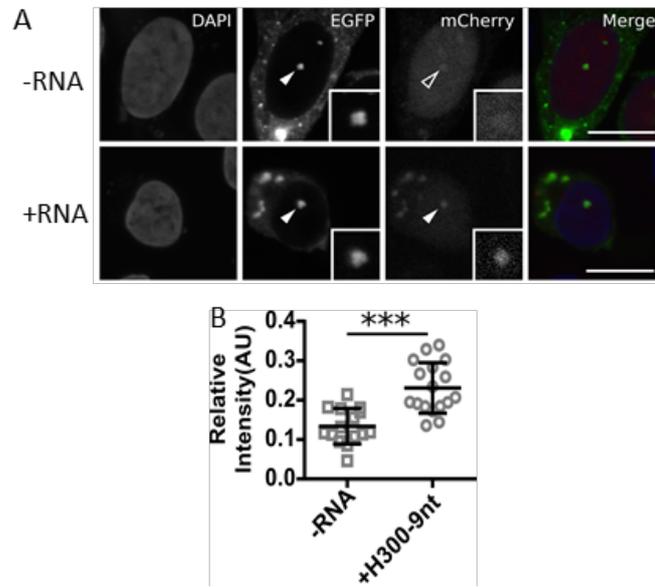
**Figure. 47 Detection of ms2-MCP interaction with the PUF RNA trap**

(A) The construction of the PUF RNA trap. The MCP of the MCP RNA trap is taken over by the PUF domain. (B) Images of the ms2-MCP interaction assay with the PUF RNA trap at the *lacO* array (Bar: 10  $\mu$ m). (C) Quantification reveals that the ms2-9nt RNA can significantly enhance the aggregation of the MCP protein at the *lacO* array (\*\*\*\*  $p < 0.0001$ ).



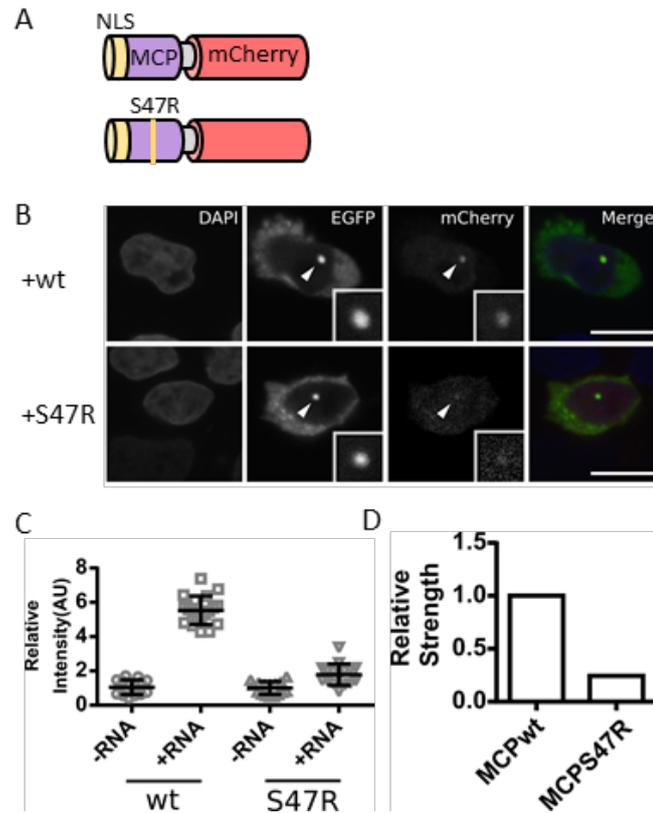
**Figure. 48 Detection of pp7-PCP interaction with the PUF RNA trap**

(A) Images of the pp7-PCP interaction assay with the PUF RNA trap at the *lacO* array (Bar: 10  $\mu\text{m}$ ). (B) Quantification demonstrates the significant enrichment of the PCP protein at the *lacO* array by the pp7-9nt RNA (\*\*\*\*  $p < 0.0001$ ).



**Figure. 49 Detection of H300-EZH2N interaction with the PUF RNA trap**

(A) Images of the H300-EZH2 interaction assay with the PUF RNA trap at the *lacO* array (Bar: 10  $\mu\text{m}$ ). (B) Quantification demonstrates the significant accumulation of the EZH2N protein at the *lacO* array by the H300-9nt RNA (\*\*\*)  $p < 0.001$ ).



**Figure. 50 Measurement of the ms2 binding site in MCP with the PUF RNA trap**

(A) The construction of the mutant MCP protein. The serine at the 47<sup>th</sup> position is replaced with an arginine. (B) Images of the wild-type/S47R MCP protein and ms2-9nt RNA interaction assays with the PUF RNA trap at the *lacO* array (Bar: 10  $\mu$ m). (C, D) Quantification suggests that the S47R mutant MCP has a significantly lower binding ability compared with the wildtype, and it is just 25% of the wild-type protein.

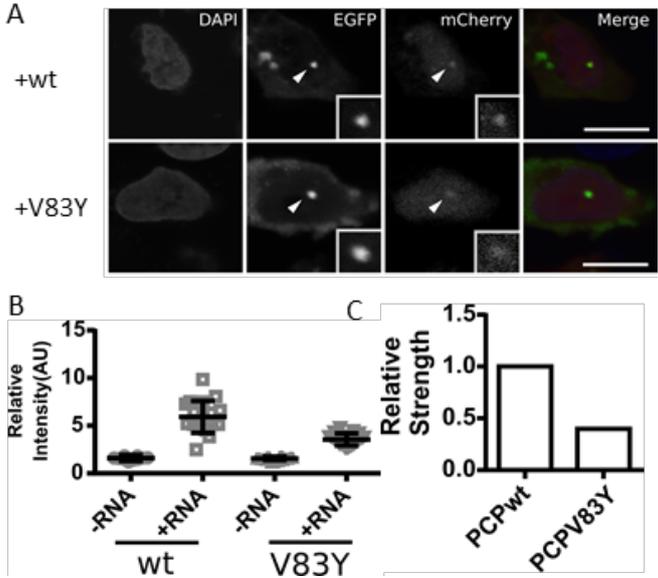


Figure. 51 Measurement of the pp7 binding site in PCP with the PUF RNA trap

(A) Images of the wild-type/V83Y PCP and pp7-9nt RNA interaction assays with the PUF RNA trap at the *lacO* array (Bar: 10  $\mu$ m). (B, C) Quantification indicates a significant fluorescence decrease between the wild-type and the mutant MCP proteins, and the value of the mutant was only 40% of the wild-type.

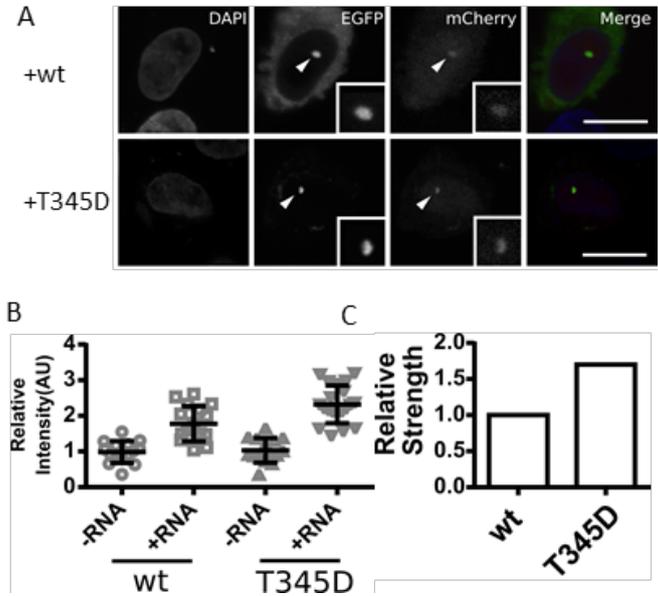


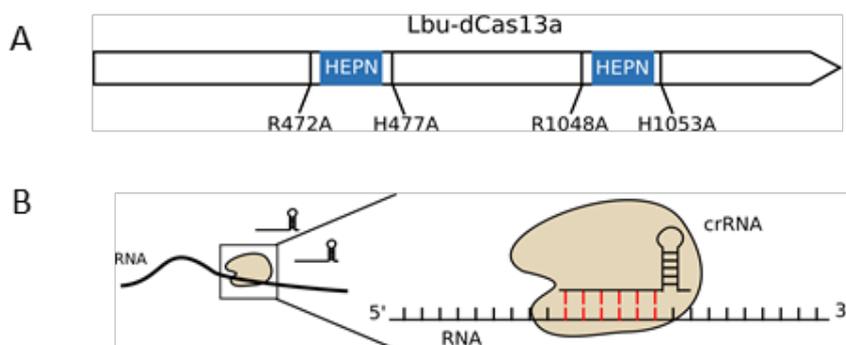
Figure. 52 Measurement of the HOTAIR binding site in EZH2 with the PUF RNA trap

(A) Images of the wild-type/T345D EZH2N protein and H300 RNA interaction assays with the PUF RNA trap at the *lacO* array (Bar: 10  $\mu$ m). (B, C) Quantification reveals a significant fluorescence enhancement

between the wild-type and the T345D mutant EZH2N proteins, and the value of the mutant was 65% higher than the wild-type.

### 3.6.2 dCas13a based RNA trap

Cas13a belongs to the class II type IV CRISPR/Cas system, which takes part in the recognition and digestion of the infected RNA molecules. The deactivated Cas13a (dCas13a) has been developed and used in RNA visualization and manipulation. As the target RNA molecule of the dCas13a is defined by the guide RNA, it can be used for the construction of a programmable RNA trap that can recognize different target RNAs just by conveniently using particular guide RNA molecules (Figure. 53).

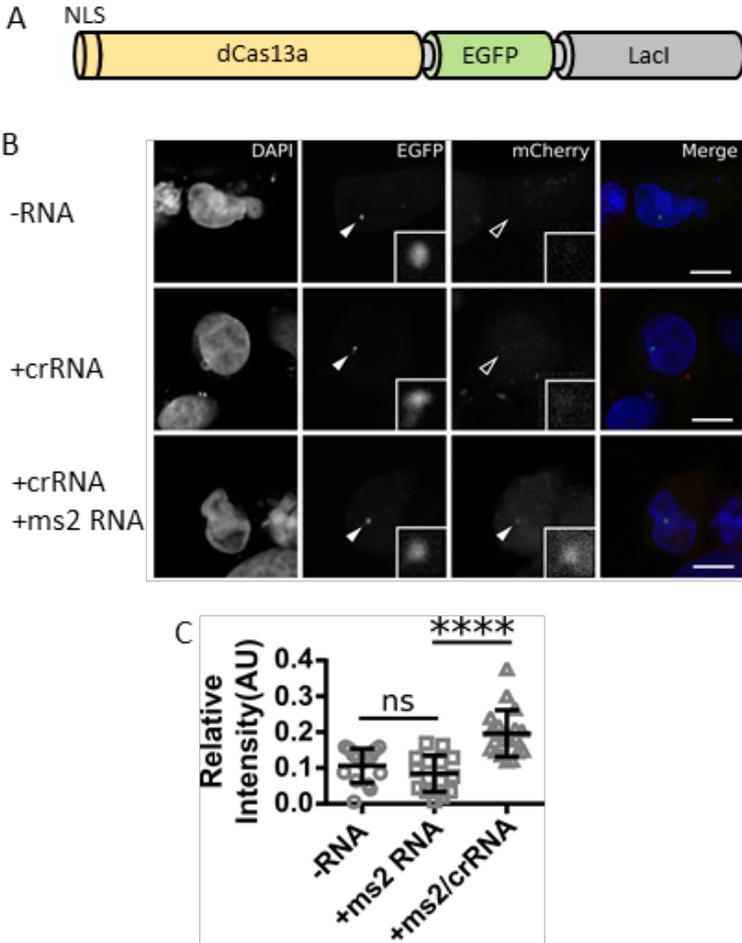


**Figure. 53 The structure of Lbu-dCas13a**

(A) The construction of deactivated Cas13a protein. The nuclease activity of the Cas13a protein is removed by four mutations (R472A, H477A, R1048A, H1053A) in two HEPN domains. (B) The dCas13a can recognize the target RNA sequence with the crRNA.

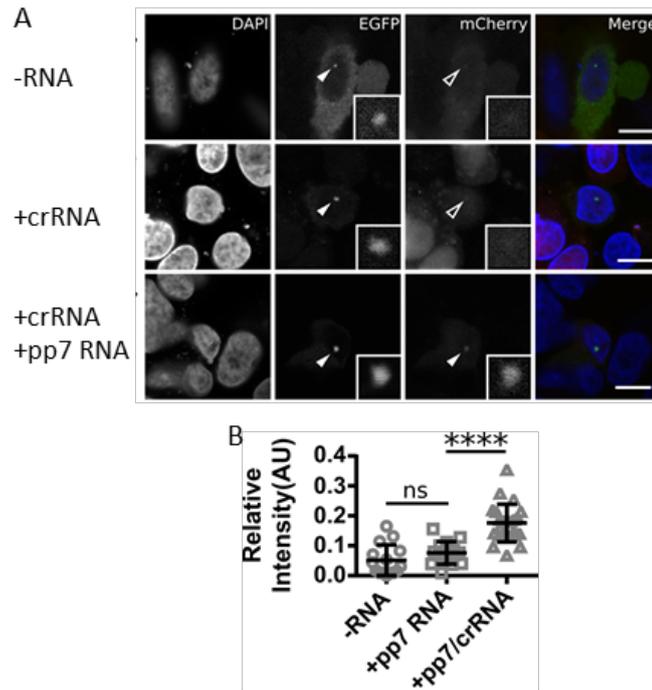
Based on this, we constructed another new RNA trap that the dCas13a molecule from *Leptotrichia buccalis* (Lbu) was used to take over the MCP in the RNA trap. And we also designed the guide RNA that contains an anti-parallel stem-loop followed by a spacer sequence (TAC CTT AGG ATC CAC TAG TTC TAG AAT C), which is reverse complement to a sequence (GAT TCT AGA ACT AGT GGA TCC TAA GGT A) that lies at the 5' end of both multiple m2 and multiple pp7 RNAs. Together with the mCherry labeled MCP and PCP protein, we tested the function of this dCas13a RNA trap in the BHK cells. The appearances of the green spots demonstrated that the dCas13a RNA trap is able to anchor at the *lacO* array first, followed by adding the ms2 and pp7 RNA, the aggregations of the red spots that co-localized

with the RNA traps showed that with the help of the guide RNA, dCas13a molecules are able to catch the target RNA and to recruit the MCP or PCP proteins, which also proves that the dCas13a RNA trap can be used for RNA-protein interaction measurement (Figure. 54, 55).



**Figure. 54** Detection of the ms2-MCP interaction with the dCas13a RNA trap

(A) The construction of the dCas13a RNA trap. The MCP in the MCP RNA trap is replaced with the Lbu-dCas13a. (B) Images of the ms2-MCP interaction assay at the multiple *lacO* loci with the dCas13a RNA trap (Bar: 10  $\mu$ m). (C) Quantification shows that the recruitment of the MCP at the *lacO* can be significantly enhanced by the coexistence of the crRNA and ms2 RNA (\*\*\*\* p < 0.0001).



**Figure. 55 Detection of the ms2-MCP interaction with the dCas13a RNA trap**

(A) Images of the pp7-PCP interaction assay at the multiple *lacO* loci with the dCas13a RNA trap (Bar: 10  $\mu\text{m}$ ). (B) Quantification demonstrates that the recruitment of the MCP at the *lacO* can be significantly enhanced by the coexistence of the crRNA and ms2 RNA (\*\*\*\*  $p < 0.0001$ ).

### 3.7 Measurement of RNA-protein binding kinetics

So far, the RNA-protein interactions we characterized were all in fixed samples; one of the advantages of our RNA assay is that it can also be applied in live cells. We performed Fluorescence Recovery After Photobleaching (FRAP) assay to measure the RNA binding kinetics of proteins. Fluorescence of the bound protein was photo-bleached with laser, and the fluorescence recovery kinetic was recorded over time. We compared the recovery abilities of wild-type MCP and the mutant S47R MCP that bind to the trapped multiple ms2 RNA by recording the changes of relative fluorescence after bleaching at the *lacO* array in the BHK cells. The half recovery times ( $t_{1/2}$ ) of the wild-type and mutant MCP are 3.8s and 1.6s, respectively, and the wild-type MCP showed far more immobile fraction than the S47R mutant, indicating a higher affinity of the wild-type than the mutant (Figure. 56).

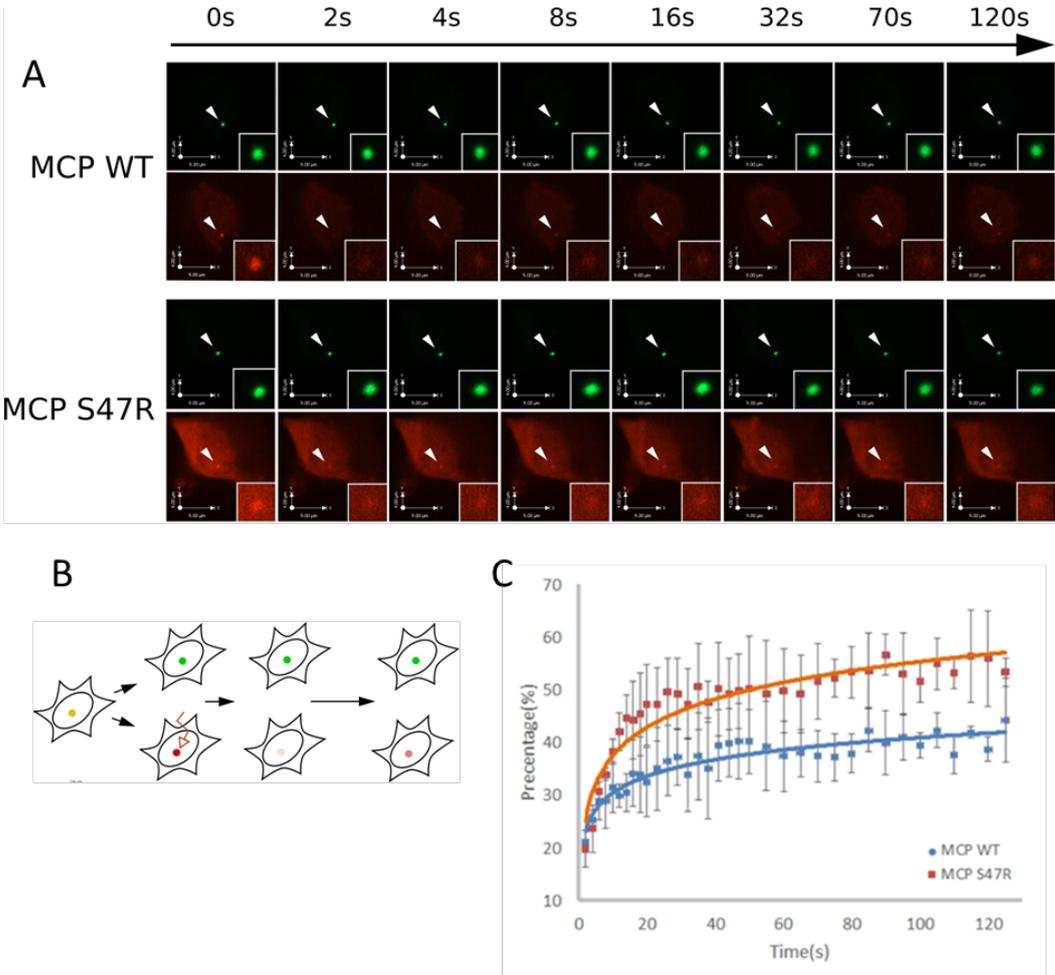


Figure. 56 Measurement of the ms2-MCP binding kinetics

(A) The principle of the FRAP. During the test, the red fluorescence at the *lacO* array will be bleached by a laser while the green fluorescence is kept, and the recovery of the red fluorescence will be recorded until it is almost stable. (B) Images of the fluorescence changes during the interaction tests of wild-type/S47R MCP protein and the ms2 RNA at the *lacO* array from 0s to 120s after bleaching. (C) The relative fluorescence curves of the wild-type and the mutant MCP proteins indicate that the mutant has a shorter half recovery time and higher recovery ratio than the wild-type.

## 4 Discussion



#### 4.1 Fluorescence Tri-hybrid Method for RNA-protein Interaction Assay

In the past decades, a series of techniques have been established to study the interactions between RNAs and proteins. These methods are developed on different principles and have distinct advantages and disadvantages. As a fluorescence-based technology for RNA-protein interaction (RPI) study, the RNA trap based fluorescence three-hybrid (F3H) method utilizes a specially designed RNA trap to anchor the test RNA and its binding protein to a specific site within the cell, such as an artificial chromatin locus in the nucleus, and then measure the interaction between them. This method integrates fluorescence and RNA labeling technology, as well as the eukaryotic expression system, to study RPI in a convenient way.

Our F3H approach is a combination of some existing methods. On the one side, our method has relationships with the fluorescence two-hybrid and tri-hybrid methods, which have been widely used in protein-protein interaction (PPI) studies [Zolghadr, Mortusewicz et al. 2008, Herce, Deng et al. 2013]. Similar to these two methods, we also employ the measurement process in a genetically modified cell line, with an integrated plasmid containing multiple repeats of bacterial *lac* operator (*lacO*) DNA sequence. The *lac* repressor (LacI) is able to bind this artificial chromatin so that it can be used for chromatin tracing. In the beginning, the artificial *lacO* array was developed to study the organization of the chromatin and visualize the activity of gene expression [Robinett, Straight et al. 1996, Tsukamoto, Hashiguchi et al. 2000], and recently, the formation of kinetochores was also studied with this system [Gascoigne, Takeuchi et al. 2011]. Based on this, a fluorescence two-hybrid system was established by our group, which contains two major parts: a triple fusion RFP-LacI-bait, and a two-components GFP-prey protein. The prey protein will be recruited and co-localized with the bait protein at the *lacO* array, and the interaction can be visualized with fluorescent proteins [Zolghadr, Mortusewicz et al. 2008]. And to expand its application, the two-hybrid system has been upgraded to a tri-hybrid version: the RFP-LacI-bait protein was taken over by a protein trap, which is made of GBP (GFP binding protein) and LacI and can recruit the GFP labeled protein to the *lacO* array directly, to allow the method measure the interaction between any GFP and RFP fused proteins [Herce, Deng et al. 2013].

On the other side, our method is also similar to the yeast tri-hybrid system, which has been used for RPI study for a long time. In both ways, RNAs of interest are all needed to be labeled with specific tags such as ms2 stem-loops. For yeast tri-hybrid method, the ms2 tagged RNA

can be recognized by the MCP fused DNA binding domain (DB), meanwhile, the RNA will also recruit the RNA binding protein fused activation domain (AD), to activate the expression of the reporter gene [Jaeger, Eriani et al. 2004]. For our method, the ms2 tagged RNA will be recognized by the MCP RNA trap and anchored at the *lacO* array in the nucleus to ensure that the interaction between RNA and protein can be estimated by measuring the strength of the two co-localized fluorescences.

As a new method for the RPI study, our RNA trap based F3H assay has several particular features. First, our approach is performed in the mammalian cell system, which allows us to study the interaction *in vivo*. It is able to maintain the natural structure and the modification status of each test component, which is much simpler and more precise than the traditional *in vitro* biochemical methods and is more suitable for the proteins from higher eukaryotes. Second, using fluorescence as the detection signal instead of the yield of the reporter gene in yeast hybrid system can simplify the measurement process and minimize the bias. Also, our method is able to indicate the binding affinity between RNA and protein by measuring the signal ratio at the *lacO* spot according to the fluorescence intensities, so that it can provide not only qualitative analysis but also semi-qualitative binding data for RNA and proteins. Moreover, our method can visualize RPI in real-time in living cells, and this feature provides the ability to study the binding kinetics between the two components. However, measurement precision still needs improvement. As the RPI F3H system contains three components, their compositions in the cell may influence the efficiency of the measurement. At first, we adjusted the amount of RNA trap. The result indicated that the negative signal decreased as the reduction of the RNA trap, and the positive signal also decreased to keep the positive/negative ratio stable. Then we adjusted the amount of test protein. The negative signal decreased as the dropping of the test protein, but the positive signal was enhanced, and both consequences led to a considerable enhancement of the positive/negative ratio. The result revealed that fewer test proteins in the system could cause a higher positive/negative ratio. And this role was also applied in our future experiments.

## 4.2 RNA Expression Cassettes

In biological research, it is necessary to express the external genes in cells so that we can study their functions better. So different RNA expression cassettes have been constructed to synthesize the target RNA molecules when they are needed.

The promoter of cytomegalovirus, or the CMV promoter, is one of the most widely used promoters for external gene expression in mammalian cells. Compared with other mammalian promoters such as SV40 and RSV, the CMV promoter has a stronger ability to start transcription [Cheng, Ziegelhoffer et al. 1993]. The upper enhancer sequence of the CMV promoter, which can recruit transcriptional factors, can significantly increase the expression level of CMV promoter [Schmidt, Christoph et al. 1990]. RNA pol II is recruited by the CMV promoter and responsible for RNA synthesis. In order to terminate the transcription process when the total length RNA has been synthesized, an SV40 poly(A) signal sequence, which provides the AAUAAA motif that can be recognized by the polyadenylase, is placed downstream of the RNA sequence [Fitzgerald and Shenk 1981]. Moreover, in order to further increase the expression level, the CAG promoter, which combines the early enhancer of the CMV promoter and the promoter sequence of chicken  $\beta$ -actin, has been developed and widely used [Alexopoulou, Couchman et al. 2008].

As the measurement process of RNA-protein interaction is performed in the nucleus, the ideal condition is that the test components can be kept in the nucleus, especially the test RNA. However, the polyadenylation process that is induced by the RNA pol II, is a strong exporting signal, which will reduce the amount of the test RNA in the nucleus. In order to overcome this, we tried to use RNA pol III, which does not rely on the poly(A) signal for termination and will not recruit the polyadenylase during the transcription, to take over RNA pol II [Borchert, Lanier et al. 2006]. The RNA pol III is typically used for the expression of siRNA and guide RNA and under the control of a U6 cassette, which contains an upstream U6 promoter and downstream oligo(T) sequence that allows the polymerase to release from the template [Miyagishi and Taira 2002]. Since the natural RNA pol III products are generally not over 300 nucleotides, it is normally used for short RNA expression, such as siRNA and guide RNA in biological researches [Miyagishi and Taira 2002, White 2004].

In our study, we mainly used the CMV cassette to generate the test RNAs for the F3H test. The post-transcriptional RNA process and transport followed the transcription by RNA pol II

may influence the amount and quality of test RNA in the nucleus, and then affect the result of the measurement. Choosing an expression system that will not cause the post-transcriptional processes may contribute to the maintenance of test RNA in the nucleus and to improve the measurement. Hence, we compared CMV and U6 cassettes in the pp7-PCP test and found that the U6 cassette products have a higher concentration in the nucleus and this may be the reason that leads to a better result than the CMV cassette. However, the weakness of the U6 cassette, which is not suitable for the generations of those long transcripts, would limit its applications in some studies.

### 4.3 Alternative Anchor Sites

One limitation of the RNA trap based F3H method is that it has to be performed only in the *lacO* array integrated cell line. In order to extend its application, it is necessary and essential to look for other suitable anchor sites, which not only can be used to aggregate enough amount of the test RNAs and proteins to make a significant fluorescence enrichment but also distribute widely in general cell lines.

In the eukaryotic cells, the nuclear and cytoplasmic components are separated by the nuclear envelope (NE), which consists of the outer and inner nuclear membranes (ONM and INM). The nuclear lamina locates at the inner side of the INM, which is a network of lamin protein polymers and lamin-binding proteins [Stuurman, Heins et al. 1998]. Compared with the A-type lamins, the B type-lamins exist in all the cell types and have a stronger and more stable INM binding ability, which largely depends on the isoprenylation modification [Gruenbaum, Margalit et al. 2005]. These properties make Lamin B a useful anchor site for INM localization, for example, Lamin B1 has been used as a widespread target molecule, which exists in any cell type, to localize the cargos to the INM [Herce, Deng et al. 2013].

The eukaryotic cell contains different subcellular structures and the eukaryotic nucleus also has subnuclear organelles [Handwerger and Gall 2006]. Nuclear body is one kind of subnuclear organelles and considered to play important roles in many cell processes like RNA maturation. The spliceosomal small ribonucleoprotein particles (snRNPs) are essential pre-mRNA splicing factors, and the final steps of their maturation take place in the Cajal bodies (CBs), which are non-membrane inclusions present in the nucleus of most cells [Ogg and

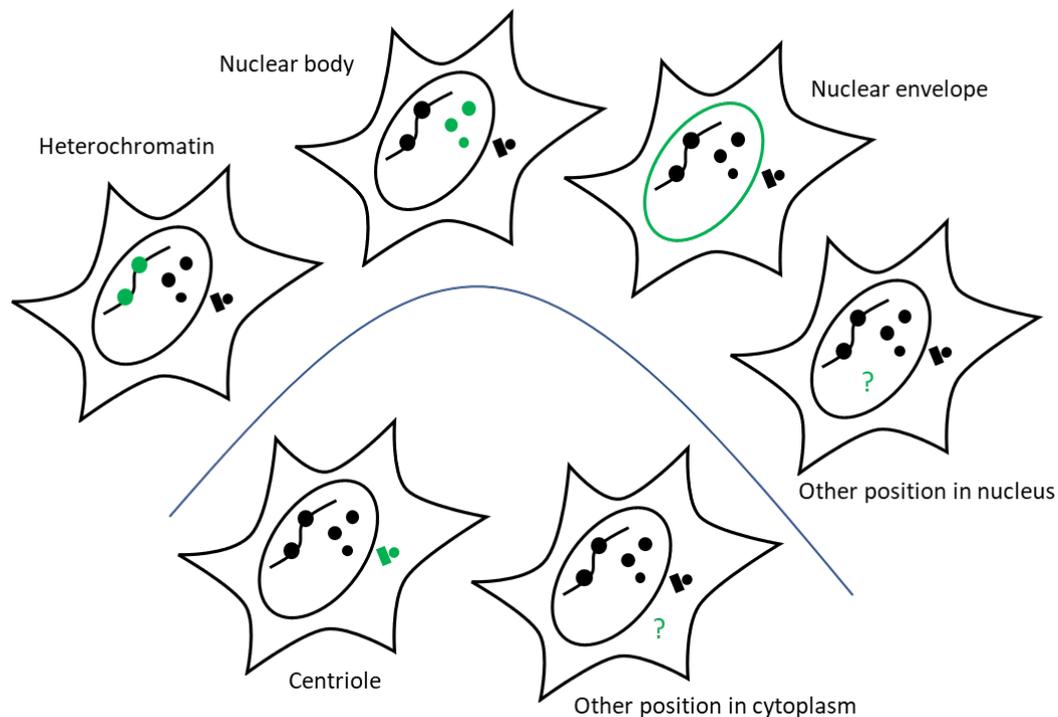
Lamond 2002]. Coilin is the marker protein of CBs, its N-terminal amino acid sequence mediates self-interaction and is required for CN targeting. Although the size and number of CBs vary among cell types, typically there are 2-3 CBs existing in the nucleus of HeLa cells [Morris 2008], which is suitable for RNA-protein interaction measurement.

The *lacO* array, which we used as the anchor point for the RNA trap, was constructed by the stable integration of 256 copies of the *lac* operator sequence into the genome of BHK cell [Tsukamoto, Hashiguchi et al. 2000], and can be recognized by the Lac inhibitor (LacI) fused molecules. In the nucleus, the multiple *lacO* array reveals a condensed appearance, which is similar to the structure of heterochromatin, to make it suitable for visualization. However, there are several natural tandem repeat sequences similar to the *lacO* array existing in the genome, such as telomeric repeats, major and minor satellites [Padeken, Zeller et al. 2015]. The major satellite repeats are usually 6 megabases of 234 bp units, which distribute around the centromeres and exist as heterochromatin in the interphase nuclei [Guenatri, Bailly et al. 2004]. Previous researchers have visualized the major satellite repeats by using the fluorescent protein-tagged dCas9, which indicated the application of dCas9 in chromatin localization [Anton, Bultmann et al. 2014].

In our study, we tried all three nuclear anchor sites mentioned above. By using modified RNA traps, we localized the test RNAs to these sites and observed the aggregation of test proteins, indicating that all the three anchor points can be used for interaction assay. Both nuclear envelope and major satellites indicated lower backgrounds and showed higher ratios of positive/negative signals that mean more precise results. But the network structure of the lamina layer at the INM is slimmer than other anchor sites, which limits the binding of RNA trap and test molecules. We also found that the nuclear entry of the dCas9 RNA trap, which is used for major satellite repeats targeting, is not as good as other RNA traps, and may influence the measurement. This may be caused by its large molecule weight and could be improved by rearranging or adding more NLSs [Luo, Pang et al. 2004]. The CBs have complex compositions of RNAs and proteins, which may lead to strong unspecific binding with test proteins [Gall 2003], to reveal a similar background and positive/negative ratio as the *lacO* array.

Although all the anchor sites we used are in the nucleus, some structures in the cytoplasm may also be used as anchor sites, one of them is the centriole. The centrioles are the densely

staining structure of the centrosome, which are strikingly symmetrical barrel-shaped structures of nine sets of triplet microtubules [Doxsey 2001]. The centrioles are usually surrounded by the pericentriolar material, which is an interconnected meshwork of fibers and proteins [Doxsey 2001]. Some studies have shown the centrin, which is one component of the pericentriolar material, can be used as a localization tag that targets the fusion protein to the spot of the centrosome [Baron, Greenwood et al. 1992, Herce, Deng et al. 2013].



**Figure. 57 Different anchor sites in the cell**

Many cellular organelles and structures can be used as anchors. Heterochromatin, nuclear body and nuclear envelope in the nucleus and centriole in the cytoplasm showed in green color are all potential anchor sites for RPI F3H assay.

#### 4.4 Genome Targeting with CRISPR/Cas9 System

The organization of the genome and its interactions with histones and other epigenetic factors play fundamental roles in gene expression and even cell function regulation. On the one side, to understand how chromatin structure affects cell activities, Chromosome conformation capture (3C) and other methods have been successfully applied to show the folding principles and the organizations of chromatin [Woodcock and Horowitz 1995]. On the

other side, Fluorescence in situ hybridization (FISH) based technology allows us to exam how the particular chromatin sequence distributes and functions in the life process [Williamson, Berlivet et al. 2014].

CRISPR/Cas9 is an RNA-guided DNA endonuclease that functions as the virus defense system in bacterial cells. The most important application of the CRISPR/Cas9 system is a tool for chromatin targeting. Before the development of the CRISPR/Cas9 system, zinc-finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN) had been used for genome engineering [Gaj, Gersbach et al. 2013]. The programmable sequence-binding modules provide ZFN and TALEN the abilities to target a particular position of the genome by the recognition of a specific target sequence. However, the disadvantage, which is that the sequence binding modules have to be re-designed for different target sequences, restricts their application [Gaj, Gersbach et al. 2013]. Instead, the Cas9 based tools, which can recognize multiple targets by conveniently using different guide RNAs, have been quickly and widely applied in biological studies [Hsu, Lander et al. 2014].

Genome editing is the most critical application of the CRISPR/Cas9 system. Cas9 protein is able to localize at the specific site of the genome with the help of the guide RNA and cleave at the protospacer adjacent motif (PAM) site with its endonuclease activity. The double-strand break caused by Cas9 protein will lead to the repair process, and then causes random mutation or homologous recombination [Hsu, Lander et al. 2014]. The Cas9 induced mutation has been widely used as a gene knockdown strategy to permanently inhibit target gene expression. While the Cas9 induced homologous recombination can be used to insert the sequence of the donor into the target site, moreover, the MIN (Multifunctional Integrase) tag technology, which uses the Cas9 to add a Bxb recombination site containing sequence into the genome, can be used to integrate large DNA fragment into the target genome [Mulholland, Smets et al. 2015]. Genome visualization is another important application of Cas9 protein. Fluorescent protein-tagged deactivated Cas9 has been used to visualize the heterochromatin sequences such as telomeres, minor, and major satellites in mouse embryo cells [Anton, Bultmann et al. 2014]. Although the dCas9 based imaging is mainly restricted to the repeat sequence, co-expression of a guide RNA library that targets the different segments of the same sequence makes it possible to trace the single-copy locus [Chen, Gilbert et al. 2014]. In order to observe two or more genome locus at the same time, different guide RNAs

are fused with varying tags of RNA such as ms2 and pp7. By co-expressing different fluorescent proteins labeled MCP, PCP, and other RNA tag binding proteins, dCas9 can indicate different genome locus simultaneously in the nucleus [Wang, Su et al. 2016]. Moreover, dCas9 protein can also be used to bring transcription factors to regulate the expression of specific genes [Konermann, Brigham et al. 2015]. In our study, dCas9 was applied to localize the RNA trap at the major satellite sequences, which distribute around the centromeres as heterochromatin and is a suitable place for visualization and quantification of the interaction between test RNA and protein.

## 4.5 RNA Tracking Tools

RNAs play essential and diverse roles in biology, but the shortage of suitable tools limits our study on them. But recently, the discoveries and developments of PUF and Cas13a protein make significant improvements that dramatically promote researches in the RNA area.

### 4.5.1 RNA tracking with PUF protein

In eukaryotic cells, the activities of the mRNA molecules are highly controlled, usually through regulatory elements, which are located in their 3' UTRs. Proteins that bind these elements to play essential roles in controlling mRNA stability, translation, and localization [Wilkie, Dickson et al. 2003]. One family of the 3' UTR regulatory proteins is called PUF proteins. The PUF proteins exist in a wide variety of eukaryotic species, and they not only have related structures but also have similar functions: enhancing or repressing translation combinatorically with other regulatory proteins [Wickens, Bernstein et al. 2002].

The RNA binding function of PUF protein is performed by the PUF domain. It is typically characterized by the presence of eight consecutive repeats, each repeat is approximately 40 amino acids in length and folds into a three-helical unit. The second helix of each repeat mainly takes part in the base recognition. Sequence analysis indicates that the second helices of different units have highly conserved amino acid sequences. The various amino acids in the conserved sequence are critical in base recognition: cysteine and glutamine bind adenine, asparagine and glutamine bind uracil, and serine and glutamate bind guanine [Wang, McLachlan et al. 2002, Cheong and Hall 2006]. These codes have been confirmed by studies,

and the specificity of individual repeats can also be switched between different bases just by simple single amino acid mutation [Cheong and Hall 2006]. Based on this, engineered PUF domains, which are generated by arranging the order of different PUF repeats, have the ability to recognize and bind different RNA sequences of interest [Lu, Dolgner et al. 2009]. Furthermore, the number of the repeats also affects the RNA binding ability of the PUF domain: on the one hand, PUF domain with more binding repeats will have a more extended target sequence, which can increase the number of the molecular interactions between them and then enhance the binding, on the other hand, more repeats will enlarge the curved shape of the domain, and the change of the conformation may also interfere with the interactions between the bases and the amino acids [Zhang, McCann et al. 2016, Zhao, Mao et al. 2018]. By testing the binding abilities between PUF domains of different sizes and their targets, the nine repeats PUF domain keeps the best balance between the binding capacity and the specificity [Zhao, Mao et al. 2018]. According to this, in our study, a PUF domain that contains nine repeat areas, which can recognize a conserved nine nucleotide sequence in the test RNA molecules, was chosen to modify the RNA trap. In the later tests, this PUF RNA trap shows its function that can successfully recruit the ms2, pp7, and H300 RNA at the *lacO* site for their interaction measurements.

However, the use of the PUF protein as tools has still been hampered because the natural cytosine binding PUF repeat has not been found [Filipovska, Razif et al. 2011]. This sharply limits the potential targets for engineered PUFs because even for those RNAs which generated from the low guanine/cytosine composition genomes, the main octamer sequence could have one or even more cytosines, and if a small RNA or a restricted part of the RNA is the target, it is not easy to design the PUF domain for binding [Filipovska, Razif et al. 2011]. Many works have been applied to find out the cytosine-recognition code. As a result, several PUF repeats, which contain five-residue RNA interaction sequences such as SYXXR, GYXXR, and TYXXR from position 12 to 16 in the second helix, show the binding ability to the cytosine [Dong, Wang et al. 2011, Filipovska, Razif et al. 2011]. These discoveries will broadly enrich the applications of the PUF domain in RNA research.

#### 4.5.2 RNA tracking with CRISPR/Cas13a system

Most adaptive defense systems of prokaryotic cells target DNA substrates, but the type III and IV CRISPR systems are able to recognize ssRNAs. It has been demonstrated that Cas13a (known as C2c2 before) and Cas13b, which belong to the class 2 type IV CRISPR/Cas family, are single-component RNA-guided RNA targeting RNases that have both RNA processing and programmed RNA degradation abilities [Yang and Chen 2017].

In contrast to Cas9, Cas13a is an RNA-targeted RNase and does not need a PAMmer, which makes it easier to be applied in RNA engineering [Yang and Chen 2017]. At first, the Cas13a from *Leptotrichia wadei* (LwaCas13a) has been used for RNA knockdown, which was achieved through RNA cleavage in *E. coli*; then, the mammalian codon-optimized LwaCas13a, which was fused with GFP and a nuclear localization sequence (NLS), resulted in a considerable level of knockdown in mammalian cells. And the crRNA, which contains a 28 nucleotide spacer that ends with a 3' non-guanine nucleotide, could provide the Cas13a the highest silencing efficiency [Yang and Chen 2017]. Besides knockdown, Cas13a can also be used for nucleic acid detection. The RNase activity of the Cas13a can be activated by the dimer of the crRNA and the target RNA molecule, and the cleavage will be applied not only on the target RNA, but also on the non-target RNA molecules (bystander) [Terns 2018]. Based on this, a technology called Specific High-Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK) was generated and used for attomolar virus detection [Abudayyeh, Gootenberg et al. 2017].

The deactivated Cas13 platform, in which the nuclease function of the HEPN domain has been eliminated, has already been shown to be a useful RNA targeting tool, and it could be used directly as the inhibitors of activators of pre-mRNA splicing or translation, or to affect the functions of mRNA and non-coding RNA [Terns 2018]. Fusing different effector proteins to dCas13a to expand its role beyond RNA cleavage has already been considered as an effective strategy. Similar to the fluorescent protein labeled dCas9 that can be used for DNA sequence visualization with guide RNA [Anton, Bultmann et al. 2014], the RNA visualization method *in vivo* with fluorescent protein labeled dCas13a has also been established [Abudayyeh, Gootenberg et al. 2017], and the tracking target can be easily changed by using different crRNAs. Besides tracking, RNA editing is another essential application of dCas13. The deaminase domain of adenosine deaminases acting on RNA (ADAR<sub>DD</sub>) fused to dCas13b has demonstrated RNA editing for programmable adenosine to inosine (A to I) replacement

(REPAIR) [Cox, Gootenberg et al. 2017]. The design of alternative splicing, as well as the generation of specific protein isoforms, can all be programmably altered by coupling dCas13 protein with certain splicing regulators [Konermann, Lotfy et al. 2018]. Moreover, the epitope or affinity-tagged dCas13 molecules have the possibilities to impact on RNA biology by providing a specific way to isolate and characterize particular RNAs and their associated protein. In our study, we developed a dCas13a containing RNA trap, which can recognize and recruit specific RNA molecules at the *lacO* array with certain crRNA. We confirmed that this RNA trap can successfully demonstrate the RNA-protein interactions in both ms2-MCP and pp7-PCP tests. Compared with the PUF RNA trap, the crRNAs can provide longer recognition sequence, which can provide dCas13a RNA trap a higher accuracy of RNA capture. However, the dCas13a RNA trap also has nuclear entry problems, which is possibly related to its large molecular weight, and it may be improved by changing the position of the NLS or adding an additional NLS. Even though, the development of CRISPR-based RNA tracking tools will still contribute to the progression of RNA biology.

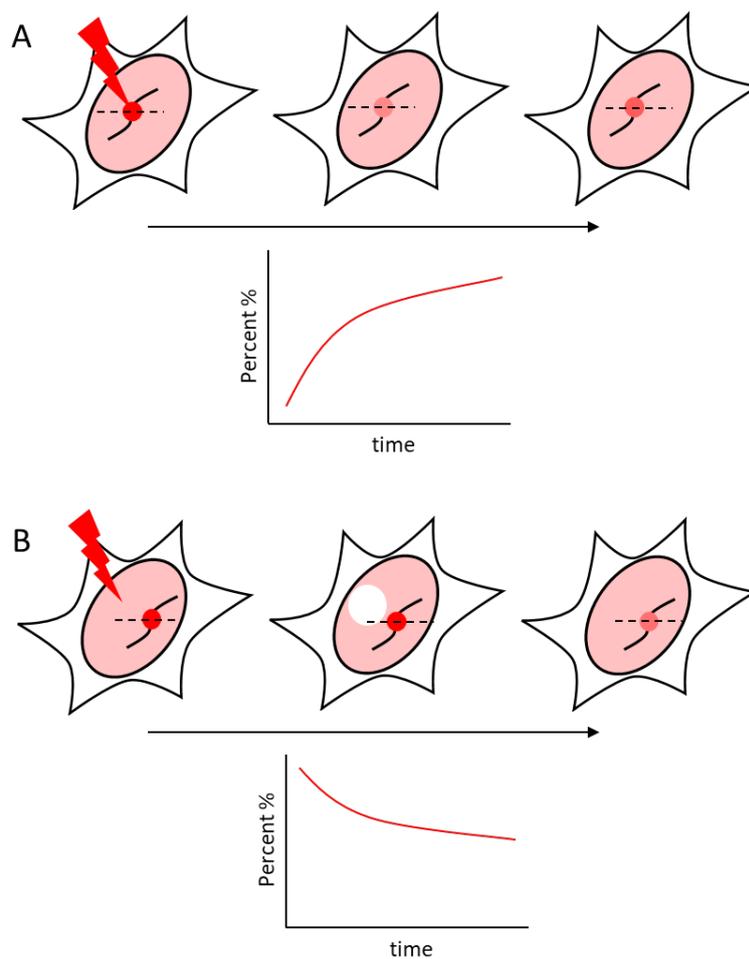
#### 4.6 Measurement of RNA-protein Binding Kinetics

The mobility of a fluorescent protein can be measured by using the photobleaching technology fluorescence recovery after photobleaching (FRAP). In general, the fluorescent molecules within a small area of the cell can be irreversibly photobleached by a high-power laser beam, and then the movement of the surrounding unbleached molecules can be detected by a low-power laser beam [Reits and Neefjes 2001]. The fluorescence intensities in the bleaching area need to be recorded for a period of time from a few seconds before bleaching to several minutes after bleaching until the brightness becomes stable. The mobile fraction  $M_f$  and the diffusion constant  $D$ , which can be obtained from the FRAP curve, are normally used to analyze the mobility of the test protein.

FRAP technology allows us to study the protein dynamics in real-time in living cells, and it can be used to explain a series of questions. Compared with the original *in vitro* RNA-protein interaction assay such as EMSA, our fluorescence three-hybrid method can largely retain the native properties of RNA molecules and reflect their interaction with proteins in the cell environment. However, the measurement process is still the fixation of test cells, which still

cannot reflect the real binding situations. As our RNA trap is able to aggregate the test RNA as well as the test protein, it provides the possibility to determine the interaction in a more precise FRAP way. The test protein will gather at the anchor site in the nucleus, and the binding and release will be balanced. After bleaching, the fluorescence at the anchor site will recover, and the recovery speed and final brightness largely depend on the interaction between test RNA and test protein.

To test the theory that the serine at 47<sup>th</sup> position of MCP protein is a potential ms2 RNA binding site, wild-type and S47R mutant MCP proteins are expressed in test cells separately with RNA trap and ms2 RNA, and the recovery curves of both proteins are recorded after bleaching. From the curves, the mutant protein demonstrated a higher exchange rate and faster exchange speed than the wild-type protein, which indicates the weaker binding ability of the mutant protein and the importance of the serine in RNA binding.



**Figure. 58 Study of RNA-protein binding kinetics with FRAP**

(A) The typical FRAP method is using a high-power laser to bleach the fluorescence of proteins at the anchor point and then recording the fluorescence recovery. (B) But sometimes the fluorescence of unbound proteins can also be bleached, and the fluorescence decrease at the anchor site is recorded and analyzed.

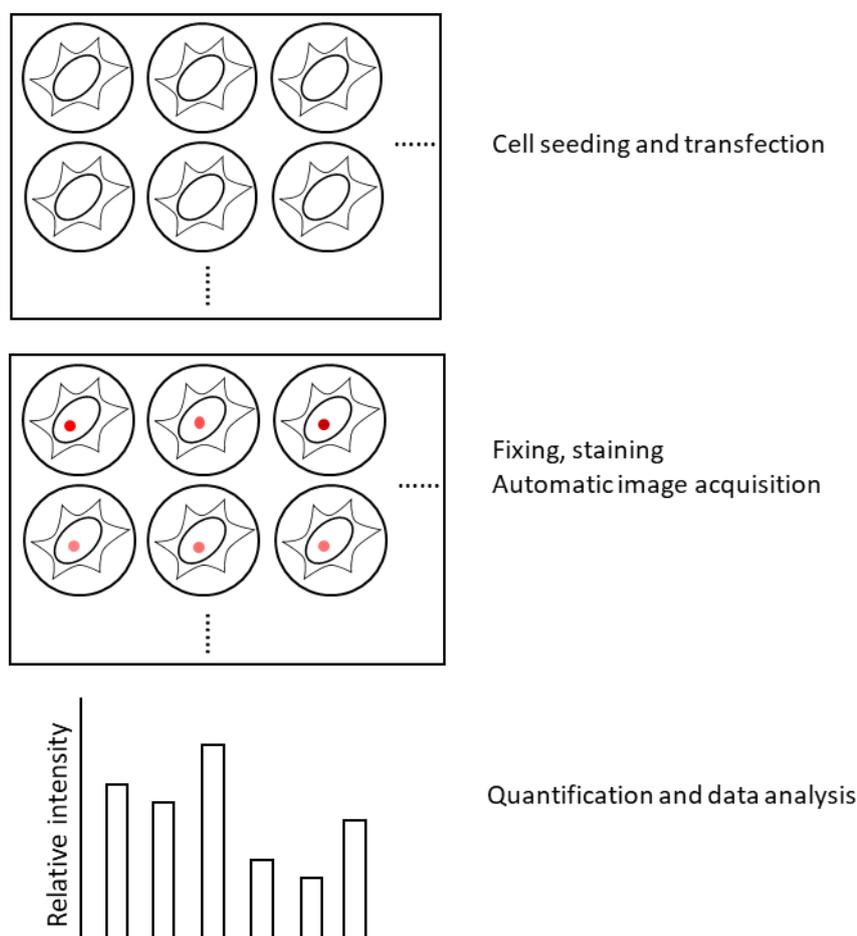
**4.7 High-throughput Screening in RNA-protein Interaction Assay**

Large-scale and high-throughput methods are becoming more and more essential for modern biological studies. For protein-protein interaction studies, high-throughput technologies have been already widely used. For example, the large-scale yeast two-hybrid screen technology has been established to screen the potential binder of the known proteins. A library of test proteins needs to be developed first, and then be used to bind known protein, the interaction can be detected by checking the products of the reporter gene on large plates [Uetz, Giot et al. 2000]. Large-scale screening of protein-protein interaction can also be performed with fluorescence tri-hybridization method, by measuring the relative fluorescence at the *lacO* array in the nucleus of the cells in 96-well plates [Herce, Deng et al. 2013].

The interaction between RNA and protein can also be measured in a high-throughput way. Similar to other high-throughput methods, library establishment is a necessary preparation. According to the screening targets, different types of libraries are needed. To find out the target RNAs of a specific protein, or the target proteins of a particular RNA molecule, a simple ms2 tag labeled RNA library or a fluorescent protein marked protein library needs to be constructed. To find out the RNA binding sites of a specific protein, a library that contains the test protein with different mutants is necessary to be built. Or to discover the precise sequence within an RNA molecule that can interact with a specific protein, a library of the fragmented RNA molecules is also needed.

As this RNA trap assay is a fluorescence-based method at the cellular level, the screening protocol can be designed as follows. The *lacO* array integrated BHK cells are seeded in 96-well plates, followed by transfected with RNA trap and the test libraries, fixed and stained. In order to acquire the images efficiently, an Operetta platform can be applied to collect the signals of the nucleus (DAPI), the RNA trap (GFP), and the test protein (RFP) of each cell as images. The

image data can be used for further automatic analysis, which has similar steps as the protocol that was established by the previous studies [Herce, Deng et al. 2013]. In general, the cells, which are in good condition and successfully express all three components, are picked up, and the fluorescence of GFP and RFP at the *lacO* loci of each cell are measured for relative intensity calculation. After that, we can easily arrange the mutant proteins or RNA fragments in the library according to their relative intensities and find out the molecules that have significant influences on the RNA-protein interaction. And these candidates may reflect the binding areas of test protein or RNA.



**Figure. 59 High-throughput screening of RNA-protein interaction**

Cells are seeded, transferred, fixed, and stained in the 96-well plates. Images of DAPI, GFP and RFP channels are acquired with a high-content screening system. The data are further analyzed by software to obtain the relative fluorescence intensity of each well that can be used to compare their binding abilities.

## 5 Annex



## 5.1 Abbreviations

3C	Chromosome conformation capture
ATP	Adenosine triphosphate
AD	Activation domain
Cas13a	CRISPR-associated protein 13 a
Cas9	CRISPR-associated protein 9
CB	Cajal body
CIP	Cold-Induced protein
CLIP	Crosslinking and immunoprecipitation
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
CSD	Cold-chock domain
DB	DNA binding domain
dCas13a	Deactivated Cas13a
dCas9	Deactivated Cas9
DDRP	DNA-dependent RNA polymerase
DMHBI	3, 5-diMethoxy-4-hydroxybenzlidene imidazolinone
DNA	Deoxyribonucleic acid
DPSS	Diode-pumped solid-state
dsRBD	Double-strand RNA binding domain
EGFP	Enhanced green fluorescent protein
EMSA	Electrophoresis mobility shift assay
ER	Endoplasmic reticulum
EZH2	Enhancer of zeste homolog 2
F3H	Fluorescence three-hybrid
FISH	Fluorescence in situ hybrid
FH	Fluorescence hybridization
FMN	Flavin mononucleotide
FP	Fluorescent protein
FRAP	Fluorescence recovery after photobleaching
GBP	GFP binding protein

GTP	Guanosine triphosphate
HA	Hemagglutinin
HAT	Histone acetyltransferase
HBI	4-hydroxybenzlidene imidazolinone
HCV	Hepacivirus C
HDAC	Histone deacetylase
HDV	Hepacivirus D
HEPN	Higher eukaryotes and prokaryotes nucleotide-binding domain
HIV	Human immunodeficiency virus
HMT	Histone methyltransferase
HOTAIR	HOX transcript antisense RNA
IF	Immunofluorescence
INM	Inner nuclear membrane
IRES	Internal ribosome entry site
KH	K-homology
LacI	Lac inhibitor
<i>lacO</i>	Lac operon
LSD1	Lysine-specific demethylase 1
MCP	MS2 coat protein
MIN	Multifunctional integrase
NB	Nuclear body
ncRNA	Non-coding RNA
NE	Nuclear envelope
NES	Nuclear export sequence
NGD	No-go decay
NORAD	Non-coding RNA activated by DNA damage
NPC	Nuclear pore complex
NSD	No-stop decay
ONM	Outer nuclear membrane
ORF	Open reading frame
PABP	Poly(A) binding protein
PAM	Protospacer adjacent motif

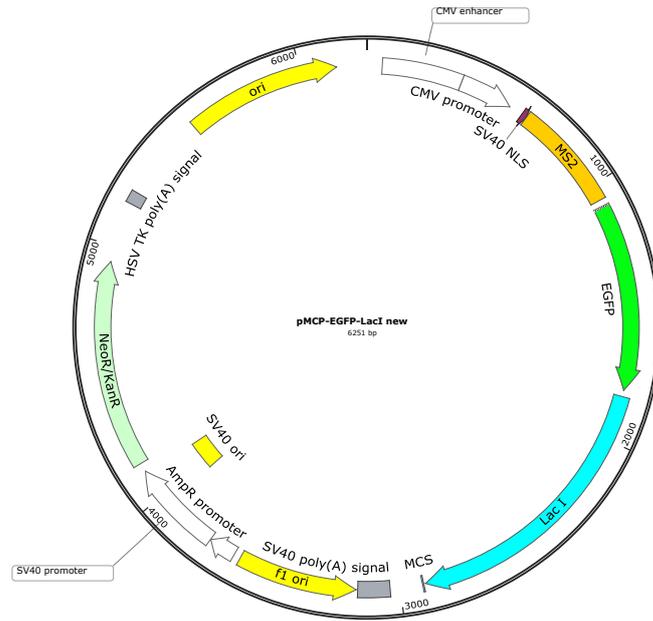
PCP	PP7 coat protein
PMT	Photomultiplier tube
PPI	Protein-protein interaction
PRC2	Polycomb repression complex 2
PUF	Pumilio and FBF homology
PUM2	Pumilio RNA binding family member 2
RDDP	RNA-dependent DNA polymerase
RDRP	RNA-dependent RNA polymerase
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RPI	RNA-protein interaction
RRM	RNA recognition motif
SELEX	Systematic evolution of ligands by exponential enrichment
snRNP	Small nuclear ribonucleoprotein
SV40	Simian vacuolating virus 40
TALEN	Transcription activator-like effector nuclease
TRAMP	Trf4/Air2/Mtr4p polyadenylation complex
TREX	Transcription/export complexes
UTR	Untranslational region
ZFN	Zinc finger nuclease



## 5.2 Plasmids used in this study

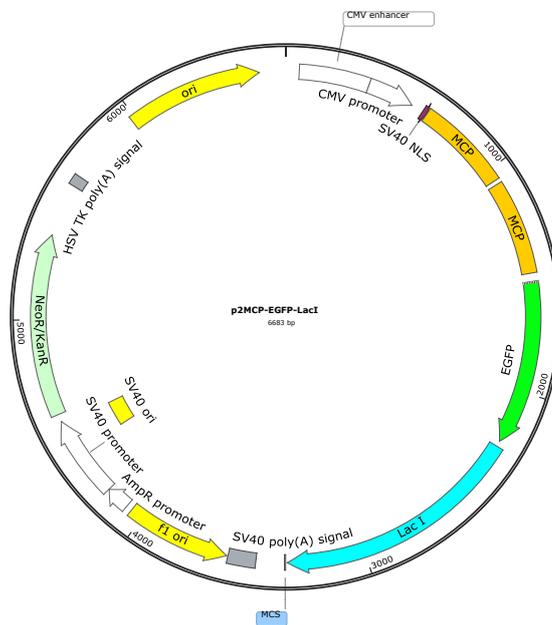
Catalog	Name
pc4555	pMCP-EGFP-LacI
pc4556	p2MCP-EGFP-LacI
pc4557	pMCP-EGFP-Lamin B1
pc4558	pMCP-EGFP-dCas9
pc4559	pMCP-EGFP-Coilin
pc4561	pPUF-EGFP-LacI
pc4562	pdCas13a-EGFP-LacI
pc4563	pCMV-ms2-H300
pc4564	pCMV-ms2
pc4565	pCMV-ms2-pp7
pc4566	pCMV-ms2-H1-100
pc4567	pCMV-ms2-H101-200
pc4568	pCMV-ms2-H201-300
pc4569	pCMV-ms2-H1-200
pc4571	pCMV-ms2-H101-300
pc4572	pCMV-ms2-NORAD
pc4573	pCMV-ms2-PABPC1
pc4574	pEx-A-U6-ms2
pc4575	pEx-A-U6-ms2-pp7
pc4576	pEx-A-U6-pp7
pc4577	pMCP-mCherry
pc4578	pMCP-mScarlet-i
pc4579	pMCPS47R-mCherry
pc4581	pPCP-mCherry
pc4582	pPCP-mScarlet-i
pc4583	pPCPV83Y-mCherry
pc4584	pEZH2-mCherry
pc4585	pEZH2N-mCherry
pc4586	pEZH2NT345A-mCherry
pc4587	pEZH2NT345D-mCherry
pc4588	pPABPC1-mCherry
pc4589	pPUM2-mCherry
pc4591	pHA-MCP

Created with SnapGene®

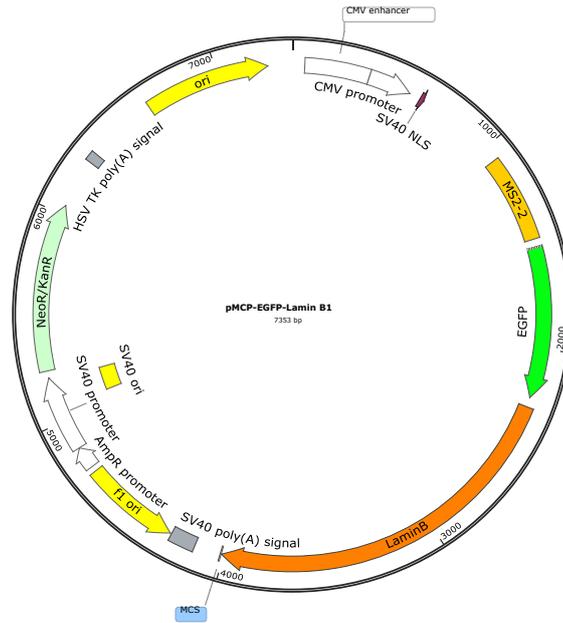


pc4555 / pMCP-EGFP-LacI

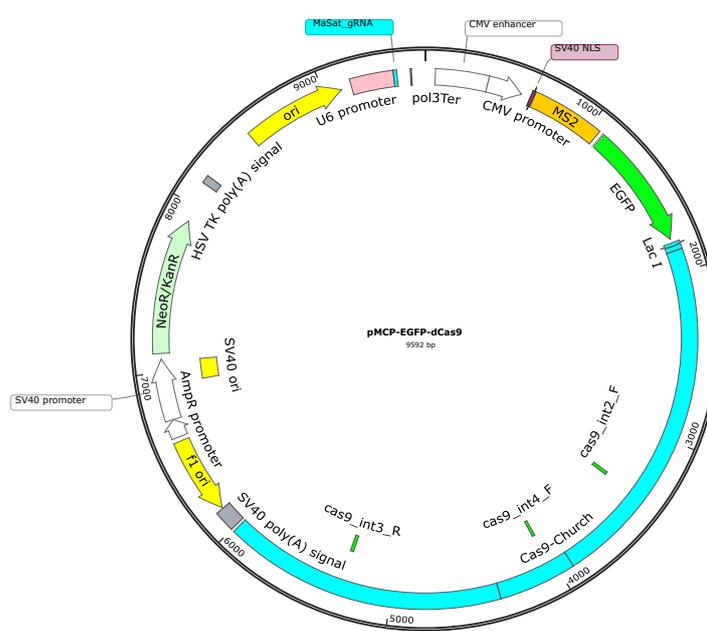
Created with SnapGene®



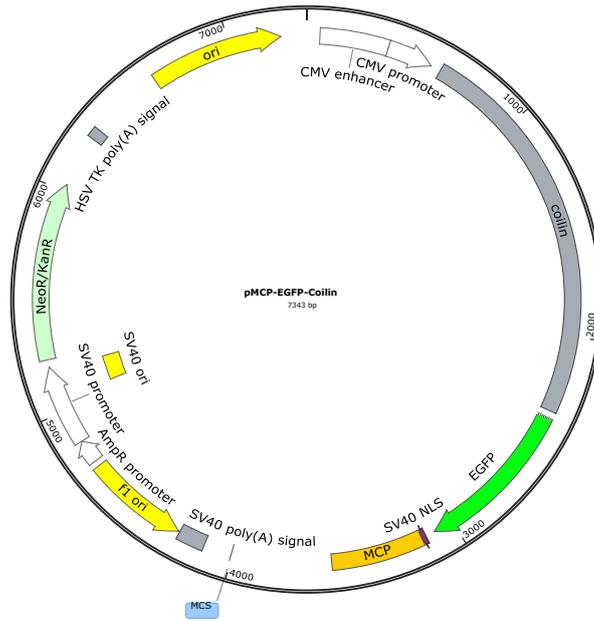
pc4556 / p2MCP-EGFP-LacI



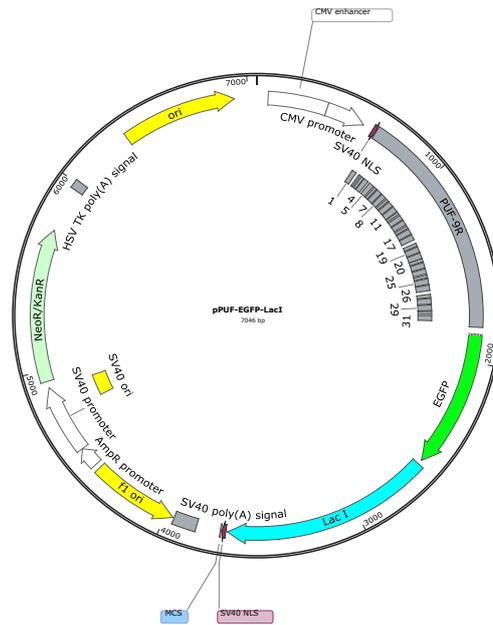
pc4557 / pMCP-EGFP-Lamin B1



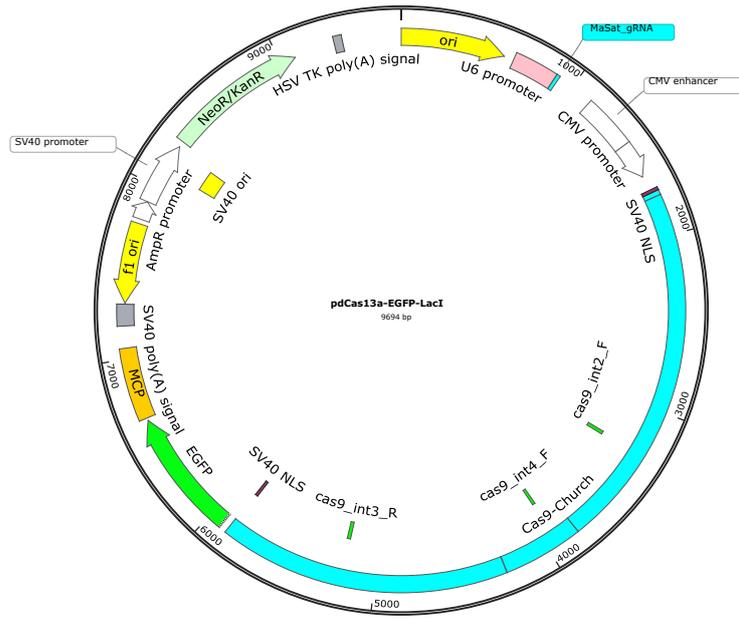
pc4558 / pMCP-EGFP-dCas9



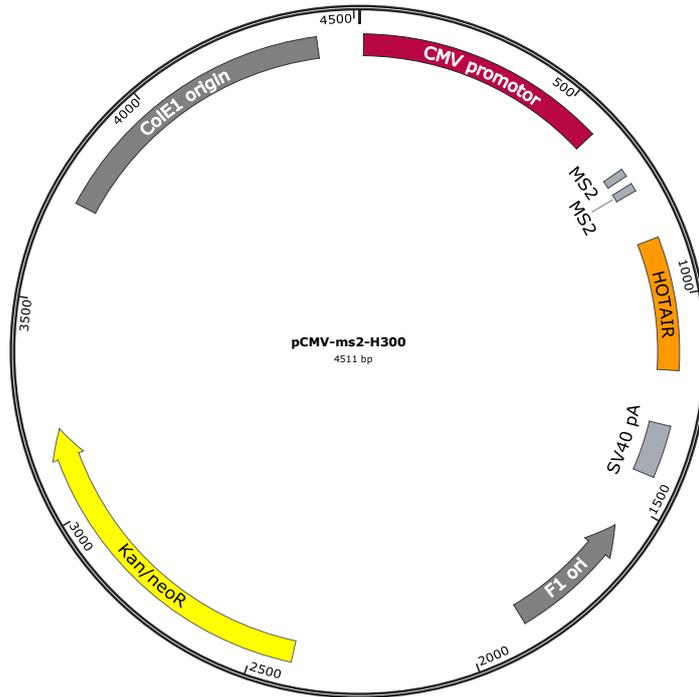
pc4559 / pMCP-EGFP-Coilin



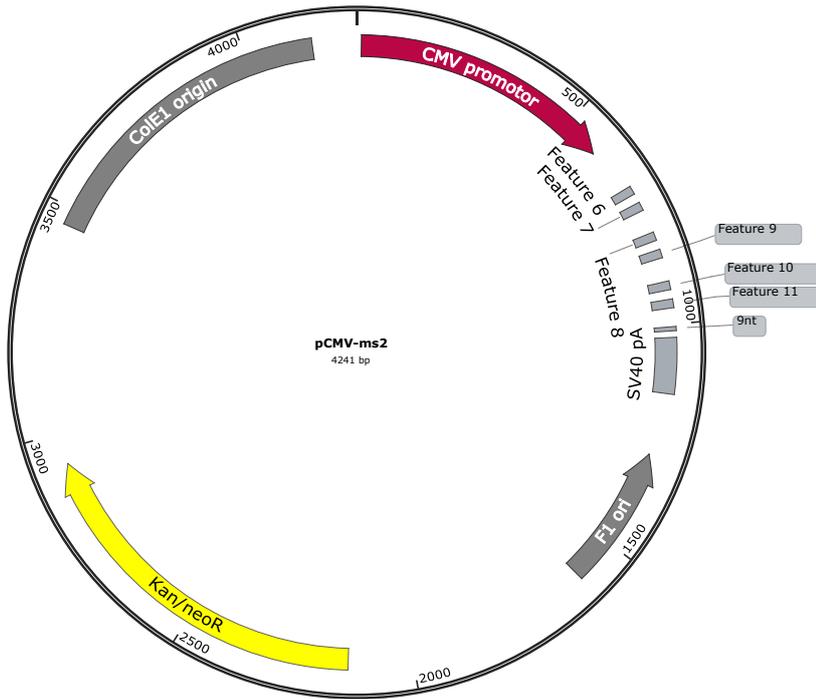
pc4561 / pPUF-EGFP-LacI



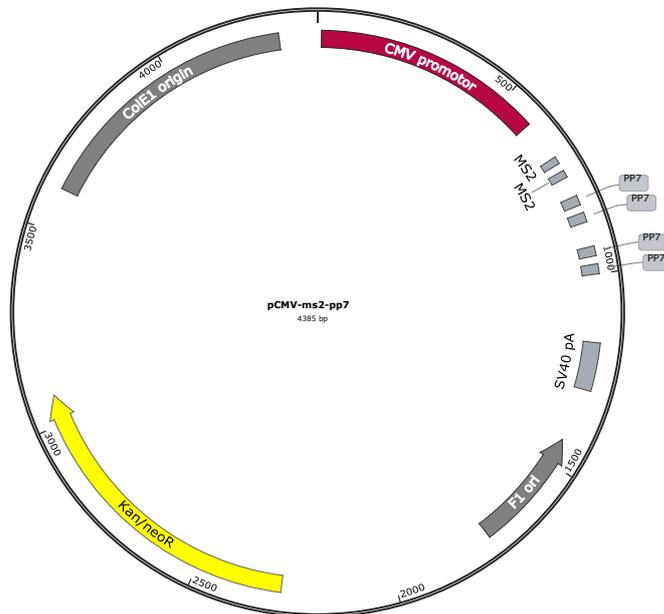
pc4562 / pdCas13a-EGFP-LacI



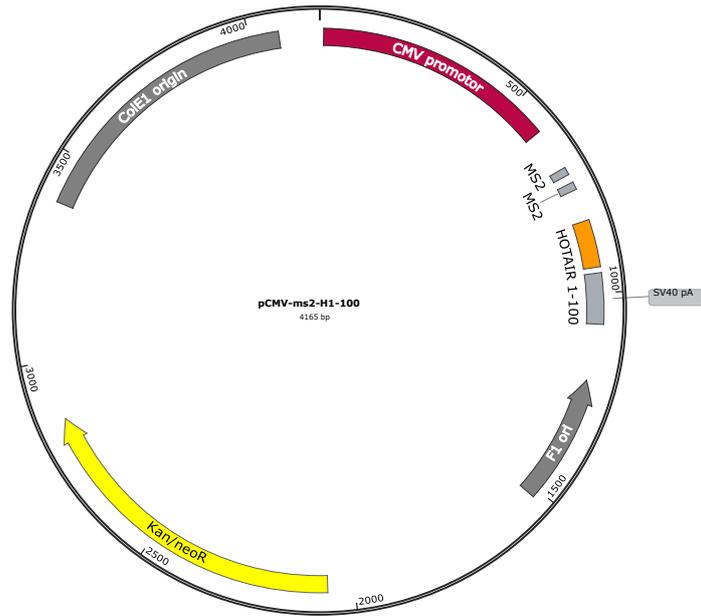
pc4563 / pCMV-ms2-H300



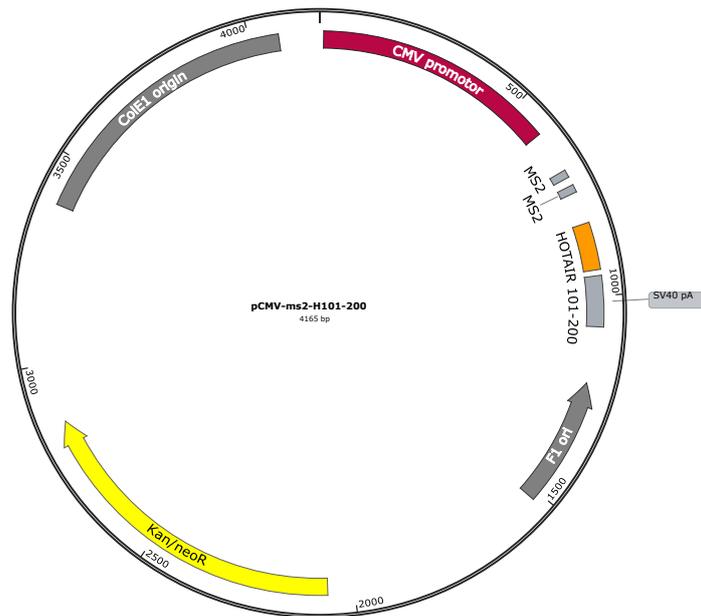
pc4564 / pCMV-ms2



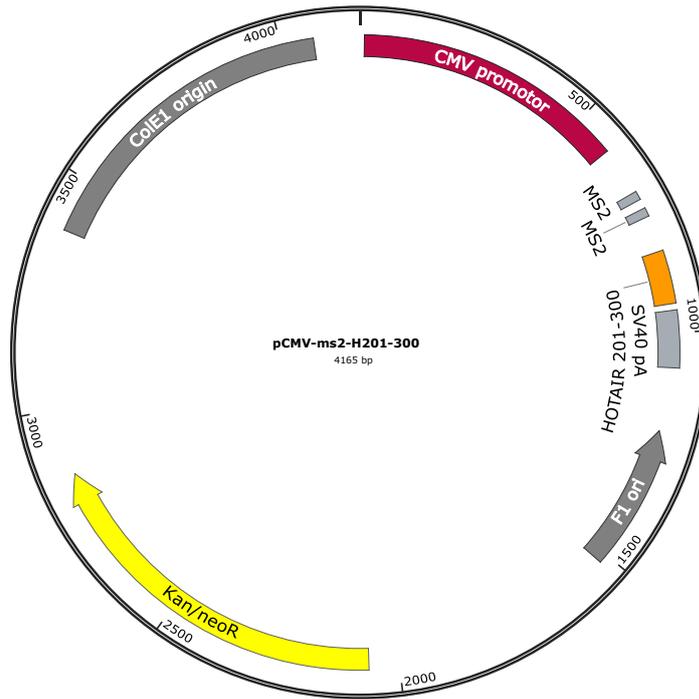
pc4565 / pCMV-ms2-pp7



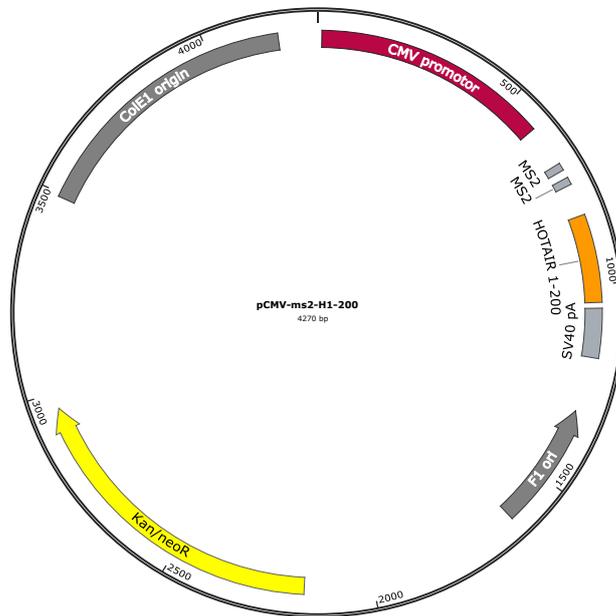
pc4566 / pCMV-ms2-H1-100



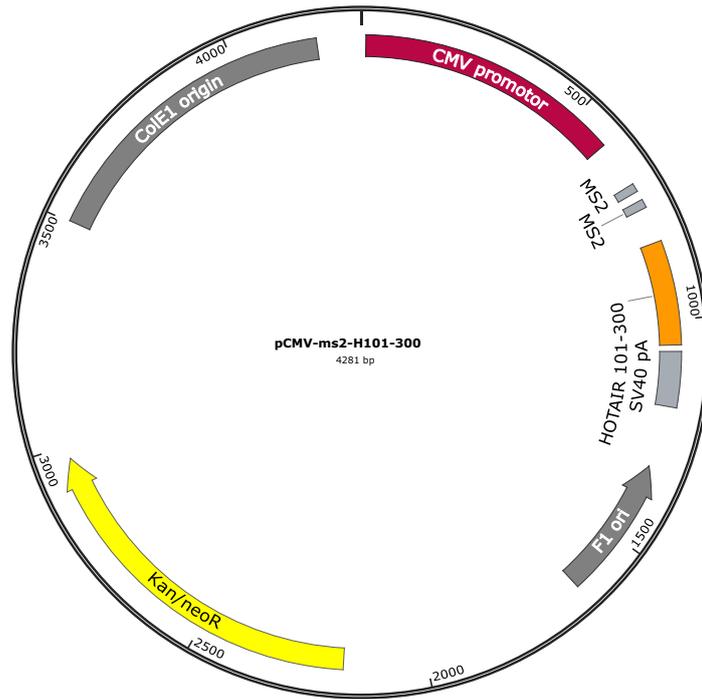
pc4567 / pCMV-ms2-H101-200



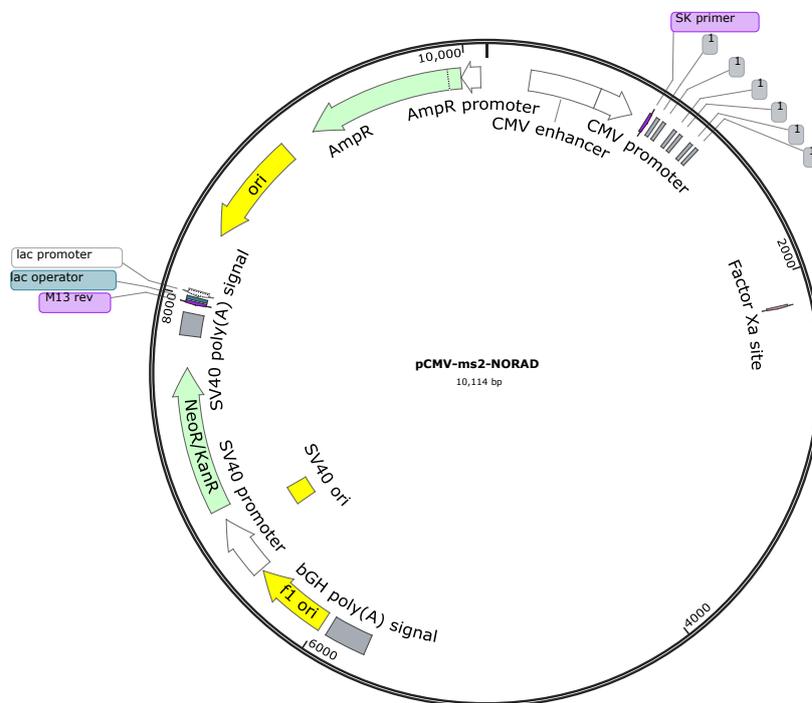
pc4568 / pCMV-ms2-H201-300



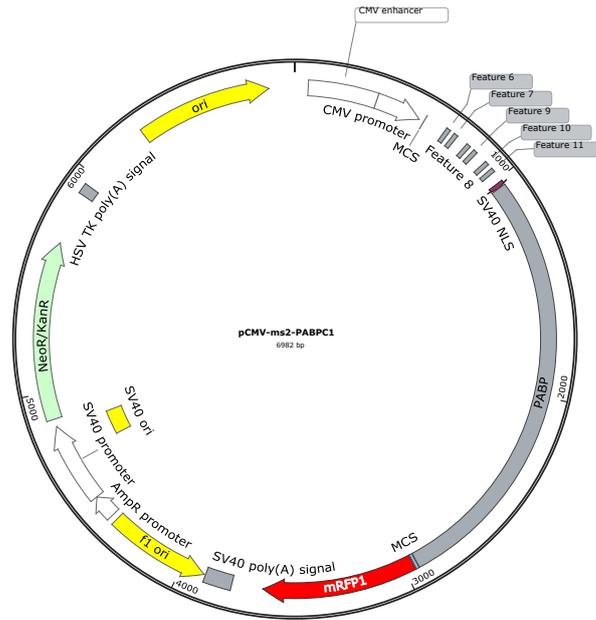
pc4569 / pCMV-ms2-H1-200



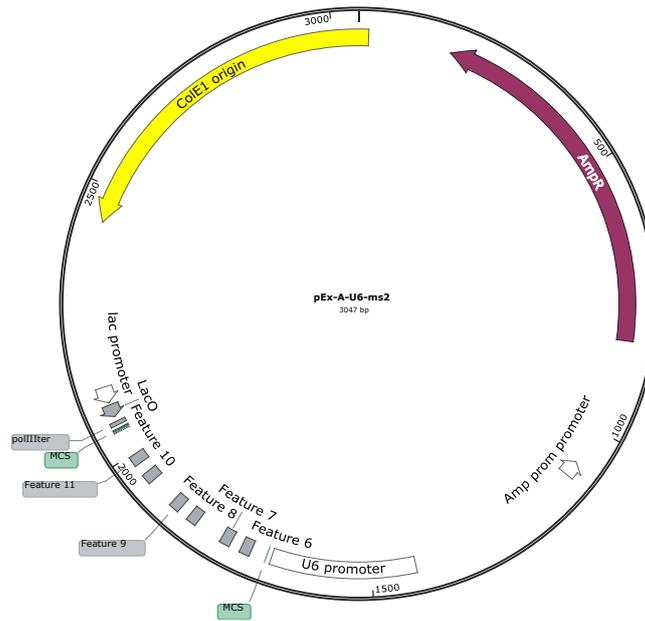
pc4571 / pCMV-ms2-H101-300



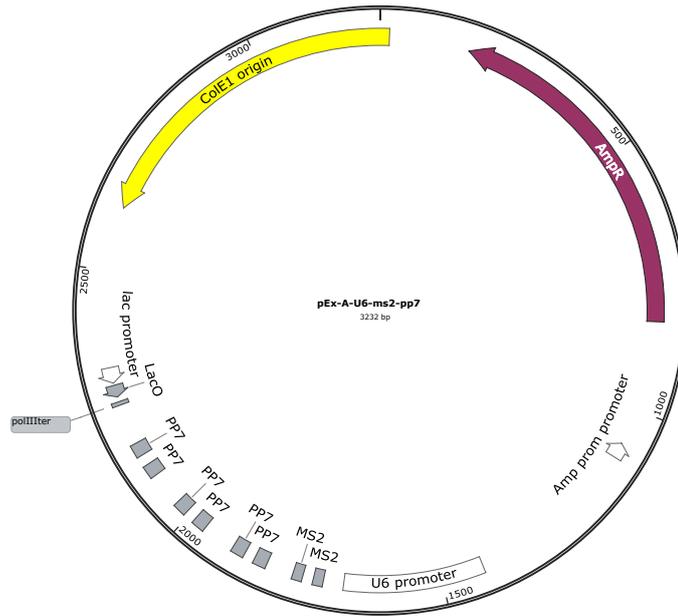
pc4572 / pCMV-ms2-NORAD



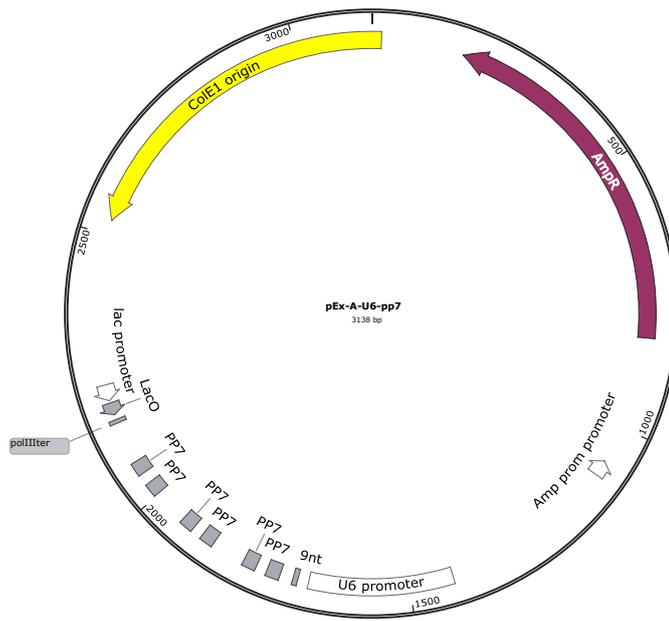
pc4573 / pCMV-ms2-PABPC1



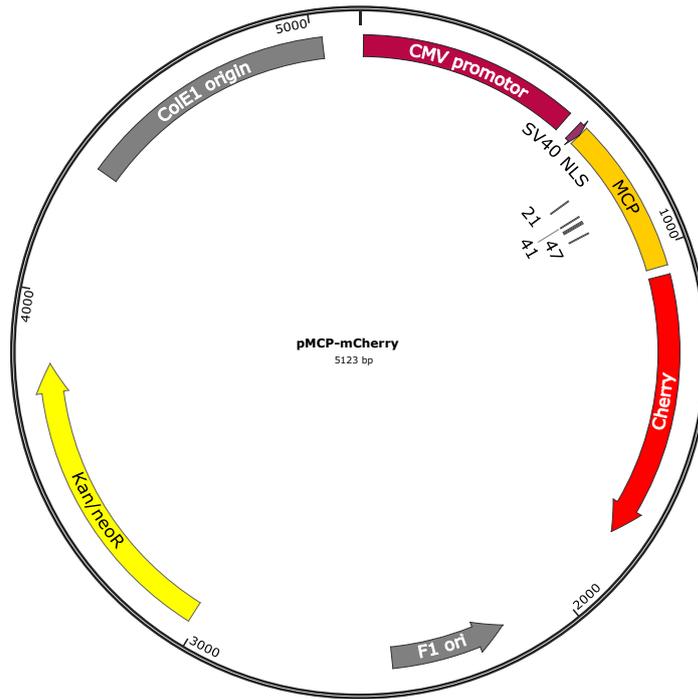
pc4574 / pEx-A-U6-ms2



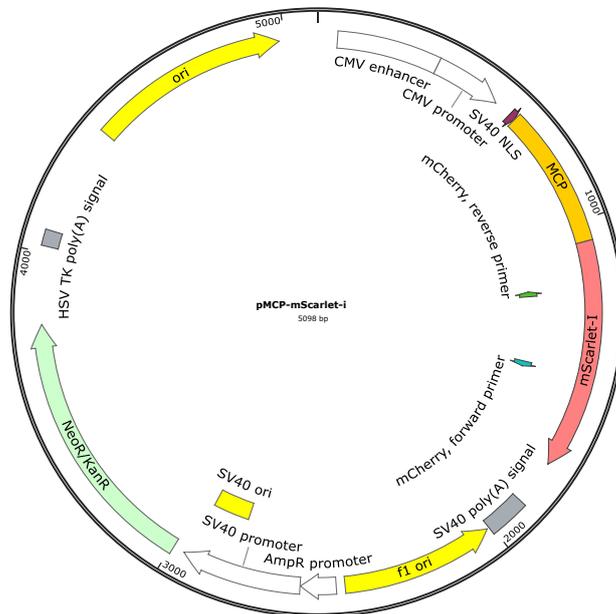
pc4575 / pEx-A-U6-ms2-pp7



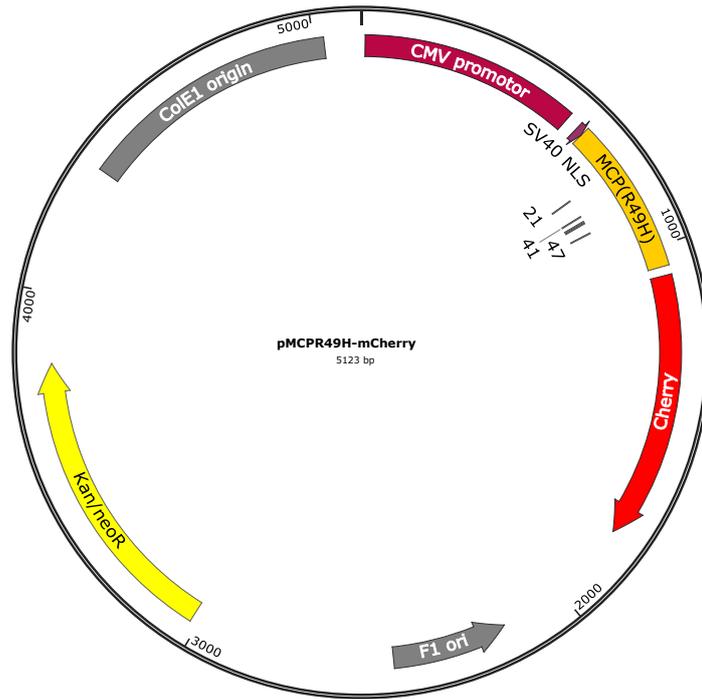
pc4576 / pEx-A-U6-pp7



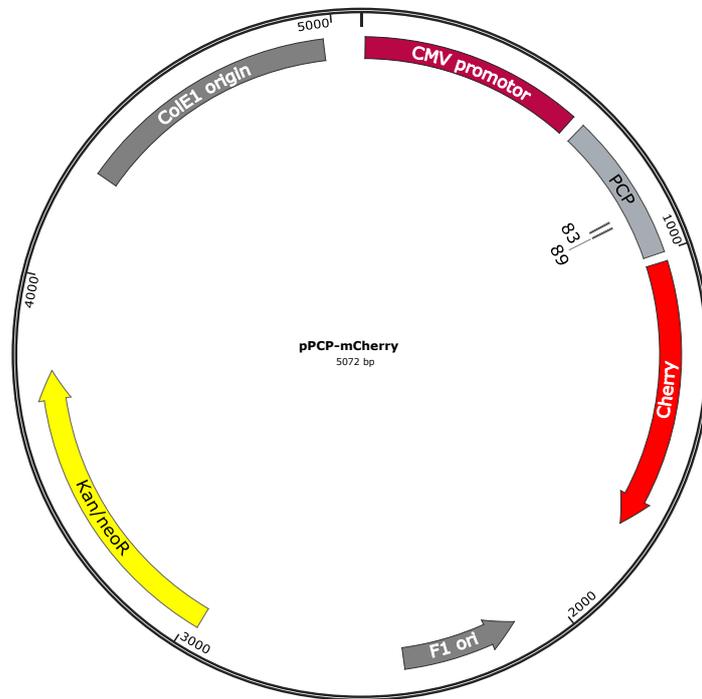
pc4577 / pMCP-mCherry



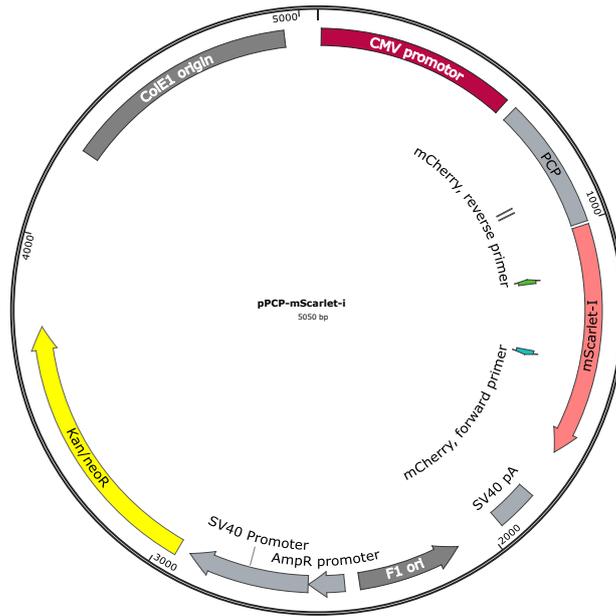
pc4578 / pMCP-mScarlet-i



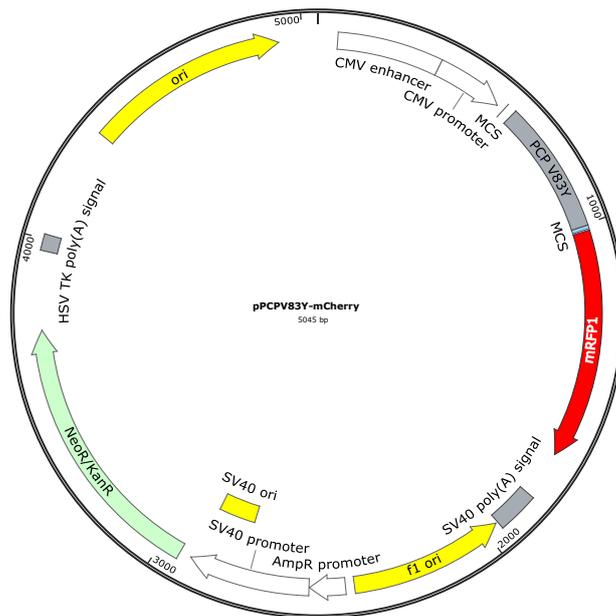
pc4579 / pMCPR49H-mCherry



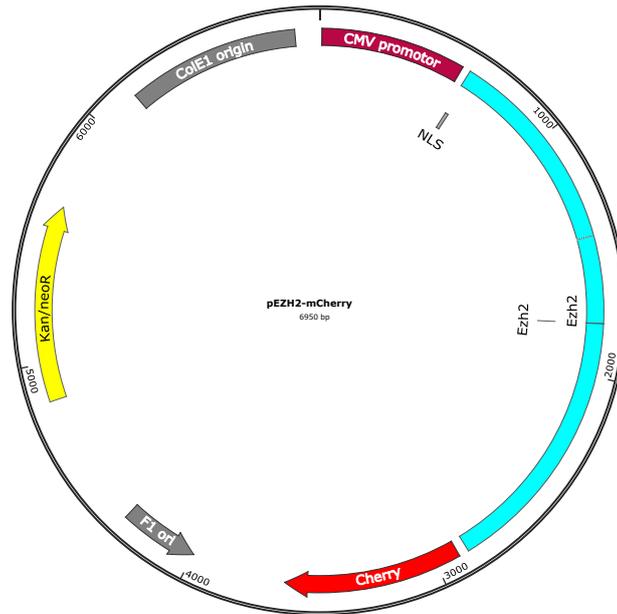
pc4581 / pPCP-mCherry



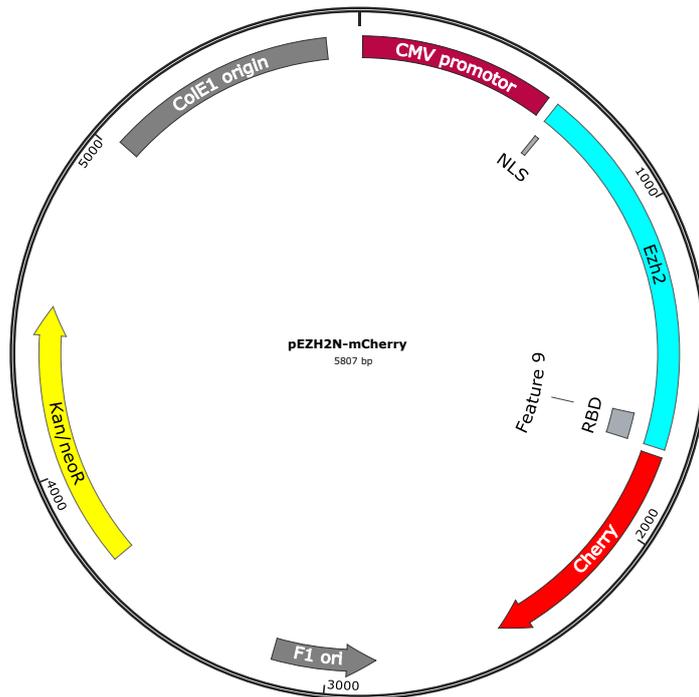
pc4582 / pPCP-mScarlet-i



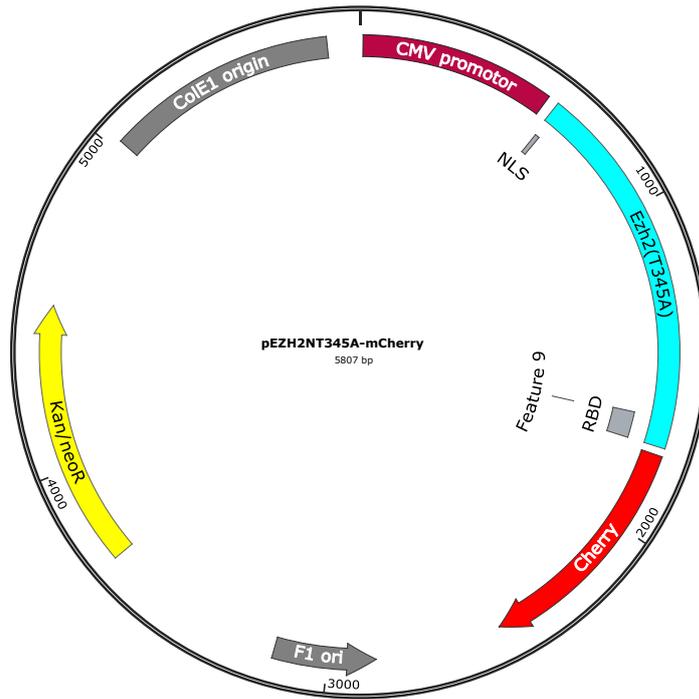
pc4583 / pPCPV83Y-mCherry



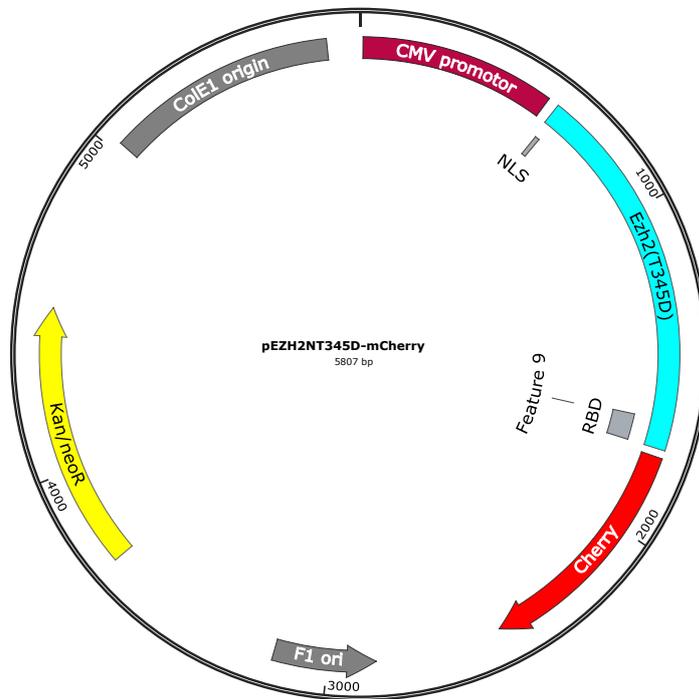
pc4584 / pEZH2-mCherry



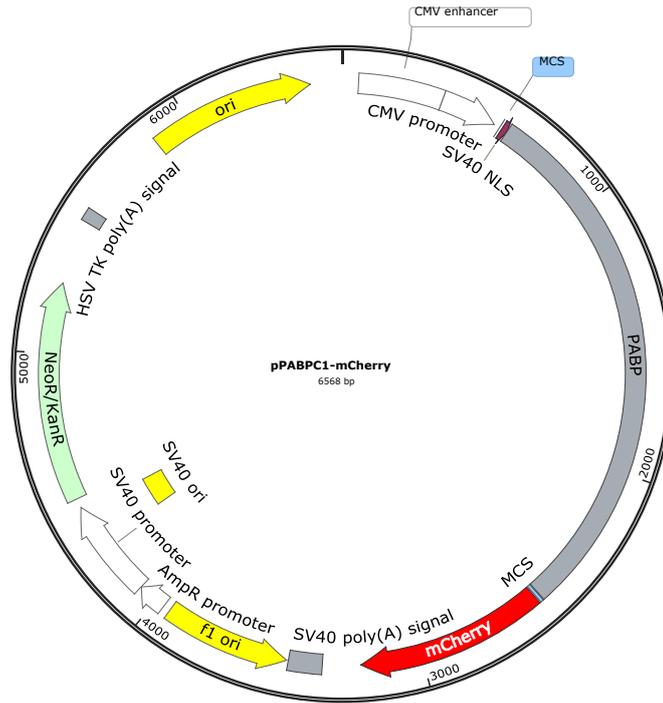
pc4585 / pEZH2N-mCherry



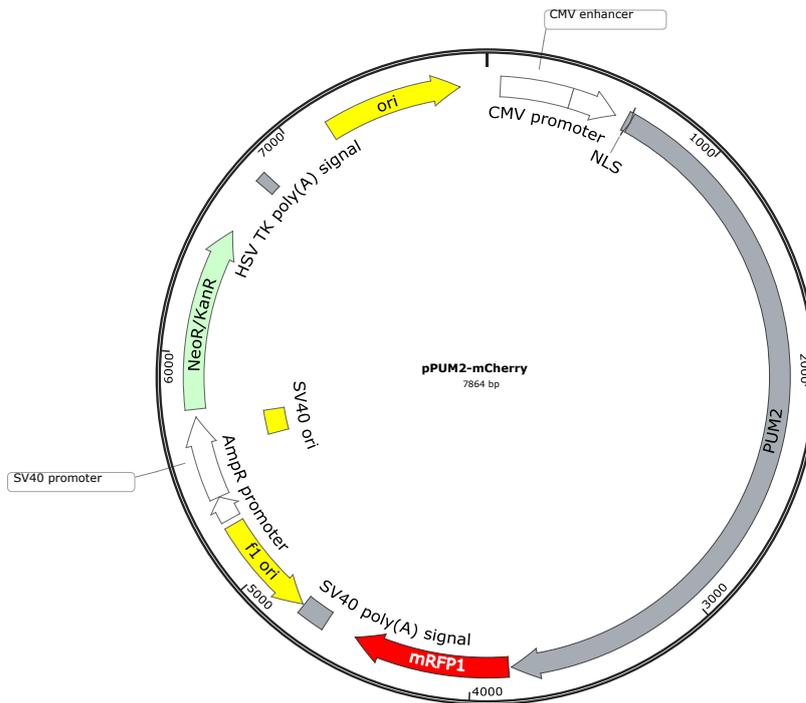
pc4586 / pEZH2NT345A-mCherry



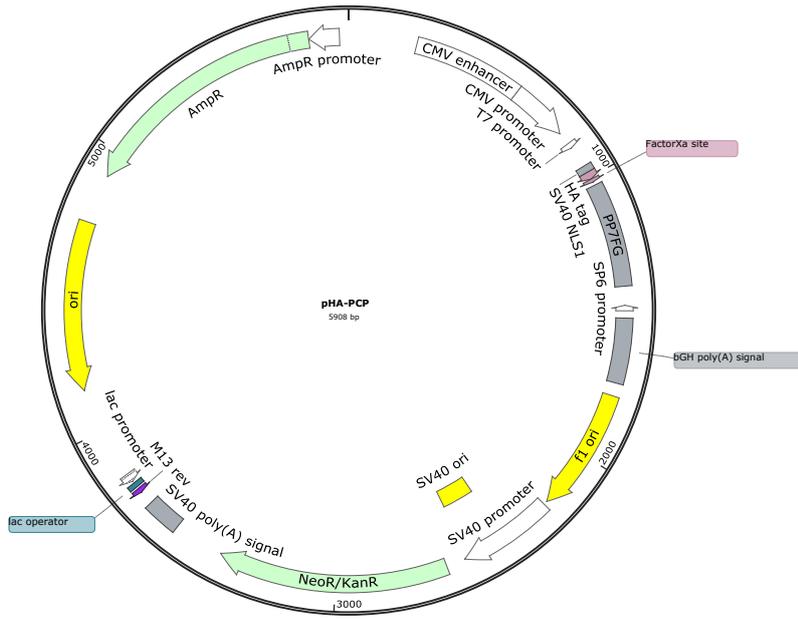
pc4587 / pEZH2NT345D-mCherry



pc4588 / pPABPC1-mCherry



pc4589 / pPUM2-mCherry



pc4591 / pHA-PCP

### 5.3 References

- Abudayyeh, O. O., J. S. Gootenberg, P. Essletzbichler, S. Han, J. Joung, J. J. Belanto, V. Verdine, D. B. Cox, M. J. Kellner and A. Regev (2017). "RNA targeting with CRISPR–Cas13." *Nature* **550**(7675): 280.
- Alber, F., S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprpto, O. Karni-Schmidt, R. Williams and B. T. Chait (2007). "The molecular architecture of the nuclear pore complex." *Nature* **450**(7170): 695.
- Alexopoulou, A. N., J. R. Couchman and J. R. Whiteford (2008). "The CMV early enhancer/chicken  $\beta$  actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors." *BMC cell biology* **9**(1): 2.
- Allmang, C., P. Mitchell, E. Petfalski and D. Tollervey (2000). "Degradation of ribosomal RNA precursors by the exosome." *Nucleic acids research* **28**(8): 1684-1691.
- Anton, T., S. Bultmann, H. Leonhardt and Y. Markaki (2014). "Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system." *Nucleus* **5**(2): 163-172.
- Armitage, B. A. (2011). "Imaging of RNA in live cells." *Current opinion in chemical biology* **15**(6): 806-812.
- Arts, G. J., S. Kuersten, P. Romby, B. Ehresmann and I. W. Mattaj (1998). "The role of exportin-t in selective nuclear export of mature tRNAs." *The EMBO journal* **17**(24): 7430-7441.
- Auffinger, P. and E. Westhof (1997). "Rules governing the orientation of the 2'-hydroxyl group in RNA." *Journal of molecular biology* **274**(1): 54-63.
- Auweter, S. D., F. C. Oberstrass and F. H.-T. Allain (2006). "Sequence-specific binding of single-stranded RNA: is there a code for recognition?" *Nucleic acids research* **34**(17): 4943-4959.
- Babendure, J. R., S. R. Adams and R. Y. Tsien (2003). "Aptamers switch on fluorescence of triphenylmethane dyes." *Journal of the American Chemical Society* **125**(48): 14716-14717.
- Baron, A. T., T. M. Greenwood, C. W. Bazinet and J. L. Salisbury (1992). "Centrin is a component of the pericentriola lattice." *Biology of the Cell* **76**: 383-388.

- Beau, I., A. Esclatine and P. Codogno (2008). "Lost to translation: when autophagy targets mature ribosomes." *Trends in cell biology* **18**(7): 311-314.
- Belasco, J. G. (1993). *mRNA degradation in prokaryotic cells: an overview*, Academic Press, New York.
- Ben-Shem, A., N. G. de Loubresse, S. Melnikov, L. Jenner, G. Yusupova and M. Yusupov (2011). "The structure of the eukaryotic ribosome at 3.0 Å resolution." *Science* **334**(6062): 1524-1529.
- Benkovic, S. J. and S. Hammes-Schiffer (2003). "A perspective on enzyme catalysis." *Science* **301**(5637): 1196-1202.
- Bentley, D. (1999). "Coupling RNA polymerase II transcription with pre-mRNA processing." *Current opinion in cell biology* **11**(3): 347-351.
- Bergman, N. H., N. C. Lau, V. Lehnert, E. Westhof and D. P. Bartel (2004). "The three-dimensional architecture of the class I ligase ribozyme." *RNA* **10**(2): 176-184.
- Bertrand, E., P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer and R. M. Long (1998). "Localization of ASH1 mRNA particles in living yeast." *Molecular cell* **2**(4): 437-445.
- Björk, G. R., J. U. Ericson, C. E. Gustafsson, T. G. Hagervall, Y. H. Jönsson and P. M. Wikström (1987). "Transfer RNA modification." *Annual review of biochemistry* **56**(1): 263-285.
- Blackburn, E. H. (1991). "Structure and function of telomeres." *Nature* **350**(6319): 569.
- Borchert, G. M., W. Lanier and B. L. Davidson (2006). "RNA polymerase III transcribes human microRNAs." *Nature structural & molecular biology* **13**(12): 1097.
- Bouvet, P., K. Matsumoto and A. P. Wolffe (1995). "Sequence-specific RNA recognition by the *Xenopus* Y-box proteins. An essential role for the cold shock domain." *The Journal of biological chemistry* **270**(47): 28297-28303.
- Bouvier, N. M. and P. Palese (2008). "The biology of influenza viruses." *Vaccine* **26**: D49-D53.
- Brennicke, A., A. Marchfelder and S. Binder (1999). "RNA editing." *FEMS Microbiology Reviews* **23**(3): 297-316.
- Brown, R. S. (2005). "Zinc finger proteins: getting a grip on RNA." *Current opinion in structural biology* **15**(1): 94-98.
- Bullock, S. L. and D. Ish-Horowicz (2001). "Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis." *Nature* **414**(6864): 611.

- Butcher, S. E. and A. M. Pyle (2011). "The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks." *Accounts of chemical research* **44**(12): 1302-1311.
- Cabili, M. N., M. C. Dunagin, P. D. McClanahan, A. Biaisch, O. Padovan-Merhar, A. Regev, J. L. Rinn and A. Raj (2015). "Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution." *Genome biology* **16**(1): 20.
- Cai, Y., X. Yu, S. Hu and J. Yu (2009). "A brief review on the mechanisms of miRNA regulation." *Genomics, proteomics & bioinformatics* **7**(4): 147-154.
- Casey, J. L. and J. L. Gerin (1995). "Hepatitis D virus RNA editing: specific modification of adenosine in the antigenomic RNA." *Journal of Virology* **69**(12): 7593-7600.
- Cech, T. R. (2012). "The RNA worlds in context." *Cold Spring Harbor perspectives in biology* **4**(7): a006742.
- Chambon, P. (1975). "Eukaryotic nuclear RNA polymerases." *Annual review of biochemistry* **44**(1): 613-638.
- Chao, J. A., Y. Patskovsky, S. C. Almo and R. H. Singer (2008). "Structural basis for the coevolution of a viral RNA–protein complex." *Nature structural & molecular biology* **15**(1): 103.
- Chekanova, J. A. and D. A. Belostotsky (2003). "Evidence that poly (A) binding protein has an evolutionarily conserved function in facilitating mRNA biogenesis and export." *Rna* **9**(12): 1476-1490.
- Chen, B., L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman and L. S. Qi (2014). "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system." *Cell* **156**(1): 373.
- Cheng, L., P. R. Ziegelhoffer and N. S. Yang (1993). "In vivo promoter activity and transgene expression in mammalian somatic tissues evaluated by using particle bombardment." *Proceedings of the National Academy of Sciences* **90**(10): 4455-4459.
- Cheong, C.-G. and T. M. T. Hall (2006). "Engineering RNA sequence specificity of Pumilio repeats." *Proceedings of the National Academy of Sciences* **103**(37): 13635-13639.

- Coates, P. and P. Hall (2003). "The yeast two-hybrid system for identifying protein–protein interactions." *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **199**(1): 4-7.
- Cornish-Bowden, A. (2014). *Principles of enzyme kinetics*, Elsevier.
- Cox, D. B. T., J. S. Gootenberg, O. O. Abudayyeh, B. Franklin, M. J. Kellner, J. Joung and F. Zhang (2017). "RNA editing with CRISPR-Cas13." *Science* **358**(6366): 1019-1027.
- Cox, K. H., D. V. DeLeon, L. M. Angerer and R. C. Angerer (1984). "Detection of mRNAs in sea urchin embryos by in situ hybridization using asymmetric RNA probes." *Developmental biology* **101**(2): 485-502.
- Crick, F. (1966). "Codon-anticodon pairing: the wobble hypothesis."
- Crick, F. (1970). "Central dogma of molecular biology." *Nature* **227**(5258): 561.
- Daubner, G. M., A. Cléry and F. H. Allain (2013). "RRM–RNA recognition: NMR or crystallography... and new findings." *Current opinion in structural biology* **23**(1): 100-108.
- Davidovich, C. and T. R. Cech (2015). "The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2." *Rna* **21**(12): 2007-2022.
- Davis, I. and D. Ish-Horowicz (1991). "Apical localization of pair-rule transcripts requires 3' sequences and limits protein diffusion in the *Drosophila* blastoderm embryo." *Cell* **67**(5): 927-940.
- Deutscher, M. P. (1984). "Processing of tRNA in Prokaryotes and Eukaryote." *Critical Reviews in Biochemistry* **17**(1): 45-71.
- Didier, D. K., J. Schifffenbauer, S. L. Woulfe, M. Zacheis and B. D. Schwartz (1988). "Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box." *Proceedings of the National Academy of Sciences* **85**(19): 7322-7326.
- Doherty, E. A. and J. A. Doudna (2001). "Ribozyme structures and mechanisms." *Annual review of biophysics and biomolecular structure* **30**(1): 457-475.
- Dong, S., Y. Wang, C. Cassidy-Amstutz, G. Lu, R. Bigler, M. R. Jezyk, C. Li, T. M. T. Hall and Z. Wang (2011). "Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains." *Journal of Biological Chemistry* **286**(30): 26732-26742.

- Doxsey, S. (2001). "Re-evaluating centrosome function." *Nature Reviews Molecular Cell Biology* **2**(9): 688-698.
- Doyle, M. and M. F. Jantsch (2002). "New and old roles of the double-stranded RNA-binding domain." *Journal of structural biology* **140**(1-3): 147-153.
- Draper, D. E., I. C. Deckman and J. V. Vartikar (1988). [13] Physical studies of ribosomal protein—RNA interactions. *Methods in enzymology*, Elsevier. **164**: 203-220.
- Draper, D. E., D. Grilley and A. M. Soto (2005). "Ions and RNA folding." *Annu. Rev. Biophys. Biomol. Struct.* **34**: 221-243.
- Dunkle, J. A., L. Wang, M. B. Feldman, A. Pulk, V. B. Chen, G. J. Kapral, J. Noeske, J. S. Richardson, S. C. Blanchard and J. H. D. Cate (2011). "Structures of the bacterial ribosome in classical and hybrid states of tRNA binding." *Science* **332**(6032): 981-984.
- Eddy, S. R. (2001). "Non-coding RNA genes and the modern RNA world." *Nature Reviews Genetics* **2**(12): 919.
- Eichler, D. C. and N. Craig (1994). Processing of eukaryotic ribosomal RNA. *Progress in nucleic acid research and molecular biology*, Elsevier. **49**: 197-239.
- Ekland, E., J. Szostak and D. Bartel (1995). "Structurally complex and highly active RNA ligases derived from random RNA sequences." *Science* **269**(5222): 364-370.
- Ellington, A. D. and J. W. Szostak (1990). "In vitro selection of RNA molecules that bind specific ligands." *nature* **346**(6287): 818.
- Ermolenko, D. and G. Makhatadze (2002). "Bacterial cold-shock proteins." *Cellular and Molecular Life Sciences CMLS* **59**(11): 1902-1913.
- Fedor, M. J. (2002). "The role of metal ions in RNA catalysis." *Current opinion in structural biology* **12**(3): 289-295.
- Fedor, M. J. and J. R. Williamson (2005). "The catalytic diversity of RNAs." *Nature reviews Molecular cell biology* **6**(5): 399.
- Fields, S. and O.-k. Song (1989). "A novel genetic system to detect protein—protein interactions." *Nature* **340**(6230): 245.
- Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. M. Jou, F. Molemans, A. Raeymaekers and A. Van den Berghe (1976). "Complete nucleotide sequence

of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene." *Nature* **260**(5551): 500.

Filipovska, A., M. F. Razif, K. K. Nygård and O. Rackham (2011). "A universal code for RNA recognition by PUF proteins." *Nature chemical biology* **7**(7): 425.

Filonov, G. S., J. D. Moon, N. Svensen and S. R. Jaffrey (2014). "Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution." *Journal of the American Chemical Society* **136**(46): 16299-16308.

Fitzgerald, M. and T. Shenk (1981). "The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs." *Cell* **24**(1): 251-260.

Freed, E. O. (2001). "HIV-1 replication." *Somatic cell and molecular genetics* **26**(1-6): 13-33.

Gaj, T., C. A. Gersbach and C. F. Barbas III (2013). "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering." *Trends in biotechnology* **31**(7): 397-405.

Gall, J. G. (2003). "The centennial of the Cajal body." *Nature reviews Molecular cell biology* **4**(12): 975.

Gascoigne, K. E., K. Takeuchi, A. Suzuki, T. Hori, T. Fukagawa and I. M. Cheeseman (2011). "Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes." *Cell* **145**(3): 410-422.

Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss (2008). "RNA-binding proteins and post-transcriptional gene regulation." *FEBS letters* **582**(14): 1977-1986.

Gomez, D. E., R. G. Armando, H. G. Farina, P. L. Menna, C. S. Cerrudo, P. D. Ghiringhelli and D. F. Alonso (2012). "Telomere structure and telomerase in health and disease." *International journal of oncology* **41**(5): 1561-1569.

Gopinath, S. C. B. (2007). "Methods developed for SELEX." *Analytical and bioanalytical chemistry* **387**(1): 171-182.

Graumann, P. L. and M. A. Marahiel (1998). "A superfamily of proteins that contain the cold-shock domain." *Trends in biochemical sciences* **23**(8): 286-290.

Greenberg, J. R. (1979). "Ultraviolet light-induced crosslinking of mRNA to proteins." *Nucleic acids research* **6**(2): 715-732.

- Gruenbaum, Y., A. Margalit, R. D. Goldman, D. K. Shumaker and K. L. Wilson (2005). "The nuclear lamina comes of age." *Nature reviews Molecular cell biology* **6**(1): 21.
- Guenatri, M., D. Bailly, C. Maison and G. Almouzni (2004). "Mouse centric and pericentric satellite repeats form distinct functional heterochromatin." *The Journal of cell biology* **166**(4): 493-505.
- Hall, T. M. T. (2005). "Multiple modes of RNA recognition by zinc finger proteins." *Current opinion in structural biology* **15**(3): 367-373.
- Hammes, G. G., S. J. Benkovic and S. Hammes-Schiffer (2011). "Flexibility, diversity, and cooperativity: pillars of enzyme catalysis." *Biochemistry* **50**(48): 10422-10430.
- Han, K. Y., B. J. Leslie, J. Fei, J. Zhang and T. Ha (2013). "Understanding the photophysics of the spinach–DFHBI RNA aptamer–fluorogen complex to improve live-cell RNA imaging." *Journal of the American Chemical Society* **135**(50): 19033-19038.
- Handwerger, K. E. and J. G. Gall (2006). "Subnuclear organelles: new insights into form and function." *Trends in cell biology* **16**(1): 19-26.
- Hansen, J. L., P. B. Moore and T. A. Steitz (2003). "Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit." *Journal of molecular biology* **330**(5): 1061-1075.
- Hellman, L. M. and M. G. Fried (2007). "Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions." *Nature protocols* **2**(8): 1849.
- Henras, A. K., C. Plisson-Chastang, M. F. O'Donohue, A. Chakraborty and P. E. Gleizes (2015). "An overview of pre-ribosomal RNA processing in eukaryotes." *Wiley Interdisciplinary Reviews: RNA* **6**(2): 225-242.
- Herce, H. D., W. Deng, J. Helma, H. Leonhardt and M. C. Cardoso (2013). "Visualization and targeted disruption of protein interactions in living cells." *Nature communications* **4**: 2660.
- Higgins, N. and N. Cozzarelli (1982). "The binding of gyrase to DNA: analysis by retention by nitrocellulose filters." *Nucleic acids research* **10**(21): 6833-6847.
- Hodges, P. E. and J. D. Beggs (1994). "RNA splicing: U2 fulfils a commitment." *Current Biology* **4**(3): 264-267.

- Houseley, J., J. LaCava and D. Tollervey (2006). "RNA-quality control by the exosome." *Nature reviews Molecular cell biology* **7**(7): 529.
- Houseley, J. and D. Tollervey (2009). "The many pathways of RNA degradation." *Cell* **136**(4): 763-776.
- Hsu, P. D., E. S. Lander and F. Zhang (2014). "Development and applications of CRISPR-Cas9 for genome engineering." *Cell* **157**(6): 1262-1278.
- Huang, Y., R. Gattoni, J. Stévenin and J. A. Steitz (2003). "SR splicing factors serve as adapter proteins for TAP-dependent mRNA export." *Molecular cell* **11**(3): 837-843.
- Ishihama, A. (2000). "Functional modulation of Escherichia coli RNA polymerase." *Annual Reviews in Microbiology* **54**(1): 499-518.
- Isken, O. and L. E. Maquat (2008). "The multiple lives of NMD factors: balancing roles in gene and genome regulation." *Nature reviews Genetics* **9**(9): 699.
- Jaeger, S., G. Eriani and F. Martin (2004). "Results and prospects of the yeast three-hybrid system." *FEBS letters* **556**(1-3): 7-12.
- James, W. (2007). "Aptamers in the virologists' toolkit." *Journal of General Virology* **88**(2): 351-364.
- Jensen, K. B. and R. B. Darnell (2008). CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *RNA-Protein Interaction Protocols*, Springer: 85-98.
- Jiang, W., Y. Hou and M. Inouye (1997). "CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone." *Journal of Biological Chemistry* **272**(1): 196-202.
- Johansson, M. K., H. Fidder, D. Dick and R. M. Cook (2002). "Intramolecular dimers: a new strategy to fluorescence quenching in dual-labeled oligonucleotide probes." *Journal of the American Chemical Society* **124**(24): 6950-6956.
- Joho, K. E., M. K. Darby, E. T. Crawford and D. D. Brown (1990). "A finger protein structurally similar to TFIIIA that binds exclusively to 5S RNA in Xenopus." *Cell* **61**(2): 293-300.
- Joyce, G. F. (1989). "RNA evolution and the origins of life." *Nature* **338**(6212): 217.
- Joyce, G. F. (2002). "The antiquity of RNA-based evolution." *nature* **418**(6894): 214.

- Kaneko, S., G. Li, J. Son, C.-F. Xu, R. Margueron, T. A. Neubert and D. Reinberg (2010). "Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA." *Genes & development* **24**(23): 2615-2620.
- Keiler, K. C. (2011). "RNA localization in bacteria." *Current opinion in microbiology* **14**(2): 155-159.
- Kim, V. N. (2018). "RNA-targeting CRISPR comes of age." *Nature biotechnology* **36**(1): 44.
- Klappenbach, J. A., J. M. Dunbar and T. M. Schmidt (2000). "rRNA operon copy number reflects ecological strategies of bacteria." *Appl. Environ. Microbiol.* **66**(4): 1328-1333.
- Knott, G. J., A. East-Seletsky, J. C. Cofsky, J. M. Holton, E. Charles, M. R. O'Connell and J. A. Doudna (2017). "Guide-bound structures of an RNA-targeting A-cleaving CRISPR–Cas13a enzyme." *Nature structural & molecular biology* **24**(10): 825.
- Knowles, J. R. (1991). "Enzyme catalysis: not different, just better." *Nature* **350**(6314): 121.
- Konermann, S., M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg and H. Nishimasu (2015). "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex." *Nature* **517**(7536): 583.
- Konermann, S., P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev and P. D. Hsu (2018). "Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors." *Cell* **173**(3): 665-676. e614.
- Koning, R., S. van den Worm, J. R. Plaisier, J. van Duin, J. P. Abrahams and H. Koerten (2003). "Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2." *Journal of molecular biology* **332**(2): 415-422.
- Kressler, D., E. Hurt and J. Baßler (2010). "Driving ribosome assembly." *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* **1803**(6): 673-683.
- Krug, R. M., F. V. Alonso-Caplen, I. Julkunen and M. G. Katze (1989). Expression and replication of the influenza virus genome. *The influenza viruses*, Springer: 89-152.
- Le Hir, H., A. Nott and M. J. Moore (2003). "How introns influence and enhance eukaryotic gene expression." *Trends in biochemical sciences* **28**(4): 215-220.

- Leary, J. J., D. J. Brigati and D. C. Ward (1983). "Rapid and sensitive colorimetric method for visualizing biotin-labeled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: Bio-blots." *Proceedings of the National Academy of Sciences* **80**(13): 4045-4049.
- Lidmar, J., L. Mirny and D. R. Nelson (2003). "Virus shapes and buckling transitions in spherical shells." *Physical Review E* **68**(5): 051910.
- Lim, F., T. P. Downey and D. S. Peabody (2001). "Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7." *Journal of Biological Chemistry* **276**(25): 22507-22513.
- Liu, L., X. Li, J. Ma, Z. Li, L. You, J. Wang, M. Wang, X. Zhang and Y. Wang (2017). "The molecular architecture for RNA-guided RNA cleavage by Cas13a." *Cell* **170**(4): 714-726. e710.
- Long, Y., B. Bolanos, L. Gong, W. Liu, K. J. Goodrich, X. Yang, S. Chen, A. R. Gooding, K. A. Maegley and K. S. Gajiwala (2017). "Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2." *Elife* **6**: e31558.
- Lu, G., S. J. Dolgner and T. M. T. Hall (2009). "Understanding and engineering RNA sequence specificity of PUF proteins." *Current opinion in structural biology* **19**(1): 110-115.
- Luo, M., C. M. Pang, A. E. Gerken and T. G. Brock (2004). "Multiple nuclear localization sequences allow modulation of 5-lipoxygenase nuclear import." *Traffic* **5**(11): 847-854.
- Mackereth, C. D. and M. Sattler (2012). "Dynamics in multi-domain protein recognition of RNA." *Current opinion in structural biology* **22**(3): 287-296.
- Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." *Science* **339**(6121): 823-826.
- Marcia, M. and A. M. Pyle (2012). "Visualizing group II intron catalysis through the stages of splicing." *Cell* **151**(3): 497-507.
- Margueron, R. and D. Reinberg (2011). "The Polycomb complex PRC2 and its mark in life." *Nature* **469**(7330): 343.
- Maris, C., C. Dominguez and F. H. T. Allain (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." *The FEBS journal* **272**(9): 2118-2131.
- Marshall, K. A. and A. D. Ellington (2000). "[14] In vitro selection of RNA aptamers."

- Maslah, G., P. Barraud and F. H.-T. Allain (2013). "RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence." *Cellular and Molecular Life Sciences* **70**(11): 1875-1895.
- Masuda, S., R. Das, H. Cheng, E. Hurt, N. Dorman and R. Reed (2005). "Recruitment of the human TREX complex to mRNA during splicing." *Genes & development* **19**(13): 1512-1517.
- Masutomi, K., Y. Y. Evan, S. Khurts, I. Ben-Porath, J. L. Currier, G. B. Metz, M. W. Brooks, S. Kaneko, S. Murakami and J. A. DeCaprio (2003). "Telomerase maintains telomere structure in normal human cells." *Cell* **114**(2): 241-253.
- McClure, W. R. (1985). "Mechanism and control of transcription initiation in prokaryotes." *Annual review of biochemistry* **54**(1): 171-204.
- Melnikov, S., A. Ben-Shem, N. G. De Loubresse, L. Jenner, G. Yusupova and M. Yusupov (2012). "One core, two shells: bacterial and eukaryotic ribosomes." *Nature structural & molecular biology* **19**(6): 560.
- Meyer, M., H. Nielsen, V. Oliéric, P. Roblin, S. D. Johansen, E. Westhof and B. Masquida (2014). "Speciation of a group I intron into a lariat capping ribozyme." *Proceedings of the National Academy of Sciences* **111**(21): 7659-7664.
- Miyagishi, M. and K. Taira (2002). "U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells." *Nature biotechnology* **20**(5): 497.
- Moradpour, D., F. Penin and C. M. Rice (2007). "Replication of hepatitis C virus." *Nature reviews microbiology* **5**(6): 453.
- Morris, G. E. (2008). "The cajal body." *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1783**(11): 2108-2115.
- Moy, T. I. and P. A. Silver (2002). "Requirements for the nuclear export of the small ribosomal subunit." *Journal of Cell Science* **115**(14): 2985-2995.
- Mroczek, S. and J. Kufel (2008). "Apoptotic signals induce specific degradation of ribosomal RNA in yeast." *Nucleic acids research* **36**(9): 2874-2888.
- Mulholland, C. B., M. Smets, E. Schmidtman, S. Leidescher, Y. Markaki, M. Hofweber, W. Qin, M. Manzo, E. Kremmer and K. Thanisch (2015). "A modular open platform for systematic functional studies under physiological conditions." *Nucleic acids research* **43**(17): e112-e112.

- Murata, A., S.-i. Sato, Y. Kawazoe and M. Uesugi (2011). "Small-molecule fluorescent probes for specific RNA targets." *Chemical Communications* **47**(16): 4712-4714.
- Nagai, K., C. Oubridge, T. H. Jessen, J. Li and P. R. Evans (1990). "Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A." *Nature* **348**(6301): 515.
- Nath, J. and K. L. Johnson (2000). "A review of fluorescence in situ hybridization (FISH): current status and future prospects." *Biotechnic & Histochemistry* **75**(2): 54-78.
- Nevins, J. R. (1983). "The pathway of eukaryotic mRNA formation." *Annual review of biochemistry* **52**(1): 441-466.
- Nevo-Dinur, K., S. Govindarajan and O. Amster-Choder (2012). "Subcellular localization of RNA and proteins in prokaryotes." *Trends in Genetics* **28**(7): 314-322.
- Niranjanakumari, S., T. Stams, S. M. Crary, D. W. Christianson and C. A. Fierke (1998). "Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA." *Proceedings of the National Academy of Sciences* **95**(26): 15212-15217.
- O'Connell, M. R. (2019). "Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems." *Journal of molecular biology* **431**(1): 66-87.
- Ogg, S. C. and A. I. Lamond (2002). "Cajal bodies and coilin—moving towards function." *The Journal of cell biology* **159**(1): 17-21.
- Padeken, J., P. Zeller and S. M. Gasser (2015). "Repeat DNA in genome organization and stability." *Current opinion in genetics & development* **31**: 12-19.
- Paige, J. S., K. Y. Wu and S. R. Jaffrey (2011). "RNA mimics of green fluorescent protein." *Science* **333**(6042): 642-646.
- Pascarella, S. and F. Negro (2011). "Hepatitis D virus: an update." *Liver International* **31**(1): 7-21.
- Portela, A. n. and P. Digard (2002). "The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication." *Journal of general virology* **83**(4): 723-734.
- Pratt, A. J. and I. J. MacRae (2009). "The RNA-induced silencing complex: a versatile gene-silencing machine." *Journal of Biological Chemistry* **284**(27): 17897-17901.
- Putz, U., P. Skehel and D. Kuhl (1996). "A tri-hybrid system for the analysis and detection of RNA-protein interactions." *Nucleic acids research* **24**(23): 4838-4840.

- Pyle, A. M., F. L. Murphy and T. R. Cech (1992). "RNA substrate binding site in the catalytic core of the Tetrahymena ribozyme." *Nature* **358**(6382): 123.
- Ramakrishnan, V. (2002). "Ribosome structure and the mechanism of translation." *Cell* **108**(4): 557-572.
- Ramos, A., S. Grünert, J. Adams, D. R. Micklem, M. R. Proctor, S. Freund, M. Bycroft, D. St Johnston and G. Varani (2000). "RNA recognition by a Staufien double-stranded RNA-binding domain." *The EMBO journal* **19**(5): 997-1009.
- Rauhut, R. and G. Klug (1999). "mRNA degradation in bacteria." *FEMS microbiology reviews* **23**(3): 353-370.
- Reed, R. and H. Cheng (2005). "TREX, SR proteins and export of mRNA." *Current opinion in cell biology* **17**(3): 269-273.
- Reits, E. A. J. and J. J. Neefjes (2001). "From fixed to FRAP: measuring protein mobility and activity in living cells." *Nature Cell Biology* **3**(6): E145-E147.
- Rinn, J. L. and H. Y. Chang (2012). "Genome regulation by long noncoding RNAs." *Annual review of biochemistry* **81**: 145-166.
- Robinett, C. C., A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray and A. S. Belmont (1996). "In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition." *The Journal of cell biology* **135**(6): 1685-1700.
- Ross, A. F., Y. Oleynikov, E. H. Kislaukis, K. L. Taneja and R. H. Singer (1997). "Characterization of a beta-actin mRNA zipcode-binding protein." *Molecular and cellular biology* **17**(4): 2158-2165.
- Rougemaille, M., T. Villa, R. K. Gudipati and D. Libri (2008). "mRNA journey to the cytoplasm: attire required." *Biology of the Cell* **100**(6): 327-342.
- Sachs, A. B. (1993). "Messenger RNA degradation in eukaryotes." *Cell* **74**(3): 413-421.
- Schmidt, E. V., G. Christoph, R. Zeller and P. Leder (1990). "The cytomegalovirus enhancer: a pan-active control element in transgenic mice." *Molecular and Cellular Biology* **10**(8): 4406-4411.

- Schwede, A., L. Ellis, J. Luther, M. Carrington, G. Stoecklin and C. Clayton (2008). "A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells." *Nucleic acids research* **36**(10): 3374-3388.
- Sentenac, A. (1985). "Eukaryotic RNA polymerase." *Critical Reviews in Biochemistry* **18**(1): 31-90.
- Sharp, P. A. and C. B. Burge (1997). "Classification of introns: U2-type or U12-type." *Cell* **91**(7): 875-879.
- Shen, L., Z. Cai and I. Tinoco Jr (1995). "RNA structure at high resolution." *The FASEB journal* **9**(11): 1023-1033.
- Shiman, R. and D. E. Draper (2000). "Stabilization of RNA tertiary structure by monovalent cations." *Journal of molecular biology* **302**(1): 79-91.
- Silva, I. J., M. Saramago, C. Dressaire, S. Domingues, S. C. Viegas and C. M. Arraiano (2011). "Importance and key events of prokaryotic RNA decay: the ultimate fate of an RNA molecule." *Wiley Interdisciplinary Reviews: RNA* **2**(6): 818-836.
- Simmonds, A. J., I. Livne-Bar and H. M. Krause (2001). "Apical localization of wingless transcripts is required for wingless signaling." *Cell* **105**(2): 197-207.
- Singer, R. H. (1993). "RNA zipcodes for cytoplasmic addresses." *Current Biology* **3**(10): 719-721.
- Soto-Rifo, R. and T. Ohlmann (2013). "The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism." *Wiley Interdisciplinary Reviews: RNA* **4**(4): 369-385.
- Sparano, B. A. and K. Koide (2005). "A strategy for the development of small-molecule-based sensors that strongly fluoresce when bound to a specific RNA." *Journal of the American Chemical Society* **127**(43): 14954-14955.
- Srivastava, A. K. and D. Schlessinger (1990). "Mechanism and regulation of bacterial ribosomal RNA processing." *Annual review of microbiology* **44**(1): 105-129.
- St Johnston, D. (1995). "The intracellular localization of messenger RNAs." *Cell* **81**(2): 161-170.
- Stewart, M. (2010). "Nuclear export of mRNA." *Trends in biochemical sciences* **35**(11): 609-617.

- Stoltenburg, R., C. Reinemann and B. Strehlitz (2007). "SELEX—a (r) evolutionary method to generate high-affinity nucleic acid ligands." *Biomolecular engineering* **24**(4): 381-403.
- Stuurman, N., S. Heins and U. Aebi (1998). "Nuclear lamins: their structure, assembly, and interactions." *Journal of structural biology* **122**(1-2): 42-66.
- Sunbul, M. and A. Jäschke (2013). "Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer." *Angewandte Chemie International Edition* **52**(50): 13401-13404.
- Sunbul, M. and A. Jäschke (2018). "SRB-2: a promiscuous rainbow aptamer for live-cell RNA imaging." *Nucleic acids research* **46**(18): e110-e110.
- Swiger, R. R. and J. D. Tucker (1996). "Fluorescence in situ hybridization: a brief review." *Environmental and molecular mutagenesis* **27**(4): 245-254.
- Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida and H. Okayama (1991). "Structure and organization of the hepatitis C virus genome isolated from human carriers." *Journal of virology* **65**(3): 1105-1113.
- Tambe, A., A. East-Seletsky, G. J. Knott, J. A. Doudna and M. R. O'Connell (2018). "RNA binding and HEPN-nuclease activation Are decoupled in CRISPR-Cas13a." *Cell reports* **24**(4): 1025-1036.
- Terns, M. P. (2018). "CRISPR-based technologies: impact of RNA-targeting systems." *Molecular cell* **72**(3): 404-412.
- Theunissen, O., F. Rudt, U. Guddat, H. Mentzel and T. Pieler (1992). "RNA and DNA binding zinc fingers in *Xenopus* TFIIIA." *Cell* **71**(4): 679-690.
- Thompson, D. M., C. Lu, P. J. Green and R. Parker (2008). "tRNA cleavage is a conserved response to oxidative stress in eukaryotes." *Rna* **14**(10): 2095-2103.
- Tsai, M.-C., O. Manor, Y. Wan, N. Mosammamparast, J. K. Wang, F. Lan, Y. Shi, E. Segal and H. Y. Chang (2010). "Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes." *Science* **329**(5992): 689-693.
- Tsukamoto, T., N. Hashiguchi, S. M. Janicki, T. Tumber, A. S. Belmont and D. L. Spector (2000). "Visualization of gene activity in living cells." *Nature cell biology* **2**(12): 871.

- Tyagi, S. (2009). "Imaging intracellular RNA distribution and dynamics in living cells." *natuRe methods* **6**(5): 331.
- Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields and J. M. Rothberg (2000). "A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*." *Nature* **403**(6770): 623-627.
- Ule, J., K. Jensen, A. Mele and R. B. Darnell (2005). "CLIP: A method for identifying protein–RNA interaction sites in living cells." *Methods* **37**(4): 376-386.
- Valegård, K., L. Liljas, K. Fridborg and T. Unge (1990). "The three-dimensional structure of the bacterial virus MS2." *Nature* **345**(6270): 36.
- Valencia-Burton, M., R. M. McCullough, C. R. Cantor and N. E. Broude (2007). "RNA visualization in live bacterial cells using fluorescent protein complementation." *Nature methods* **4**(5): 421.
- Venables, J. P., M. Ruggiu and H. J. Cooke (2001). "The RNA-binding specificity of the mouse Dazl protein." *Nucleic acids research* **29**(12): 2479-2483.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans and R. A. Holt (2001). "The sequence of the human genome." *science* **291**(5507): 1304-1351.
- Vernizzi, G. and M. O. de la Cruz (2007). "Faceting ionic shells into icosahedra via electrostatics." *Proceedings of the National Academy of Sciences* **104**(47): 18382-18386.
- Vicens, Q., E. a. Mondragón, F. E. Reyes, P. Coish, P. Aristoff, J. Berman, H. Kaur, K. W. Kells, P. Wickens and J. Wilson (2018). "Structure–activity relationship of flavin analogues that target the flavin mononucleotide riboswitch." *ACS chemical biology* **13**(10): 2908-2919.
- Vignali, M., A. H. Hassan, K. E. Neely and J. L. Workman (2000). "ATP-dependent chromatin-remodeling complexes." *Molecular and cellular biology* **20**(6): 1899-1910.
- Wang, S., J.-H. Su, F. Zhang and X. Zhuang (2016). "An RNA-aptamer-based two-color CRISPR labeling system." *Scientific reports* **6**: 26857.
- Wang, X., J. McLachlan, P. D. Zamore and T. M. T. Hall (2002). "Modular recognition of RNA by a human pumilio-homology domain." *Cell* **110**(4): 501-512.

- Warner, K. D., M. C. Chen, W. Song, R. L. Strack, A. Thorn, S. R. Jaffrey and A. R. Ferré-D'Amaré (2014). "Structural basis for activity of highly efficient RNA mimics of green fluorescent protein." *Nature structural & molecular biology* **21**(8): 658.
- West, S., N. J. Proudfoot and M. J. Dye (2008). "Molecular dissection of mammalian RNA polymerase II transcriptional termination." *Molecular cell* **29**(5): 600-610.
- White, R. J. (2004). "RNA polymerase III transcription and cancer." *Oncogene* **23**(18): 3208.
- Wickens, M., D. S. Bernstein, J. Kimble and R. Parker (2002). "A PUF family portrait: 3' UTR regulation as a way of life." *TRENDS in Genetics* **18**(3): 150-157.
- Wilkie, G. S., K. S. Dickson and N. K. Gray (2003). "Regulation of mRNA translation by 5'-and 3'-UTR-binding factors." *Trends in biochemical sciences* **28**(4): 182-188.
- Will, C. L. and R. Lührmann (2011). "Spliceosome structure and function." *Cold Spring Harbor perspectives in biology* **3**(7): a003707.
- Williamson, I., S. Berlivet, R. Eskeland, S. Boyle, R. S. Illingworth, D. Paquette, J. Dostie and W. A. Bickmore (2014). "Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization." *Genes & development* **28**(24): 2778-2791.
- Winkler, W. C., S. Cohen-Chalamish and R. R. Breaker (2002). "An mRNA structure that controls gene expression by binding FMN." *Proceedings of the National Academy of Sciences* **99**(25): 15908-15913.
- Wittmann, H. (1983). "Architecture of prokaryotic ribosomes." *Annual review of biochemistry* **52**(1): 35-65.
- Wood, T. I., D. P. Barondeau, C. Hitomi, C. J. Kassmann, J. A. Tainer and E. D. Getzoff (2005). "Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis." *Biochemistry* **44**(49): 16211-16220.
- Woodcock, C. and R. Horowitz (1995). "Chromatin organization re-viewed." *Trends in cell biology* **5**(7): 272-277.
- Wu, L., P. Murat, D. Matak-Vinkovic, A. Murrell and S. Balasubramanian (2013). "Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins." *Biochemistry* **52**(52): 9519-9527.

- Yakovchuk, P., E. Protozanova and M. D. Frank-Kamenetskii (2006). "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix." *Nucleic acids research* **34**(2): 564-574.
- Yang, L. and L.-L. Chen (2017). "Enhancing the RNA engineering toolkit." *Science* **358**(6366): 996-997.
- Yoon, J.-H., S. Srikantan and M. Gorospe (2012). "MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs." *Methods* **58**(2): 81-87.
- Yu, X. and Z. Li (2015). "Long non-coding RNA HOTAIR: A novel oncogene." *Molecular medicine reports* **12**(4): 5611-5618.
- Zahler, A. M., W. S. Lane, J. A. Stolk and M. B. Roth (1992). "SR proteins: a conserved family of pre-mRNA splicing factors." *Genes & development* **6**(5): 837-847.
- Zhang, H., T. Eom, Y. Oleynikov, S. M. Shenoy, D. Liebelt, J. Dichtenberg, R. H. Singer and G. Bassell (2001). "Neurotrophin-induced transport of a  $\beta$ -actin mRNP complex increases  $\beta$ -actin levels and stimulates growth cone motility." *Neuron* **31**(2): 261-275.
- Zhang, J., K. L. McCann, C. Qiu, L. E. Gonzalez, S. J. Baserga and T. M. T. Hall (2016). "Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA." *Nature communications* **7**: 13085.
- Zhao, Y.-Y., M.-W. Mao, W.-J. Zhang, J. Wang, H.-T. Li, Y. Yang, Z. Wang and J.-W. Wu (2018). "Expanding RNA binding specificity and affinity of engineered PUF domains." *Nucleic acids research* **46**(9): 4771-4782.
- Zheng, B., Z. Wang, S. Li, B. Yu, J.-Y. Liu and X. Chen (2009). "Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis." *Genes & development* **23**(24): 2850-2860.
- Zimmer, M. (2002). "Green fluorescent protein (GFP): applications, structure, and related photophysical behavior." *Chemical reviews* **102**(3): 759-782.
- Zolghadr, K., O. Mortusewicz, U. Rothbauer, R. Kleinhans, H. Goehler, E. E. Wanker, M. C. Cardoso and H. Leonhardt (2008). "A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells." *Molecular & Cellular Proteomics* **7**(11): 2279-2287.

## 5.4 Statutory Declaration and Statement

### Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde.

München, den 25.11.2019

---

Ningjun Duan

### Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig ohne Erfolg einer Doktorprüfung unterzogen habe.

München, den 25.11.2019

---

Ningjun Duan



## 5.5 Acknowledgements

First of all, I would like to thank my supervisor, Prof. Dr. Heinrich Leonhardt, for not only the general opportunity that allows me to study and work in this group but also the patient supervising and instructive discussion that was so helpful in my research. Furthermore, I also learned a lot in the field besides scientific research from his wise and independent opinions, which helps me to overcome the difficulties from past to future.

To Dr. Weihua Qin and Dr. Wen Den, for your kindly help with my projects on both experiments and manuscripts during the four years. With your help, I am able to go on with my projects. And I also thank you very much for the help in my life in Munich.

Thanks to the CSC, which provides me the financial support that allows me to focus on my study and research in these four years.

I would like to thank the members of our team, both the former and the present. To Tobias, Jack and Joel, for your kind and helpful discussions about the projects, to Irina, Hartmann and David, for your help with the microscopy and the computer, to Jeanette and Susanne, for your general service for the whole team, and to all the members of our team, you built a wonderful work and life atmosphere in the lab and helped me a lot during these days.

Many thanks to my friends in both China and Europe. In these years, I shared happiness with you, and you gave me so much support when I am frustrated and confused. Special thanks to Dr. Haitao Wu, for not only the 14 years friendship but also the talks that may change my future.

And finally, I would like to express my deep thanks to my family. Although we are far from each other, you always stand with me. To my parents, who support me to find my own way of life and to make my own decisions, for always helping me with practical and emotional issues. To my grandparents, uncles, aunts, cousins, brothers, and sisters, for your enthusiastic care on my life and carrier.

To everyone who gives me help, I love you all!