d'Collection

이학박사학위논문

# Deadenylase들과 poly(A) barricade 에 의한 poly(A) 길이 조절에 관한 연구

Poly(A) length regulation: deadenylases and the poly(A) barricade

2019년 2월

서울대학교 대학원

생명과학부

박좋아

# Abstract

## Poly(A) length regulation: deadenylases and the poly(A) barricade

Joha Park

School of Biological Sciences

The Graduate School

Seoul National University

Polyadenylation takes place at the 3′ ends of most eukaryotic messenger RNAs (mRNAs). The resulting poly(A) tail tightly associates with poly(A)-binding proteins, and serves as an essential component for mRNA function; it prevents premature decay and promotes translation of the host mRNA. To degrade mRNAs, poly(A) tails should be removed first by a process called deadenylation. Deadenylation is often a rate-limiting step in canonical mRNA degradation pathway, thus the tail provides a central hub for many post-transcriptional regulatory mechanisms including microRNA-mediated gene silencing, AU-rich element-mediated mRNA destabilisation, cytoplasmic polyadenylation during the maternal-to-zygotic transition.

Despite its importance in gene expression regulation, the biology of deadenylation remains largely unexplored, mainly due to technical difficulties in profiling poly(A) sequences, particularly in a genome-wide scale. Here, combined with RNA interference, I apply the recently developed genome-wide poly(A) length measurement techniques to dissect the role and specificity of human deadenylases. Firstly, I confirm that the widely accepted concept of 'biphasic deadenylation' holds true for most mRNAs; the PAN2-PAN3 complex (PAN2/3) comes to trim very long poly(A)s, and the CCR4-NOT complex (CNOT) takes over to complete deadenylation. Notably, however, PAN2/3 trimming was not an indispensable prerequisite for CNOT deadenylation, as the depletion of PAN2/3 had only a minimal impact on the transcriptome, implying that CNOT can largely compensate for PAN2/3 function. By statistical analysis, I show that most mRNA tails are elongated upon the CNOT depletion, and the magnitude of elongation positively correlates with the cytosolic localisation of mRNAs. Altogether, I establish CNOT as a predominant and non-specific cytosolic deadenylase.

Next, I further investigate the poly(A) length distribution at the steady state. I find that genes can be grouped into several clusters by the poly(A) length distribution at equilibrium. The steady-state poly(A) length is strongly associated with mRNA features such as abundance, stability, 3′ UTR length, nuclear enrichment, and translation. Counter-intuitively, stable mRNAs have short poly(A) tails, implying that deadenylation of such mRNAs is uncoupled from the following degradation process especially for those mRNAs. To resolve this paradoxical relationship between mRNA stability and poly(A) length, I come up with an idea of the poly(A) barricade which impedes complete deadenylation, leaving

stable but short-tailed mRNAs.

To discover the poly(A) barricade, I started with identifying the PABPC1 interacting partners with an assumption that PABPC1 constitutes a structural basis of the barricade. By employing liquid chromatography and tandem mass spectrometry for the PABPC1 co-immunoprecipitated (co-IPed) proteins, PABPC1 interacting partners were revealed. Among the PABPC1 interactants, the La-related protein family members, LARP1 and LARP4/4B, emerged as the most prominent candidates for the barricade components. Global poly(A) profiling after RNA co-IP or knockdown of each candidate demonstrate that LARP1 binds and protects ~30–60 nt poly(A)s of most mRNAs, sculpting a periodic pattern that is reminiscent of PABP footprints. Moreover, LARP1 shows strong preferential binding to the mRNAs that have 5′ terminal oligopyrimidine motif which is shared by most ribosomal protein-coding mRNAs, suggesting the higher stability but the shorter poly(A) tail of those house-keeping mRNAs can be partly explained by LARP1 binding. Furthermore, I confirm the poly(A)-binding activity and the function of LARP1 in inhibiting deadenylation by a series of *in vitro* experiments, establishing LARP1 as a core element that constitutes the poly(A) barricade. On the other hand, LARP4/4B non-specifically associates with longer poly(A)s ranged ~70–190 nt long. This implies that LARP1 and LARP4/4B may act on the mRNAs in different stages of their metabolism.

In summary, in this dissertation, I explore the poly(A) length regulation by the transcriptomic analysis on poly(A) tails. This study challenges and revises the biphasic deadenylation model and confirms CNOT as the major cytosolic deadenylase complex that is responsible for non-

specific bulk deadenylation. In addition, I hypothesise the existence of the poly(A) barricade that can resolve the paradoxical phenomenon, that is highly expressed stable mRNAs have short poly(A) tails, by unveiling LARP1 as a position-specific deadenylase inhibitor on the poly(A) that uncouples deadenylation from the following decay. The discovery of the poly(A) barricade expands the players in poly(A) length regulation, which broadens our understanding about the fundamentals of gene expression regulation.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

BPA         bulk poly(A) assay

CNOT        the CCR4-NOT complex

co-IP       co-immunoprecipitation

EMSA        electrophoretic mobility shift assay

Hire-PAT    High resolution poly(A) tail assay

hLa         human La protein

IVD         *In vitro* deadenylation

KD          knockdown

LaM         La motif

LC-MS/MS    liquid chromatography and tandem mass spectrometry

mRNA        messenger RNA

nt          nucleotides

PAN2/3      the PAN2-PAN3 complex

qRT-PCR     quantitative real-time polymerase chain reaction

RBP         RNA-binding protein

RNA         ribonucleic acid

RNA IP      RNA immuprecipitation

| | |
|---|---|
| RPM | read-per-million |
| RRM | RNA recognition motif |
| siRNA | small interfering RNA |
| TE | translation efficiency |
| TOP | terminal oligopyrimidine |
| TPM | transcripts-per-million |
| UTR | untranslated region |

# 1. Introduction

Messenger RNAs (mRNAs) take a fundamental part of the central dogma of molecular biology (Crick, 1970). They serve as a template for materialising the information of life; mRNAs convey the inherited messages encoded in DNA to provide formulae for producing the proteins. In all forms of life, proteins are manufactured only by decoding the information that mRNAs carry. Thus, gene expression regulation means manipulation of the mRNA controllers. Humans have diverse and sophisticated molecular mechanisms for controlling mRNA production, decoding, and degradation to efficiently modulate the gene expression. In this dissertation, I clarify and expand the knowledge on the molecular machinery that determines the mRNA fate at the post-transcriptional level of gene expression regulation.

## 1.1 The mRNA life cycle and post-transcriptional regulation

In human, mRNAs are synthesised by RNA polymerase II (RNAP II)-mediated transcription in the nucleus. For most mRNAs, except for those that encode replication-dependent histones, newly transcribed premature transcripts undergo a series of maturation processes including 5′ capping, splicing, and 3′ cleavage followed by polyadenylation. The mature mRNAs are exported to the cytoplasm and translated to yield proteins. After a few rounds of translation, mRNAs are degraded by various ribonucle-

ases (Figure 1.1)
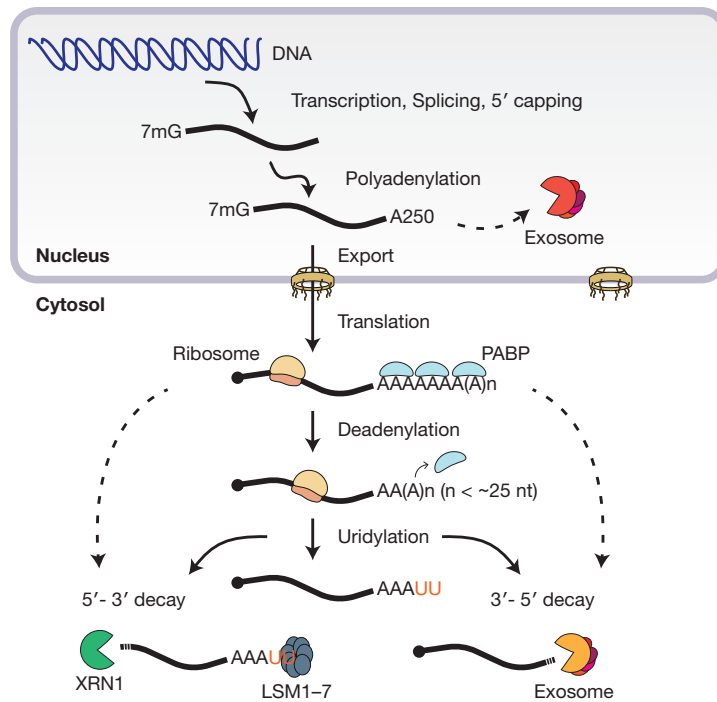


**Figure 1.1** A schematic of the mRNA life cycle in human.

During the life cycle, mRNAs are regulated by multiple layers of post-transcriptional regulation. Each step of the basic mRNA maturation processes such as splicing, export, and translation is equipped with the surveillance programmes for degrading malformed mRNAs. Also, diverse *cis-* and *trans-*acting elements and their associated effector proteins and non-coding RNAs collectively modulate mRNA stability. Consequently, mRNA half-lives greatly vary from just a few minutes to days depending on the regulatory pathways in which the mRNAs are engaged, which demonstrates the significance of the post-transcriptional regulations in determining the fate of mRNAs.

## 1.2 The role of poly(A) tails in mRNA regulation

Since deadenylation is the first and crucial step in mRNA turnover (Decker & Parker, 1993; Goldstrohm & Wickens, 2008), many post-transcriptional regulatory pathways rely on poly(A) tails to modulate mRNA stability. For instance, microRNAs destabilise their target mRNAs by facilitating shortening of the poly(A) tails by recruiting deadenylase complexes (Eulalio et al., 2008). HuR and other AU-rich element binding proteins also destabilise their targets by a similar manner (Chen & Shyu, 2011). Shortened poly(A) tails lose PABPs and become readily accessible to uridylyl transferases such as TUT4 and TUT7. Uridine-marked RNAs are recognised by the LSM complex and degraded by XRN1-mediated 5′-to-3′ decay pathway or RRP44-mediated 3′-to-5′ decay pathway. Additionally, some mRNAs with very long U tails can be directly targeted and decomposed by DIS3L2 (Thomas et al., 2015).

In several biological contexts, poly(A) tails are dynamically regulated. One of the well-known example is the maternal mRNA deposition in oocyte and reactivation in zygote. In this process, maternal mRNAs are deposited to an oocyte in a translationally inactive, short-tailed form. CPEB1 recognises the cytoplasmic polyadenylation element at the 3′ untranslated region (3′ UTR) of the maternal mRNA and nucleates the formation of a poly(A) tail-regulating complex which includes a deadenylase PARN and a terminal nucleotidyl transferase TENT2 (GLD2). The concurrent opposite activities of those catalysts result in net shortening of poly(A) tails, depositing the maternal mRNAs in a short-tailed dormant state. After fertilisation, PARN is expelled from the complex so maternal mRNAs undergo dramatic polyadenylation and start to be translated

3

(Winata & Korzh, 2018). Another example is the cytoplasmic polyadenylation by different terminal nucleotidyl transferases, TENT4A (PAPD7) and TENT4B (PAPD5). Those TENTs usually add adenosines at the 3′ end of RNAs. However, they occasionally add guanosines to further stabilise the tail, thereby shields mRNAs from rapid deadenylation (Lim et al., 2018). Taken together, the poly(A) tail serves as a regulatory platform for gene regulation in a variety of biological contexts.

# 2. Genome-wide examination of the role and specificity of the deadenylases

## 2.1 Background

In human genome, 12 deadenylases have been identified so far, and some of them are deeply conserved from yeast to human (Table 2.1). They catalyse the hydrolysis of adenosine polymers one by one in the 3′-to-5′ direction, which leads to shortening of a poly(A) stretch. Based on the conserved residues in their nuclease domains, deadenylases can be classified into two groups: DEDD nucleases and EEP nucleases. In spite of the pivotal role of poly(A) tail regulation in controlling mRNA stability and translation, the role and specificity of deadenylases remain largely unexplored, particularly in a genome-wide scale, presumably due to the lack of an advanced method for measuring poly(A) length. In this chapter, I investigate the human major deadenylases by applying RNA interference (RNAi) combined with the recently developed genome-wide poly(A) profiling techniques to reveal their poly(A) length preferences and target specificities.

### 2.1.1 The 'biphasic deadenylation' model

There has been a model claiming that deadenylation is composed of two phases which probably be conducted by different poly(A) nucleases (Yamashita et al., 2005). Using the beta-globin pulse expression system com-

**Table 2.1** Human deadenylases and their homologs. Modified from the table demonstrated in (Goldstrohm & Wickens, 2008).

| Family | Yeast | Fly | Mouse | Human |
|---|---|---|---|---|
| DEDD nucleases | | | | |
| PAN2 | Pan2 | PAN2 | PAN2 | PAN2 |
| POP2 | Pop2 | POP2 | CNOT7 | CNOT7 |
| | | | CNOT8 | CNOT8 |
| PARN | – | – | PARN | PARN |
| | | | PARNL | PARNL |
| CAF1Z | – | – | CAF1Z | CAF1Z |
| EEP nucleases | | | | |
| CCR4 | Ccr4 | CCR4 | CCR4 | CNOT6 |
| | | | CCR4L | CNOT6L |
| Nocturnin | – | NOC | NOC | NOC |
| ANGEL | Ngl1,2,3 | Angel | ANGEL1,2 | ANGEL1,2 |
| 2′PDE | 2′PDE | 2′PDE | 2′PDE | 2′PDE |

bined with RNAi against the major deadenylases, Yamashita et al. observed that the PAN2-PAN3 complex (PAN2/3) and the CCR4-NOT complex (CNOT) act on the poly(A) tails of different lengths, and proposed the 'biphasic deadenylation' model. According to this widely accepted deadenylation model, PAN2/3 initiates deadenylation by removing long poly(A) tails to ~110 nt, whereupon CNOT takes over and completes deadenylation. The later-coming CNOT is regarded as a major deadenylase complex as its depletion leads to a severe defect in deadenylation

(Tucker et al., 2001; Temme et al., 2004; Nousch et al., 2013). Of the two CNOT catalytic subunits, Ccr4, but not Caf1/Pop2, depletion drastically affects deadenylation in yeast (Tucker et al., 2001), whereas the opposite is true in animals (Temme et al., 2004; Piao et al., 2010; Nousch et al., 2013). Besides the two major players in the biphasic model of deadenylation, PARN is also implicated in mRNA deadenylation in the maternal mRNA deposition in *Xenopus* oocytes (Körner et al., 1998; Copeland & Wormington, 2001), and a discrete set of mRNAs in mouse myoblasts (Lee et al., 2012).

Our knowledge on deadenylation was primarily inferred from the studies using artificial reporter systems, a few representative genes, or bulk poly(A) measurements, which limits our understanding on how deadenylation process is regulated depending on mRNA species. In addition, genome-wide analyses defining deadenylase targets relied upon indirect RNA quantification methods such as RNA-seq or microarray (Yamashita et al., 2005), (Eulalio et al., 2008), (Aslam et al., 2009), (Mittal et al., 2011), (Lee et al., 2012), (Chen et al., 2017). Direct measurement of poly(A) tail length has been technically challenging and has not previously been applied to deadenylation studies. Thus, many critical questions, that can be addressed only with simultaneous and direct examination of mRNA species and their poly(A) length, remain unanswered despite the central importance of deadenylation in gene regulation.

First of all, it is unclear if the 'biphasic deadenylation' model, which claims that mRNAs undergo two phases of deadenylation by PAN2/3 and CNOT in an ordered manner, generally applies to all mRNAs and to what extent the first phase contributes to poly(A) removal and mRNA

turnover. Secondly, the substrate specificities of the deadenylases need to be elucidated: if the deadenylases target different mRNAs with gene specificity or if they act redundantly on overlapping sets of transcripts. Furthermore, it remains unknown if the two catalytic subunits of the CNOT complex have distinct functions and, if so, how their activities are coordinated. Lastly, the role of cytoplasmic poly(A) binding protein (PABPC) in deadenylation also remains controversial. Both stimulating and suppressive impacts of PABPC on deadenylation have been reported from *in vitro* assay (Tucker et al., 2002; Viswanathan et al., 2004; Simón & Séraphin, 2007; Funakoshi et al., 2007). However, in PABP-depleted cells, poly(A) tail becomes elongated, suggesting PABPC promotes deadenylation (Sachs & Davis, 1989; Yao et al., 2007; Fabian et al., 2009; Fukaya & Tomari, 2011).

## 2.1.2 Methods for profiling intact poly(A) tails

Measuring poly(A) length in a genome-wide scale has been challenging due to the difficulty in sequencing homopolymers in the commonly used sequencing platforms such as Illumina and Ion torrent (Chang et al., 2014). For instance, in the Illumina platform version that uses 3-channel fluorescence detection system, T signals accumulate over sequencing cycles due to the incomplete fluorescence removal reaction. As a result, it fails to accurately measure the number of homopolymeric adenosines because the T signals from poly(T), which are complementary to the poly(A) in the sequencing reaction, still remain even after an A stretch is fully sequenced (Kircher et al., 2009). Recently, however, several techniques have been developed for measuring the global poly(A) length in a large scale.

TAIL-seq and its derivative mTAIL-seq, together with Tailseeker analysis algorithm (Chang et al., 2014; Lim et al., 2016) have enabled accurate global poly(A) profiling by recalculation of the accumulating T signals. It adjusts overwhelming T signals to locate where they actually end in the base calling (Chang et al., 2014). I employed these techniques for measuring poly(A) length in a genome-wide scale to investigate the specificities of human conventional deadenylases.

To profile the intact 3′ ends of RNAs, TAIL-seq uses direct ligation of the 3′ adapter oligos to the rRNA-depleted RNAs to preserve the 3′ termini. The ligated RNAs are then made as a Illumina sequencing library by a series of steps and read in two opposite directions, one is for identifying the RNA species and the other is for measuring poly(A) tail length (Chang et al., 2014). mTAIL-seq is a derivative of TAIL-seq, which focuses on mRNAs. The major difference between mTAIL-seq and TAIL-seq lies in the 3′ adapter ligation step. mTAIL-seq employs 'splint ligation' of a custom 3′ adapter to specifically enrich mRNAs from total RNA (Lim et al., 2016). In this study, all mTAIL-seq datasets were primarily processed by the software Tailseeker and then subjected to the in-depth further analyses.

## 2.2 Genome-wide assessment of the poly(A) profiles

To dissect the role and specificity of the major human deadenylases, mTAIL-seq had been performed combined with RNAi against CNOT, PAN2/3, and PARN to assess the changes in poly(A) tail in a genome-wide scale.[1] Figure 2.1 illustrates the stoichiometry of those deadenylase complexes

---

[1]The experiments in this section were carried out by Hyerim Yi

and their catalytic subunits. Notably, CAF1a/b (CNOT7 and CNOT8), which functions as both a core catalytic subunit and a bridge that links CCR4a/b (CNOT6 and CNOT6L) to the scaffold protein CNOT1, was targeted by the small interfering RNAs (siRNAs) to impair the activities of both catalytic subunits at the same time. For PAN2/3, both PAN2 and PAN3 were depleted for the better knockdown outcome. PARN forms a homodimer so the single targeting was enough to eliminate the function. After depletion of each deadenylase, mTAIL-seq libraries were prepared and analysed to see the poly(A) length changes.
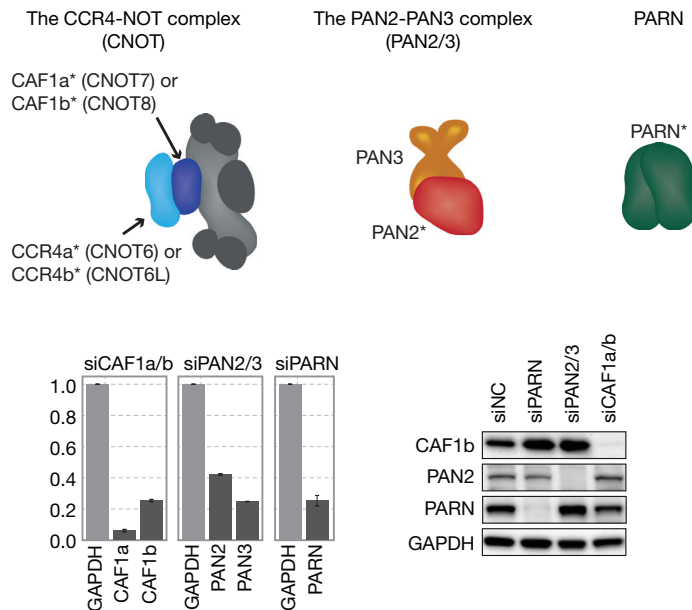


**Figure 2.1** The major human deadenylases examined in this study. The subunits targeted by siRNAs are coloured, and the catalytic subunits are marked by the asterisks. The knockdown efficiencies were confirmed by both at an RNA level (qRT-PCR, lower left) and a protein level (Western blot, lower right).

### 2.2.1 Poly(A) length changes upon knockdown of the major deadenylases

The most important advantage of applying the high-throughput poly(A) profiling method is that it makes possible to trace the poly(A) length changes for individual genes. By applying TAIL-seq, I could analyse the poly(A) length changes at individual gene-level upon knockdown of each deadenylase. The most dramatic changes were observed in the CAF1-depleted sample; most mRNAs gained greatly elongated poly(A) tails. Conversely, disruption of PAN2/3 or PARN had only minor impact on the geometric mean of poly(A) length (Figure 2.2). These results were verified for some selected genes by the Hire-PAT assay (Bazzini et al., 2012) (Figure 2.3, right). Notably, mRNAs localised to the cytoplasm are influenced by CAF1 knockdown more strongly than those that are enriched in the nucleus (Figures 2.3, left and 2.4). Together, these results indicate that the CNOT is responsible for deadenylation of the vast majority, if not all, of cytosolic mRNA population.

### 2.2.2 The 'biphasic deadenylation' model holds true for most mRNAs

Based on the artificial reporter pulse expression system, it is suggested that mRNA deadenylation consists of the two phases, one by PAN2/3 and the following phase by CNOT, which is called the 'biphasic deadenylation' model (Yamashita et al., 2005). However, it remains elusive if the model is applicable to all endogenous mRNAs. To check if the order of action and the length preferences of PAN2/3 and CNOT are valid in general, I examined the poly(A) length changes in detail by looking at the poly(A)
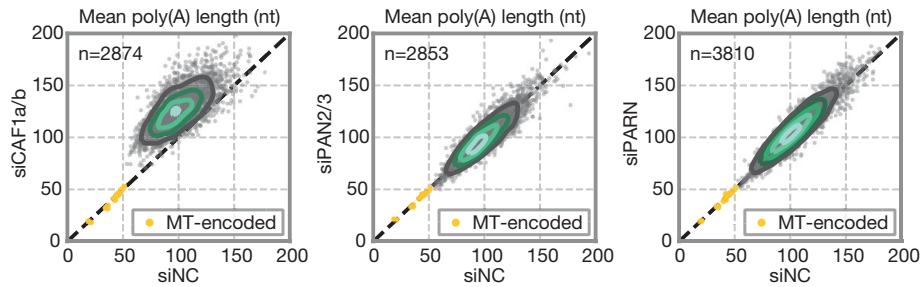
**Figure 2.2** Poly(A) length changes upon knockdown of each deadenylase. The $x$ and $y$ axes indicate the geometric mean of poly(A) length in the indicated sample. The yellow dots represent mitochondrially-encoded genes, and the grey dots represent the others. The green-to-black contours show the bivariate kernel density estimates. n: number of genes.

length distributions at an individual gene level. Indeed, consistently with the model, the results showed the clear size-specific accumulation, and this pattern was consistent among most genes; ~150 nt poly(A) tags accumulated when CNOT was disrupted while only very long poly(A)s over ~200 nt increased when PAN2/3 was depleted (Figure 2.5).

However, the degrees of accumulation were very different in the CAF1- and PAN2/3-depleted samples; upon CNOT loss, most mRNAs gained dramatically elongated tails while the effect of PAN2/3 knockdown was minimal and restricted to the very long poly(A) tails. This implies that CNOT might largely compensate the first phase PAN2/3 deadenylation, which emphasises the importance of the second phase deadenylation by CNOT.
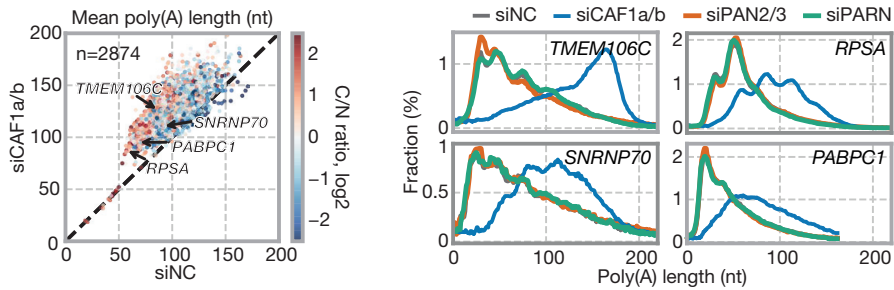
**Figure 2.3** Geometric mean poly(A) length scatter in the CAF1-depleted sample, coloured with the abundance ratio of the cytosol over the nucleus. n: number of genes (left). For the marked genes, individual poly(A) length changes in the control and knockdown samples were verified by the Hire-PAT assay (right).

## 2.3 Statistical analysis reveals the target specificity of the deadenylases

Because the impact of PAN2/3 knockdown on poly(A) was minute, the statistical test on abstract statistics such as mean poly(A) length was inappropriate to reveal the target specificity due to low power. Instead, I employed the one-tailed Mann-Whitney U test, which takes the distributional changes into account, so that it can capture the small, size-specific changes that not enough to alter the mean poly(A) length.

13
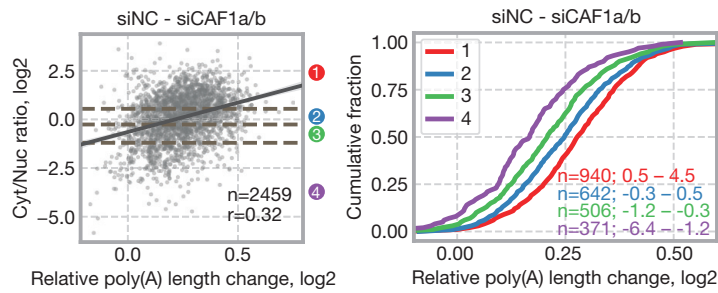
**Figure 2.4** Scatter plot for comparing the cytosol/nucleus abundance ratio with the relative poly(A) length change. The relative poly(A) length change was defined as the mean difference divided by the mean in the control sample. The dashed horizontal lines show quartiles of the cytosol over the nucleus abundance ratio. n: number of genes, r: Pearson's correlation coefficient (left). The numbers in legend represent the quartile gene groups. n: number of genes in the group (right).

## 2.3.1 Mitochondrial RNAs escape from the conventional deadenylases

The subcellular localisation of CNOT and PAN2/3 is mostly cytosolic (Yamashita et al., 2005). Thus, the physical separation of mitochondrial RNAs from the cytosol is expected to prevent the direct targeting by the conventional cytosolic deadenylases. Indeed, in the CAF1- and PAN2/3-depleted cells, the mean poly(A) length of mitochondrial mRNAs did not change (Figure 2.2). For this reason, mitochondrial mRNAs were used as a stable negative control in the following analyses.
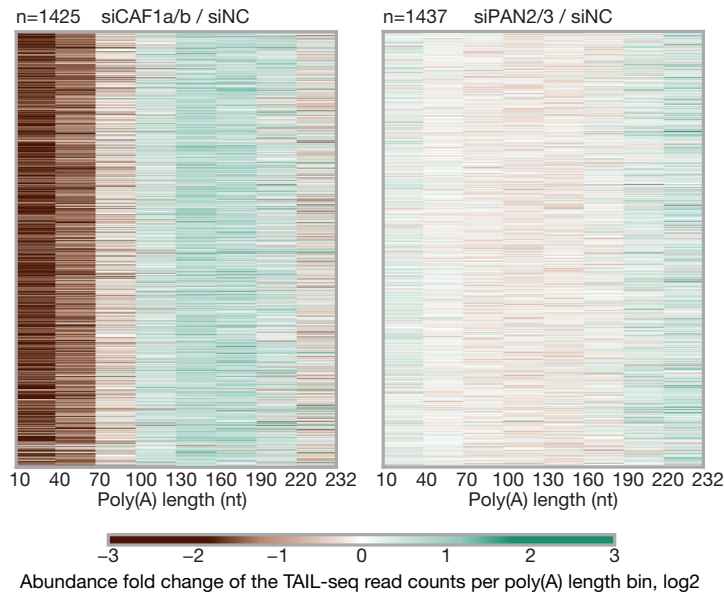
**Figure 2.5** Heat maps showing the changes in the binned poly(A) length distributions of individual genes in CAF1a/b- (left) or PAN2/3- (right) depleted HeLa cells. Each row represents individual gene. The window length of the last bin is shorter than the other bins owing to the limitation of maximum poly(A) length that can be measured by TAIL-seq in the 300-cycle setup.

## 2.3.2 CNOT is a predominant and non-specific ribonuclease

As can be expected from the great increase in the mean poly(A) length (Figure 2.2), most genes were strongly affected by the disruption of CNOT; out of 1,930 genes that were detected by >150 reads, 1,489 genes (76.6%) were called to have significantly elongated poly(A) length distributions under the criteria of the adjusted $p$-value <0.01 and the effect size >0.1 (Figure 2.6). Besides, even though not called as significant, the rest also showed somewhat elongated poly(A) tails, and the accumulating poly(A)

length range was consistent. In contrast, only 3.3% of genes had significantly elongated poly(A) tails under the same criteria in PAN2/3-depleted sample, showing the minor role of PAN2/3 in shaping the poly(A) profile. Altogether, the results suggest that CNOT is a predominant and non-specific poly(A) nuclease complex.
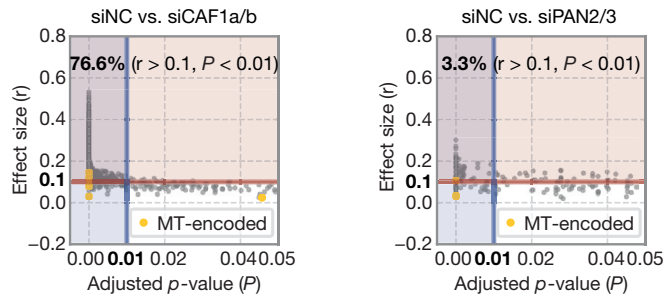


**Figure 2.6** Significantly affected genes that were identified by the one-tailed Mann-Whitney U test, under the Benjamini-Hochberg adjusted $p$-value $<0.01$, and the effect size (r) $>0.1$. Genes that have $>150$ reads with the adjusted $p$-value $<0.05$ were presented in the plots.

To further understand the functional consequences of deadenylase deficiency in mRNA turnover, we measured the steady-state mRNA levels using the mTAIL-seq read counts. To estimate the relative abundance, I normalised the read counts by the geometric mean counts of mitochondrially-encoded genes, with an assumption that mitochondrially-encoded genes are unaffected by the depletion of deadenylases. The results show that mRNA abundance increased globally in CAF1-depleted HeLa cells, whereas PAN2/3- or PARN-depleted cells showed little change, which is consistent with the changes in poly(A) length (Figure 2.7).
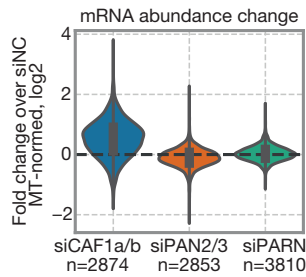
**Figure 2.7** Violin plots showing the tag abundance fold changes upon depletion of each deadenylase complex. The expression level was estimated based on the TAIL-seq read counts normalised by the geometric mean of the mitochondrial RNAs. Genes that have >100 reads were included.

## 2.4 Classification of mRNAs by the steady-state poly(A) profile

Although nearly all mRNAs were heavily affected by the CNOT depletion, some genes more strongly responded while others did not. To quantitatively analyse the degree of response to the CNOT disruption, I calculated the effect size from the U statistics of the one-tailed Mann-Whitney U test (Fritz et al., 2012). By comparing the effect size with other metrics, I found one feature that could substantially explain the difference: the steady-state mean poly(A) length in the control sample (Figure 2.8). The steady-state mean poly(A) length showed strong negative correlation (Pearson's r=-0.53) with the effect size of the siCAF1a/b sample. This is likely because the poly(A) accumulating range in the CNOT disruption is rather fixed to the ~150 nt while the distributions in the steady state are more variable among the mRNA species. In contrast, the negative correlation to the mean poly(A) length does not appear for the effect size of

the siPAN2/3 sample (Pearson's r=0.02).
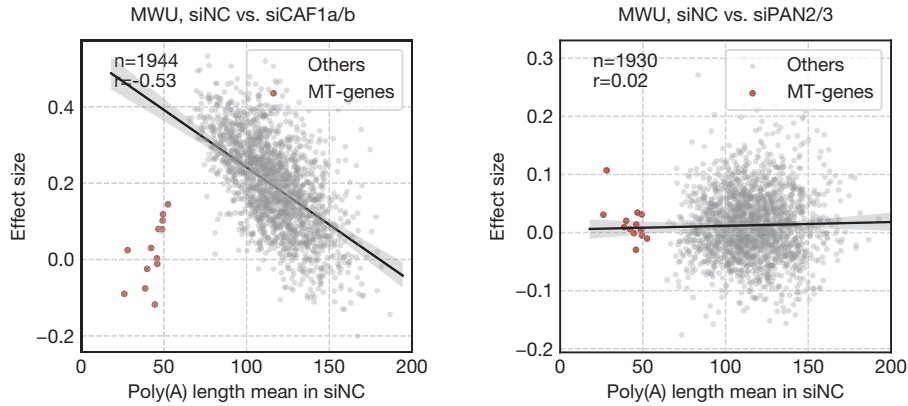


**Figure 2.8** The correlation between the effect size from the one-tailed Mann-Whitney U test and the geometric mean poly(A) length in the control sample. The effect size was calculated from the U statistics as described in (Fritz et al., 2012). MT-genes: Mitochondrially-encoded genes, r: Pearson's correlation coefficient, n: number of genes.

The stationary poly(A) length distribution is determined by the ensemble effects of the post-transcriptional regulations that control speed of deadenylation. If there are higher-order regulatory mechanisms that govern the post-transcriptional regulation for a subset of genes, grouping genes by the steady-state poly(A) profile might reveal the hidden regulatory axes that differentiate mRNAs into several groups otherwise vague by the confounding effect when being looked at in bulk. Intriguingly, through *k*-means clustering of genes by the binned poly(A) length distribution, here I show that genes can be grouped into several clusters that possess distinctive mRNA features.

### 2.4.1 Genes can be clustered by the binned poly(A) length distribution at the steady state

The poly(A) length distribution would differ in shape if mRNAs are regulated by different sets of regulatory pathways. In other words, if a group of genes is under control of the same set of regulatory pathways, the mRNA tails would be sculpted in a similar fashion. With this assumption, I binned the poly(A) length with a window size of 20 nt, and performed the clustering analysis with the *k*-means clustering algorithm. As a result, genes could be segregated into 6 clusters (Figure 2.9).

In the clustering, the most evidently distinct group was C0; this cluster comprises exclusively 14 mitochondrially-encoded genes. The poly(A) tails of mitochondrial RNAs are ~50 nt in length and relatively homogeneous, so they are readily distinguishable from the other cytosolic mRNAs. The other clusters that are worth attention are C1 and C5; C1 contains genes with the shortest poly(A) tails while C5 includes genes with the longest, suggesting that these contrasting clusters might be under the control of the most distinct post-transcriptional regulatory pathways.

### 2.4.2 Clustering is reproducible in other datasets

To check if the clustering is reproducible in other mTAIL-seq runs, I investigated the poly(A) profiles in other datasets. Firstly, the clustering appears to be reproducible in the replicate sample of the first mTAIL-seq library (Figure 2.10).[2] Overall, although the measured lengths were shorter than those from the previous experiment, the tendency of the poly(A) profiles of each cluster was reproduced well; C1 mRNAs have the

---

[2]The replicate mTAIL-seq library was produced by Hyerim Yi
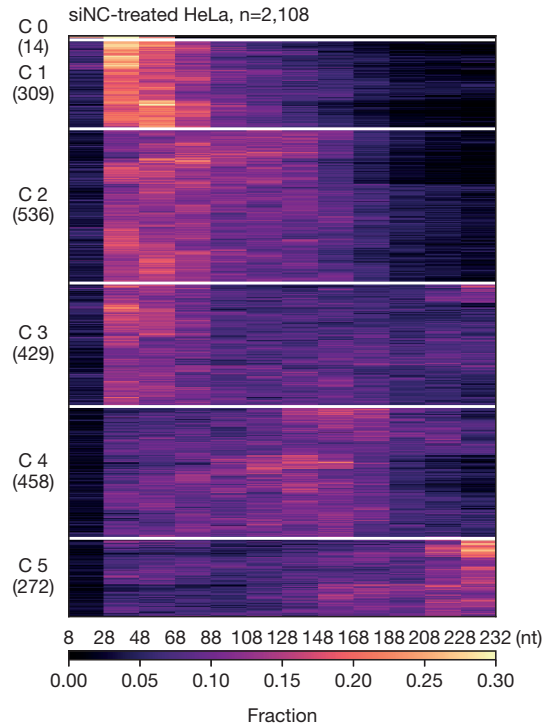
**Figure 2.9** Clustering genes by the binned poly(A) length distribution. *k*-means clustering was performed first and then hierarchical clustering was conducted for each cluster. The negative control siRNA-treated HeLa sample was used for the clustering. Each row represents the binned poly(A) length distribution of a gene. The numbers in parentheses indicate the number of genes in the corresponding cluster. The white horizontal lines partition the clusters. The last bin is smaller than the window size (20 nt) due to the technical upper limit of measuring poly(A) length by mTAIL-seq.

shortest tail while C5 mRNAs carry the longest. Moreover, the clustering was persistent in a different cell line, HCT116 cells[3], implying that the regulatory mechanisms shaping poly(A) tails are in action in other cells as well (Figure 2.11).
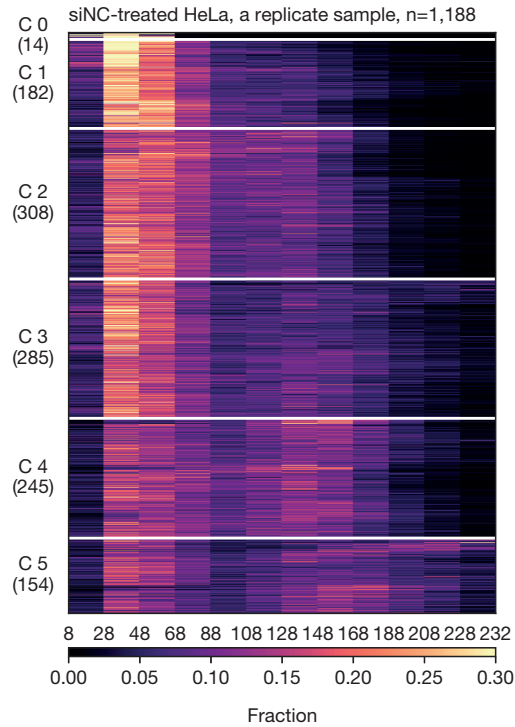


**Figure 2.10** The binned poly(A) length distributions in the HeLa replicate sample. The clustering information was preserved to see the reproducibility of the clustering. Like in Figure 2.9, each row represents the binned poly(A) length distribution of a gene. The numbers in parentheses indicate the number of genes in the corresponding cluster. The white horizontal lines partition the clusters. The last bin is smaller than the window size (20 nt) due to the technical upper limit of measuring poly(A) length by mTAIL-seq.

---

[3]The HCT116 mTAIL-seq library was produced by Hyerim Yi
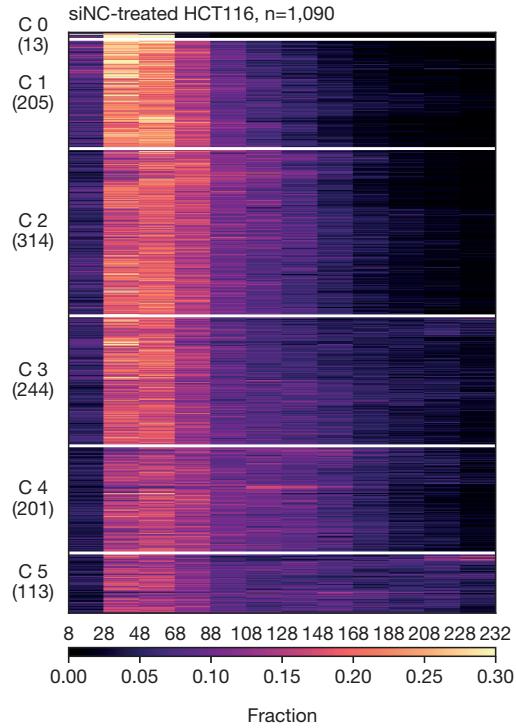
**Figure 2.11** The binned poly(A) length distributions in the siNC-treated HCT116 cells. The clustering information was preserved to see the reproducibility of the clustering. Like in Figure 2.9, each row represents the binned poly(A) length distribution of a gene. The numbers in parentheses indicate the number of genes in the corresponding cluster. The white horizontal lines partition the clusters. The last bin is smaller than the window size (20 nt) due to the technical upper limit of measuring poly(A) length by mTAIL-seq.

22

### 2.4.3 Each mRNA cluster has distinctive features

To see if mRNAs in each cluster have different characteristics, I looked into the several mRNA features: mean poly(A) length, abundance, 3′ UTR length, half-life, nuclear enrichment, and translation efficiency. Strikingly, the tendency of the cluster orders was consistent among the features, implying that the examined features were strongly intertwined. The mRNAs with shorter poly(A) tails tend to 1) be more abundant, 2) have shorter 3′ UTR, 3) be more stable, 4) be more enriched in the cytosol, 5) be translated more efficiently than those with the longer tails. This tendency is the most clearly seen when comparing C1 and C5, which are the most contrasting clusters (Figure 2.12).

Gene Ontology term enrichment analysis using GOATOOLS (Klopfenstein et al., 2018) reveals that the translation-associated terms are significantly enriched in C1 (Table 2.2). Of those translation-related genes, more than half of the ribosomal subunit-constituting genes belong to C1. Interestingly, those genes were reported to be specifically regulated in the context of cell cycle or serum starvation (Park et al., 2016). This special subset shares common characteristics such as short 3′ UTR, optimal codons, and the 5′ terminal oligopyrimidine (TOP) motif. This supports the initial anticipation that grouping genes by the poly(A) length distribution would reveal post-transcriptional regulatory units that differentiate a subset of mRNAs from the rest. On the other hand, no significantly enriched terms were found in other clusters except for C0 which exclusively contains mitochondrially-encoded genes (Table 2.3).
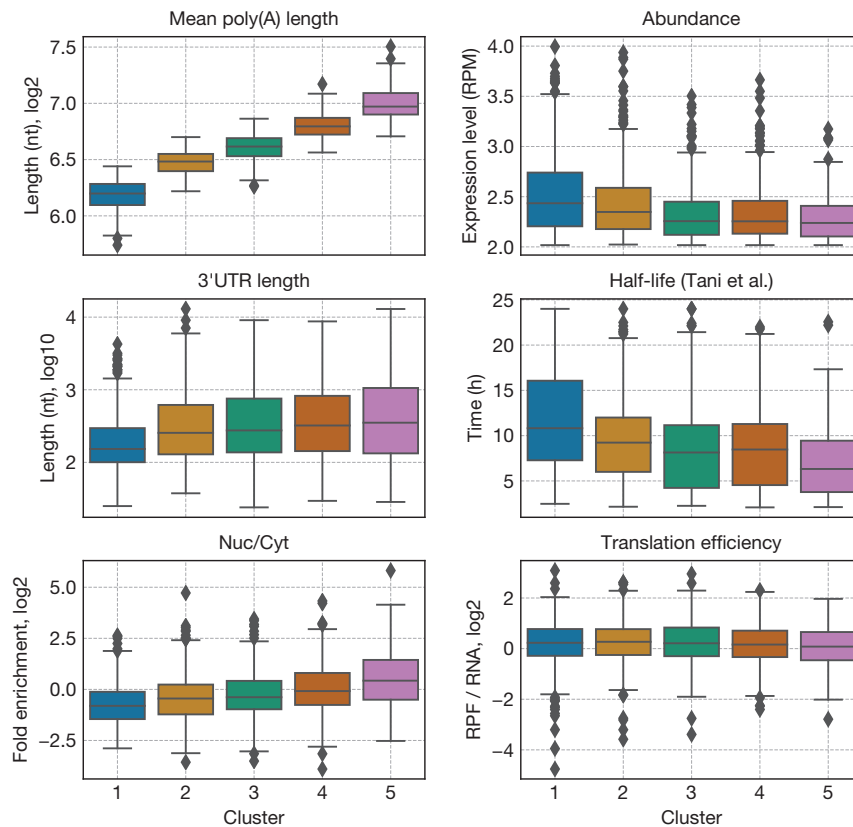
**Figure 2.12** Box plots that show the characteristics of mRNAs in each cluster. Poly(A) length and abundance were estimated from mTAIL-seq reads. 3′ UTR length is the median of those lengths from the multiple isoforms registered in GENCODE v.24. For the half-life, only the genes that have half-lives shorter than 24 hours were selected for higher accuracy (Tani et al., 2012). Translation efficiency (TE) simply means the ribosome density: abundance in RPF tags divided by RNA-seq tags. The TE datasets were adopted from (Park et al., 2016). To calculate the nuclear enrichment, ENCODE subcellular fractionation RNA-seq datasets (ENCSR000CPQ, ENCSR000CPP) were used (ENCODE, 2012)

**Table 2.2** Gene Ontology term enrichment analysis for C1.

| GO ID | NS | Name | Subset | Population | P value | Depth | Count | Adjusted P value |
|---|---|---|---|---|---|---|---|---|
| GO:0006414 | BP | translational elongation | 36/309 | 67/2018 | 1.01e-13 | 6 | 36 | 4.67e-10 |
| GO:0006413 | BP | translational initiation | 42/309 | 88/2018 | 1.64e-13 | 3 | 42 | 4.67e-10 |
| GO:0006415 | BP | translational termination | 35/309 | 65/2018 | 2.11e-13 | 6 | 35 | 4.67e-10 |
| GO:0000184 | BP | nuclear-transcribed mRNA catabolic process, no... | 36/309 | 72/2018 | 1.81e-12 | 10 | 36 | 2.60e-09 |
| GO:0006614 | BP | SRP-dependent cotranslational protein targetin... | 38/309 | 79/2018 | 1.96e-12 | 12 | 38 | 2.60e-09 |
| GO:0019083 | BP | viral transcription | 35/309 | 70/2018 | 3.79e-12 | 5 | 35 | 4.19e-09 |
| GO:0019058 | BP | viral life cycle | 38/309 | 81/2018 | 5.26e-12 | 5 | 38 | 4.98e-09 |
| GO:0016032 | BP | viral process | 63/309 | 201/2018 | 1.44e-09 | 4 | 63 | 9.54e-07 |
| GO:0006412 | BP | translation | 51/309 | 151/2018 | 5.00e-09 | 7 | 51 | 2.76e-06 |
| GO:0010467 | BP | gene expression | 79/309 | 324/2018 | 2.94e-06 | 4 | 79 | 1.50e-03 |
| GO:0044267 | BP | cellular protein metabolic process | 57/309 | 215/2018 | 7.84e-06 | 5 | 57 | 3.72e-03 |
| GO:0005840 | CC | ribosome | 42/309 | 96/2018 | 6.57e-12 | 5 | 42 | 5.45e-09 |
| GO:0022625 | CC | cytosolic large ribosomal subunit | 23/309 | 42/2018 | 2.37e-09 | 6 | 23 | 1.43e-06 |
| GO:0005829 | CC | cytosol | 135/309 | 659/2018 | 1.25e-05 | 4 | 135 | 5.53e-03 |
| GO:0070062 | CC | extracellular exosome | 130/309 | 658/2018 | 1.64e-04 | 5 | 130 | 6.72e-02 |
| GO:0022627 | CC | cytosolic small ribosomal subunit | 12/309 | 26/2018 | 1.72e-04 | 6 | 12 | 6.72e-02 |
| GO:0003735 | MF | structural constituent of ribosome | 44/309 | 110/2018 | 7.92e-11 | 2 | 44 | 5.85e-08 |
| GO:0003723 | MF | RNA binding | 104/309 | 505/2018 | 2.01e-04 | 4 | 104 | 7.43e-02 |

**Table 2.3** Gene Ontology term enrichment analysis for C0.

| GO ID | NS | Name | Subset | Population | P value | Depth | Count | Adjusted P value |
|---|---|---|---|---|---|---|---|---|
| GO:0022904 | BP | respiratory electron transport chain | 11/14 | 77/2018 | 4.03e-14 | 5 | 11 | 2.67e-10 |
| GO:0044237 | BP | cellular metabolic process | 11/14 | 93/2018 | 3.59e-13 | 2 | 11 | 1.19e-09 |
| GO:0006120 | BP | mitochondrial electron transport, NADH to ubiq... | 6/14 | 30/2018 | 1.76e-08 | 6 | 6 | 2.34e-05 |
| GO:0042773 | BP | ATP synthesis coupled electron transport | 3/14 | 3/2018 | 2.66e-07 | 6 | 3 | 2.52e-04 |
| GO:0044281 | BP | small molecule metabolic process | 11/14 | 335/2018 | 5.17e-07 | 2 | 11 | 4.29e-04 |
| GO:0055114 | BP | oxidation-reduction process | 7/14 | 115/2018 | 4.01e-06 | 2 | 7 | 2.66e-03 |
| GO:1902600 | BP | proton transmembrane transport | 4/14 | 38/2018 | 9.36e-05 | 8 | 4 | 4.78e-02 |
| GO:0005743 | CC | mitochondrial inner membrane | 11/14 | 184/2018 | 7.80e-10 | 6 | 11 | 1.73e-06 |
| GO:0005747 | CC | mitochondrial respiratory chain complex I | 6/14 | 33/2018 | 3.26e-08 | 7 | 6 | 3.60e-05 |
| GO:0045277 | CC | respiratory chain complex IV | 3/14 | 4/2018 | 1.06e-06 | 4 | 3 | 7.82e-04 |
| GO:0005739 | CC | mitochondrion | 10/14 | 381/2018 | 2.51e-05 | 5 | 10 | 1.51e-02 |
| GO:0016021 | CC | integral component of membrane | 10/14 | 387/2018 | 2.90e-05 | 3 | 10 | 1.61e-02 |
| GO:0008137 | MF | NADH dehydrogenase (ubiquinone) activity | 6/14 | 30/2018 | 1.76e-08 | 6 | 6 | 2.34e-05 |
| GO:0004129 | MF | cytochrome-c oxidase activity | 3/14 | 17/2018 | 1.71e-04 | 8 | 3 | 8.10e-02 |

## 2.5 Discussion

In this chapter, I elucidate the specificity of human major deadenylase complexes, PAN2/3 and CNOT, by combining RNAi and the genome-wide poly(A) tail length measurement techniques. Both global and individual changes in poly(A) length confirm that the 'biphasic deadenylation' model can be generalised to most endogenous mRNAs; the size-specific accumulation of the poly(A) tags in the loss of each deadenylase indicates that PAN2/3 trims long poly(A)s of >~200 nt whereas CNOT prefers the shorter tails of <~150 nt regardless of mRNA species. Notably, while a loss of PAN2/3 did not cause meaningful alteration in overall poly(A) length, CNOT disruption results in great elongation of most, if not all, mRNAs and also severe cell death, emphasising the crucial and extensive role of CNOT in mRNA turnover.

One important remaining question is how PAN2/3 recognises only very long poly(A) tails. It is unlikely that a protein complex can measure such long stretch of poly(A)s over 200 nt, because a single protein generally spans no more than 50 nt. Therefore, I hypothesise that PAN2/3 may be recruited by an RBP which is found only in the newly synthesised mRNAs, which generally have long tranquil poly(A) tails. It is plausible that PABPN1 might be the one since it associates only with the fresh poly(A) tails and starts to be replaced by cytoplasmic PABPs as mRNAs are exported from the nucleus. Also, it is a non-specific tail binder which fits with the characteristic of PAN2/3 that functions without noticeable target specificity. It would be interesting to test the role of PABPN1 in the first phase deadenylation in future studies.

While investigating on the substrate specificity of CNOT, I found

that the poly(A) length in steady state can largely explain the magnitude of response to the CNOT loss; the effect size of poly(A) elongation negatively correlates with the mean poly(A) length in the negative control. To reveal a clue about what determines the poly(A) length landscape in steady state, I looked into the binned poly(A) length distribution and performed the clustering analysis. Intriguingly, genes can be clustered into the several groups by the steady-state poly(A) profile. Moreover, each cluster has distinctive mRNA features in terms of abundance, stability, 3′ UTR length, nuclear enrichment, and translation, suggesting that the poly(A) length at equilibrium is tightly associated with the mRNA fate. The steady-state poly(A) length distribution can serve as a window for investigating post-transcriptional regulations as it reflects modulation in deadenylation speed. Thus, dissecting the mRNA heterogeneity by poly(A) profiles would open a possibility to reveal diverse fates of mRNA at the post-transcriptional level.

Although there is a room for refinement, yet, the clustering analysis enables investigation on the post-transcriptional control for the functional group of mRNAs. The most interesting example of such functional group is C1 in the clustering. Particularly, about half of the ribosomal protein-coding genes belong to C1, implying that these specific group of genes might share a special set of post-transcriptional regulation. Interesting points that need further attention are 1) how they keep their poly(A) tails short, and 2) how they remain stable even after multiple rounds of translation by polysomes, which is reported to be coupled to mRNA degradation (Roy & Jacobson, 2013). Their strategy might involve avoiding microRNA targeting by their exceptionally short 3′ UTRs, or having superior codon optimality, which is known to be a major determinant

of mRNA stability (Presnyak et al., 2015).

# 3. The discovery of the poly(A) barricade

## 3.1 Background

### 3.1.1 Poly(A) length and the mRNA lifespan

In human, most mRNAs are polyadenylated during transcription termination, resulting in the long stretch of adenosines upto ~250 nt (Kühn et al., 2009). In the nucleus, at the early step of polyadenylation, nuclear PABP (PABPN1) binds to the tail to stimulate processive adenosine incorporation catalysed by poly(A) polymerase. Also, it provides a surveillance mechanism for proper polyadenylation and nuclear export by limiting poly(A) extension to a certain length (Bresson & Conrad, 2013; Wigington et al., 2014). When the mRNA gets out of the nucleus, nuclear PABP is replaced by cytoplasmic PABPs (PABPC1–4). The cytoplasmic PABPs promote cap-dependent translation (Kahvejian et al., 2005) and protect mRNAs from precocious decay by preventing access of the decaying enzymes (Lim et al., 2014). As an mRNA ages, the poly(A) tail is trimmed, and eventually degraded when the tail reaches a certain length (~25 nt) to which PABP can no longer bind (Figure 3.1). After PABP detachment, the decay of naked RNA is facilitated by a number of decay machinery including TUT4 and TUT7, LSM complex, DCP1/2, XRN1, and the exosome complex (Lim et al., 2014). As a consequence, it is generally considered that poly(A) length reflects the remaining lifespan of mRNAs.
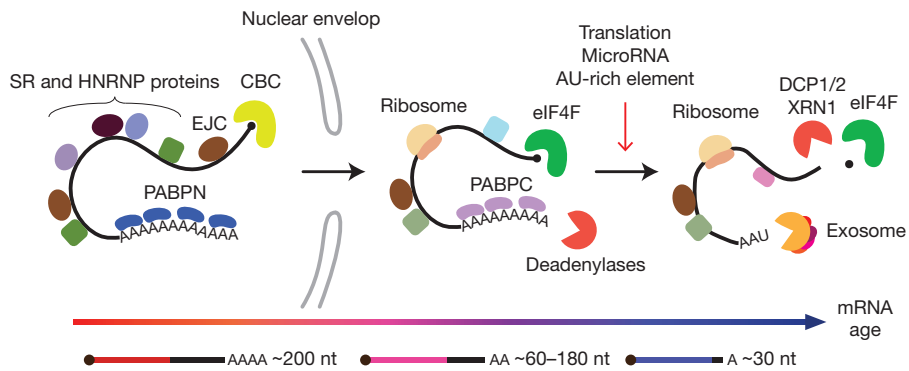
**Figure 3.1** The poly(A) tail and the associated mRNP change along with mRNA ageing.

In line with the notion, deadenylation is the first and rate-limiting step in mRNA turnover (Decker & Parker, 1993; Goldstrohm & Wickens, 2008). To modulate mRNA stability, many mRNA post-transcriptional regulatory pathways rely on the poly(A) length regulation. For instance, a microRNA destabilises its targets by facilitating deadenylation by recruiting deadenylase complexes (Eulalio et al., 2008), HuR and other AU-rich element binding proteins also destabilise their targets in a similar manner (Chen & Shyu, 2011). Shortened poly(A) tails that lost PABP binding become readily accessible to the decay-facilitating enzymes and rapidly decomposed. On the other hand, oppositely, tails can also be elongated again by cytoplasmic polyadenylation by terminal nucleotidyl transferases such as GLD2 (Winata & Korzh, 2018). TENT4A (PAPD7) and TENT4B (PAPD5) also can extend the poly(A) tail but occasionally incorporate guanosine residues during polyadenylation, resulting in a mixed tail that shields mRNAs from rapid deadenylation (Lim et al., 2018). Thus, the poly(A) tail reflects the remaining mRNA lifespan and serves

30

as a regulatory platform for many gene regulatory mechanisms which determine the mRNA fate in diverse biological contexts.

### 3.1.2 Paradoxical relationship between poly(A) length and mRNA stability

As described in the previous section, poly(A) tails are essential for mRNA as it prevents precocious decay and promotes translation. Counter-intuitively, however, the mean poly(A) length negatively correlates with mRNA half-life and translation efficiency. Recently, Lima et al. reported an intriguing observation that the highly expressed and stable messages have relatively short poly(A)s, which is consistent with the clustering analysis in this study (Figure 2.12), and this feature is deeply conserved from yeast to human (Lima et al., 2017). This relationship challenges the intuition that the longer poly(A) tail means higher stability as mRNAs are destabilised through deadenylation. To explain the paradoxical relationship between poly(A) length and mRNA stability, they proposed a model that more efficiently translated mRNAs form the more stable closed-loop owing to frequent interaction between eRF3 and PABP, thereby preventing complete deadenylation and leaving short-tailed mRNAs. However, currently, there is no evidence that eRF3 strengthens the closed-loop configuration. Besides, it is unclear whether the closed-loop configuration can stabilise mRNAs by preventing deadenylation. Thus, it still remains elusive that how stable and highly expressed mRNAs have short poly(A) tails.

### 3.1.3 The model of mRNA looping

PABPC1, which covers the poly(A) tail, directly interacts with eIF4G, the 5′ cap-binding complex component, leading to the 'mRNA looping'

31

model. The eRF3-mediated mRNA stabilisation model to explain short poly(A) tails on highly expressed mRNAs relies on this mRNA looping (Lima et al., 2017). It has been suggested that the physical interaction between the cap binder eIF4F (which contains eIF4G) with the tail binder PABP forms a closed-loop to promote translation (Tarun & Sachs, 1996; Imataka et al., 1998; Kahvejian et al., 2001). Moreover, an *in vitro* experiment that reconstituted circular mRNP using purified components (Wells et al., 1998), and another experiment that visualised circular polysomes also support the closed-loop model (Christensen et al., 1987). In addition, crosslinking and immunoprecipitation sequencing of PABPC1 reveals that PABP binding sites are found not only at the poly(A) tail but also within 5′ UTR, presumably due to the close proximity of the two extremities of an mRNA, supporting the notion of mRNA looping (Kini et al., 2016).

However, a recent study using structured illumination microscopy combined with single-molecule resolution fluorescent *in situ* hybridisation undermines the concept of closed-loop-forming mRNAs. According to this study, majority of mRNP complexes, especially those that are being translated, have open conformation, at least for those specific mRNAs examined in the study (Adivarahan et al., 2018). Translating mRNAs are rarely seen in the closed-loop configuration, indicating either that 1) mRNAs in closed-loop structure are not translated, or 2) the circularisation is transient so not being captured by the method.

In summary, although there are many indirect evidence for mRNA looping, it is not clear whether endogenous mRNAs circularise when they are being translated, and little is known how the circular configuration affects the mRNA deadenylation and decay processes.

## 3.2 Deadenylation dynamics and the steady-state poly(A) profile

Under the simple model of deadenylation where deadenylation rate is constant and the trimming occurs in a stepwise manner throughout the whole catabolic process, the steady-state poly(A) length distribution should be uniform as the probability of having a certain length of poly(A) tag is dependent only on the transition rate, the parameter assumed to be invariable. The initial input and output difference does not alter the shape of the stationary distribution but only determine the expression level at equilibrium (Figure 3.2, left). However, if there is disturbance in deadenylation, poly(A) tags will accumulate around the range where deadenylases slow down (Figure 3.2, right). In other words, if there is a change in the deadenylation rate along the way of trimming poly(A), it will be reflected in the shape of steady-state poly(A) length distribution. Thus, the steady-state poly(A) profile can be viewed as a proxy for investigating deadenylation dynamics.

### 3.2.1 Deadenylation slows down at ~30 nt position

In real datasets, the global steady-state poly(A) length distribution is not flat; it peaks at near 30 nt and skewed to the right, suggesting that deadenylation slows down at ~30 nt position. This shape is reproducible in the technically different poly(A) length measurements: mTAIL-seq, PAL-seq, and the BPA (Figure 3.3).[1] Meanwhile, the peak around 30 nt consistently appears not only in human but also in a variety of model organisms (Subtelny et al., 2014). Additionally, the accumulation of poly(A) tags at

---

[1]mTAIL-seq and BPA were performed by Hyerim Yi
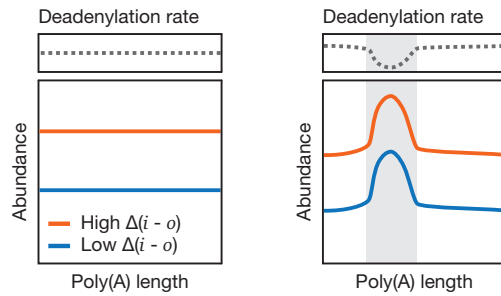
**Figure 3.2** The theoretical steady-state distribution of poly(A) tails under a simple modelling for the deadenylation process where the trimming occurs in a stepwise manner and the deadenylation rate is constant (left) or there is disturbance in the middle (right). $\Delta(i - o)$: initial input and output difference.

the short poly(A) range (~30–60 nt) can be seen at an individual gene level; when the mode of poly(A) length is plotted, the dots are densely spotted near ~30–60 nt range (Figure 3.4). This suggests that retarded deadenylation at the short poly(A) is prevalent at a transcriptomic level.

### 3.2.2 PABP and possibly other factors may be required to decelerate deadenylases

Interestingly, it seems that PABP is somehow involved in the formation of the peak around 30 nt. There are a number of reasons; firstly, the position of the peak corresponds to the size of a single PABP-spanning segment, which is ~27 nt (Smith et al., 1997; Wang et al., 1999). Also, it is plausible that the fraction upstream of the peak sharply drops in frequency, presumably due to that mRNAs with very short tails are readily decomposed. Short poly(A) tails that are <~30 nt would no longer be able to bind PABP, so they are more susceptible to be targeted by decay machin-
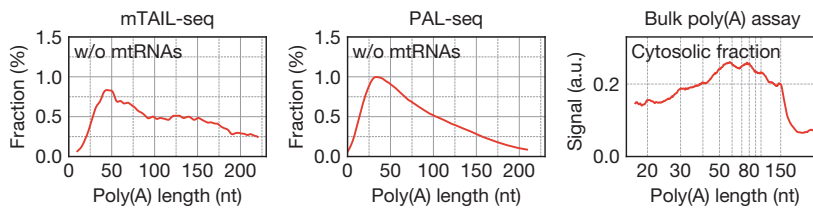
**Figure 3.3** The steady-state poly(A) length distributions measured by various methods. For Bulk poly(A) assay, cells were fractionated to obtain the cytosol fraction from the total RNA. The PAL-seq dataset were adopted from (Subtelny et al., 2014). MT: mitochondrially-encoded genes, Cyt: the cytosol fraction.

ery (Lim et al., 2014). Lastly, PABP seems to make a phasing pattern with a ~25–30 nt interval in *C.elegans*, which peaks at ~30 nt and gradually decrease in height (Lima et al., 2017). This further supports that PABP may be responsible for slowing down deadenylation at some points, particularly when the last one or two PABPs are tightly bound with little space to move along the tail.

Pab1, a fission yeast homolog of PABP, has both facilitating and blocking functions on the Ccr4 and Caf1 deadenylation *in vitro*. Using a fully reconstituted biochemical system, it was shown that Pab1 can facilitate Ccr4 activity while blocking Caf1 function. Also, Ccr4 deadenylation of Pab1-bound poly(A) delineated RRM footprints, suggesting a possibility that continual pauses of deadenylation can happen solely with PABP by itself. Therefore, I was wondering if the contradictory effects of PABP on deadenylation may be responsible for the short poly(A) accumulation in human cells. To check whether PABP is the only required factor for blocking deadenylation at around 30 nt, the *in vitro* deadenylation as-
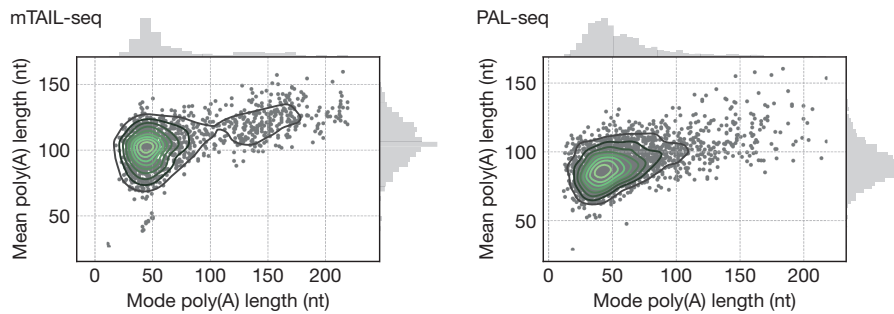
**Figure 3.4** The mean and mode of poly(A) length of individual genes in HeLa cells, which were measured by mTAIL-seq (left) or PAL-seq (right). The PAL-seq dataset was adopted from (Subtelny et al., 2014). The green-to-black contours represent the bivariate kernel density estimates.

say was performed with immunopurified FLAG-CCR4a and CAF1b. The A50 poly(A) substrate which comprises 20 nt CALM1 3′ UTR followed by 50-mer adenosines were incubated with the immunopurified wild-type or mutant FLAG-CCR4-CAF1 in the presence of absence of recombinant PABPC1, and the deadenylation rates were inferred for each condition (Figure 3.5).[2] The results showed that deadenylases do not decelerate at around 30 nt, rather, increase in speed a little as it approaches to the non-poly(A) part, in both PABPC1 presence and absence. This implies that, with only PABPC1, it may not be sufficient to explain the 30-nt deceleration of deadenylases in human cells.

---

[2]The experiment was performed by Hyerim Yi
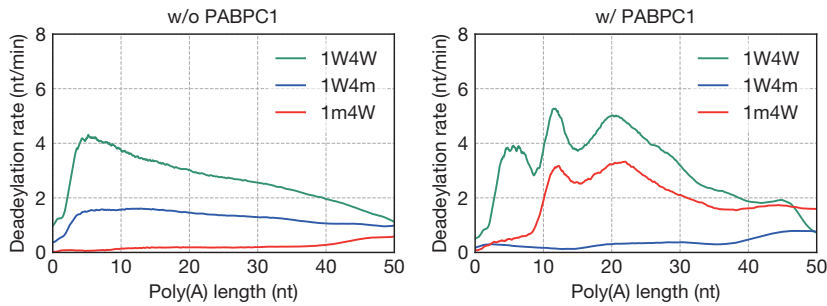
**Figure 3.5** Inference on the *in vitro* deadenylation rates. PABPC1 is pre-incubated (right) or not (left) prior to *in vitro* deadenylation assay. 1W: wild-type CAF1a 1m: catalytic mutant CAF1a, 4W: wild-type CCR4b, 4m: catalytic mutant CCR4b.

## 3.3 PABP interacting partners and their positions on the poly(A) tail

Considering that PABP is likely to be involved in reduced deadenylation at certain points, particularly around where one or two PABP are packed in, while it is not solely sufficient to exhibit such function, I suspected that there might be unknown factors that associate with PABP and block deadenylation to stabilise poly(A) tails. With an assumption that PABPC1 constitutes a structural basis for the deadenylation-blocking complex, I firstly tried listing up PABPC1 interacting partners to unveil the unknown factors.

### 3.3.1 Revealing PABPC1 interacting partners by PABPC1 co-IP followed by LC-MS/MS

To identify PABPC1 interacting partners, I analysed the liquid chromatography and tandem mass spectrometry (LC-MS/MS) data for the PABPC1

co-IPed proteins.[3] There were two trials, and overlapping genes that showed >2-fold enrichment in both trials, when using Myc co-IP sample as control, are listed in Table 3.1. PABPC1 was the top among the all enriched proteins as expected, and the protein that previously shown to interact with PABPC1 such as PAIP1 and MKRN1 were highly enriched in the PABPC1 co-IPed protein sample (Derry et al., 2006; Miroci et al., 2012).

### 3.3.2 LARPs are prominent candidates that constitute the poly(A) barricade

Among the identified PABPC1 interactants, I selected several candidates expected to constitute the deadenylation-blocking 'poly(A) barricade' for closer inspection. Considering their known biochemical properties and subcellular localisation, the finally chosen candidates were the La-related protein family members (LARPs). In human, there are seven LARP paralogs (LARP1–7), and the three of them, LARP1, LARP4, LARP4B (LARP5), appeared in the PABP-interacting protein lists in both trials. The domain structures of those LARPs are illustrated in Figure 3.6.

Human LARPs and their close relative La protein (hLa) share the La motif (LaM) accompanied by the RNA recognition motif (RRM), collectively called the La module, for their RNA binding. hLa and one of the family member LARP7 bind to UUU-3′-OH, the characteristic of RNA polymerase III (RNAP III) transcripts. Structural and biochemical studies on hLa revealed that the La module provides a fold that the terminal UUU-3′-OH can be bound. With its specific RNA binding ability and nuclear localisation, hLa is known to block precocious decay and prevent misfolding of RNAP III transcripts (Maraia et al., 2017).

---

[3]The experiment was performed by Hyerim Yi

**Table 3.1** PABPC1 co-IPed proteins identified using LC-MS/MS.

| # | Protein IDs | Replicate 1 | Replicate 2 | Note |
|---|---|---|---|---|
| | | log2 (aPABPC1/aMyc) | | |
| 1 | PABPC1 | 25.18 | 6.16 | IP target |
| 2 | FIBG | 23.88 | 5.18 | ECM component |
| 3 | FINC | 22.50 | 20.56 | ECM component |
| 4 | PABPC4 | 21.86 | 6.40 | PABPC1 paralog |
| 5 | ABCB9 | 19.84 | 19.17 | Lysosomal transmembrane |
| 6 | LARP1 | 19.28 | 6.93 | RNA binding, Cytosol |
| 7 | LARP4 | 18.91 | 5.04 | RNA binding, Cytosol |
| 8 | TPRN | 17.11 | 16.86 | Non RBP |
| 9 | PABPN1 | 16.87 | 22.69 | Nuclear PABP |
| 10 | PAIP1 | 16.86 | 20.52 | PABPC1 binding |
| 11 | MLF2 | 16.82 | 17.39 | DNA binding, Nucleus |
| 12 | HNRNPR | 16.35 | 20.21 | RNA binding, Nucleus |
| 13 | LARP4B | 15.91 | 21.60 | RNA binding, Cytosol |
| 14 | SOAT1 | 15.63 | 15.69 | Cholesterol binding |
| 15 | A2M | 15.25 | 16.10 | Proteinase inhibitor |
| 16 | MKRN1 | 15.22 | 20.23 | RNA binding, E3 ligase, Cytosol |
| 17 | CO3 | 14.14 | 16.61 | Complement C3 |
| 18 | ARPC1B | 14.12 | 14.24 | Actin polymerisation |
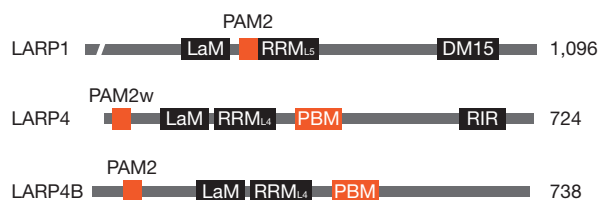
**Figure 3.6** Domain architecture of LARP1, 4, 4B. They appeared in PABPC1 co-immunoprecipitates and emerge as prominent candidates constituting the barricade on poly(A) tails. PABP-interacting domains are coloured in orange.

On the other hand, LARP1, 4, 4B, and 6 are cytoplasmic LARPs and altered their specificity from UUU-3′-OH to a poly(A) sequence; their La modules exhibit preferential binding to the poly(A) tail. Also, they acquired ability to interact with cytoplasmic poly(A)-binding protein (PABPC) through the PAM2 or PAM2-like domains (Maraia et al., 2017). Together, all these biochemical characteristics imply that the cytoplasmic LARPs play roles in mRNA tail regulation. Indeed, there have been a number of reports that demonstrated their functions in the mRNA post-transcriptional control, supposedly with the help of their poly(A)-binding abilities. Firstly, LARP1 was shown to be involved in translational control of a subset of mRNAs having the 5′ terminal oligopyrimidine (5′ TOP) motif, by simultaneously recognising the 5′ terminal 7-methylguanosine cap and the very first cytidine of the 5′ TOP motif through the LARP1-specific DM15 domain (Lahr et al., 2017; Philippe et al., 2018). Secondly, both LARP1 and LARP4 were implicated in stabilisation of mRNAs. The loss of LARP1 leads to destabilisation of 5′ TOP mRNAs (Aoki et al., 2013; Gentilella et al., 2017), and the ectopic expression of LARP1 or LARP4 result in

poly(A) tail lengthening of the ribosomal protein-coding mRNAs (Mattijssen et al., 2017). Taken together, with these hints from the previous studies, I postulate LARPs may constitute the barricade on poly(A) tails together with PABP to block complete deadenylation, thereby uncouples deadenylation from the subsequent decay processes.

## 3.4 Examination of the candidates regarding poly(A) length regulation

For the next step, I examined the potential roles of the candidates regarding poly(A) length regulation by RNA IP or knockdown followed by BPA.[4] In addition to LARPs, I included PABPC1 and eIF4G1 to check the possibility that the closed-loop configuration through PABP-eIF4F interaction might cause the deadenylation retardation for short-tailed mRNAs. To elucidate the positions where the candidates lie on poly(A) tails, cytoplasmic RNA IP followed by BPA was performed. To avoid artificial RNA-RBP interaction after cell lysis, for instance, PABPC1 binding to mitochondrial mRNAs, subcellular fractionation preceded the RNA IP. As a result, I find that the candidates have the distinct poly(A) binding profiles (Figure 3.7). Of note, PAIP1 was also tested but it did not pull down enough RNA to be analysed by BPA, presumably due to the lack of RNA binding activity, so I dropped PAIP1 in the future experiments (data not shown).

---

[4]The experiments in this section were carried out by Hyerim Yi
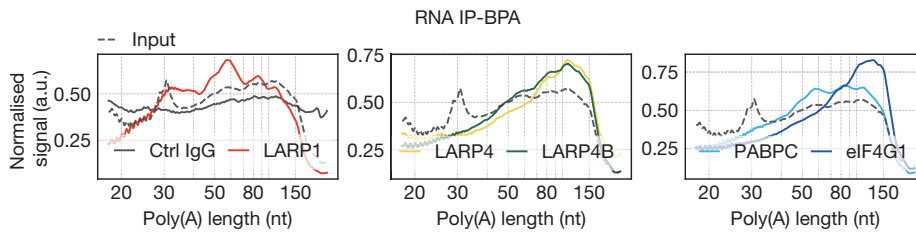
**Figure 3.7** Poly(A) binding profiles of the barricade candidates. After subcellular fractionation, RNA IP followed by the bulk poly(A) assay was performed with the cytosol fraction. The results were visualised in lanewise normalised densitogram to show the positions on the poly(A) tail for the candidates. The same cytoplasmic input sample (Input, dashed dark grey line) was drawn in all three plots for comparison. The $x$ axis grid indicates the laddering positions of the decade marker, which contains 20–100 nt plus 150 nt standards.

### 3.4.1 Poly(A) binding landscapes of the candidate proteins

The most intriguing poly(A) binding landscape was of LARP1. LARP1 co-IPed RNAs showed strong enrichment around ~30–90 nt region. Besides, it showed the clear periodic enrichment pattern with an interval of ~30 nt, which is reminiscent of PABP footprints. Considering that the positions of peaks correspond to the peaks in the phasing pattern that appear in the steady-state poly(A) length distribution, LARP1 might be the one that constitutes the poly(A) barricade. Unlike LARP1, LARP4 and 4B enriched longer poly(A) tails ranged around ~70–190 nt region. Their poly(A) binding profiles resemble each other but clearly differ from that of LARP1, implying that LARP4 and 4B may share targets but not with LARP1. On the other hand, eIF4G1-bound RNAs showed the longest poly(A) tails that

are >~80 nt, and mainly ~150 nt, while PABPC1-bound RNAs showed broad enrichment, around ~40–160 nt. Given that neither eIF4G1 nor PABPC1 strongly associate with short poly(A) tails, the mRNA looping mediated by PABP-eIF4F interaction is not likely to induce the deadenylation deceleration for preserving the short-tailed mRNAs. Rather, fresh mRNAs with long poly(A) tails that are ready for active translation seem to be more strongly bound to eIF4G1 and PABPC1.

## 3.4.2 The global poly(A) profile changes after knockdown of each candidate

In addition to RNA IP, knockdown followed by the BPA was performed to examine the consequence on poly(A) profile after depletion of each candidate (Figure 3.8). Theoretically, if a candidate protein functions as a deadenylation barricade, the candidate knockdown would result in the change in the stationary poly(A) length distribution; where the candidate protein lies would decrease in frequency because of the accelerated deadenylation on that particular region. Thus, it might look somewhat complementary to the poly(A) binding profile. In consistent with this speculation, in LARPs knockdown samples, the putative candidate-binding positions on the poly(A) showed decreased signals. Notably, in eIF4G knockdown, poly(A) tails were shortened compared to the control, suggesting that eIF4G binding to the long poly(A)-tailed mRNAs might have protective effect on the mRNAs. One exception that showed the changes that are not complementary to the poly(A) binding profile was PABPC1, presumably due to that it has both facilitating and inhibitory functions in mRNA degradation.
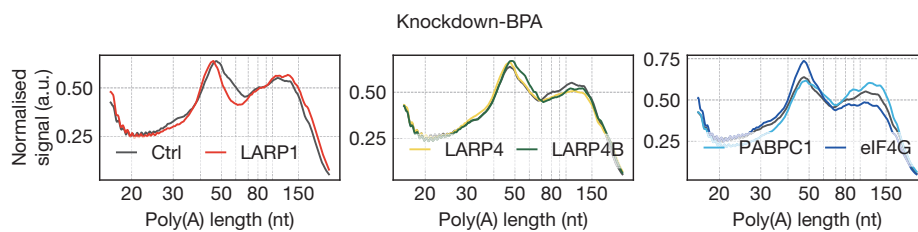
**Figure 3.8** Poly(A) profile changes after depletion of each candidate. The indicated candidates were targeted by siRNAs and the poly(A)s were visualised by the bulk poly(A) assay. For eIF4G, both eIF4G1 and eIF4G3 were depleted. The same negative control sample (Ctrl, dark grey line) was drawn in all three plots for comparison. The $x$ axis grid indicates the laddering positions of the decade marker, which contains 20–100 nt plus 150 nt standards.

### 3.4.3 Transcriptome-wide poly(A) profiling reveals the specificities of the candidates

To investigate the details of the RNAs bound to the poly(A) barricade candidates, I applied mTAIL-seq for genome-wide poly(A) profiling. Firstly, I checked whether the poly(A) binding profiles that were seen in the BPA are reproducible in the mTAIL-seq dataset. Although the degree of the enrichment was smaller than the previous experiment particularly for LARP1, in overall, the poly(A) binding landscapes were reproduced well for all examined proteins: LARP1 for ~30–60 nt (not to 90 nt in this experiment), LARP4 for ~70–190 nt, eIF4G1 for >~80 nt (Figure 3.9). Secondly, to check differential binding among mRNA species, I compared poly(A) binding profiles at the subgroup level, with the clusters defined in Figure 2.9. Despite clearly different poly(A) length distribu-

tions among the clusters in Input, the preferred binding positions on the poly(A) tail were consistent among the clusters for all candidates (Figure 3.10). This indicates that all the examined candidates basically can act on most poly(A) tails, if not all, regardless of mRNA body sequences.
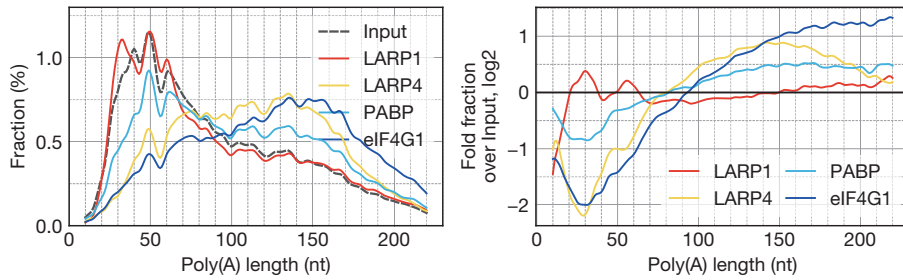


**Figure 3.9** Genome-wide poly(A) length profiling for the barricade candidates RNA IP samples using mTAIL-seq. Global poly(A) length distribution (left) and log2 fold fraction over Input (right).

For the knockdown samples, the changes in poly(A) profile were reproduced well in the mTAIL-seq dataset. The poly(A) ranges correspond to the putative poly(A) binding positions were decreased in level compared to the control, reflecting the changes in deadenylation dynamics specifically take place at the barricade-spanning regions (Figure 3.11). Consistently with the poly(A) binding landscapes in the subgroups, the poly(A) profile changes at the subgroup level were comparable among the clusters (Figure 3.12).

However, it is expected that LARP1 would have at least some level of specificity for 5′ TOP mRNAs because there are a number of reports that showed the preference of LARP1 toward mRNAs with the 5′ TOP motif (Maraia et al., 2017; Hong et al., 2017; Lahr et al., 2017). In addi-

**Figure 3.10** The cluster poly(A) length distribution in Input, and the poly(A) binding landscapes at a subgroup level. The clusters were defined in Figure 2.9, and C0 is omitted as it exclusively contains only mitochondrial mRNAs.

**Figure 3.11** Poly(A) profile changes upon depletion of the barricade candidates. Global poly(A) length distribution without mitochondrial mRNAs (left) and the fold fraction over siNC (right).

tion, there is a report that suggested LARP1 and eIF4G1 might compete with each other for binding to the 5′ cap of 5′ TOP mRNAs, as LARP1 can replace eIF4G1 for binding to the 7-methylguanosine-capped RPS6 5′ UTR *in vitro* (Lahr et al., 2017). Thus, to further elucidate the specificity for 5′ TOP mRNAs, I looked into the relative enrichment of mRNAs in the pull-down samples (Figure 3.13). In line with the previous reports, LARP1 showed the strong affinity toward the 5′ TOP mRNAs (P=3.6e-12, Fisher's exact test). Interestingly, conversely to LARP1, the 5′ TOP mRNAs were significantly depleted in the eIF4G1 pull-down sample (P=5.1e-7, Fisher's exact test), implying that LARP1 does compete with eIF4G1 for the 5′ cap binding in cells as well, particularly for those 5′ TOP mRNAs. In contrast, LARP4 and PABPC1 did not show any statistically significant binding preference for those mRNAs. Taken together, the transcriptomic analyses suggest that the poly(A) barricade can act on any mRNAs that having poly(A) tails, but the LARP1 barricade has some level of specificity particularly for the 5′ TOP mRNAs.
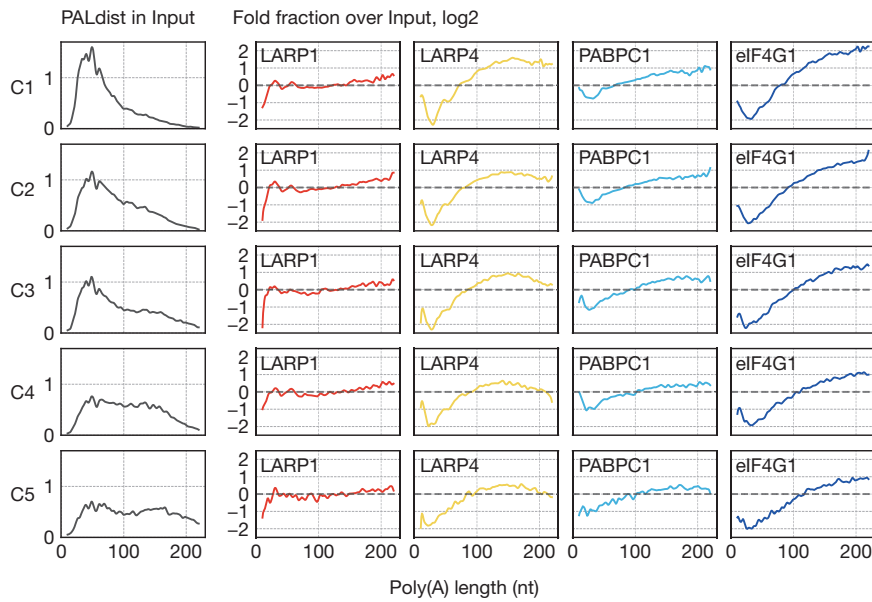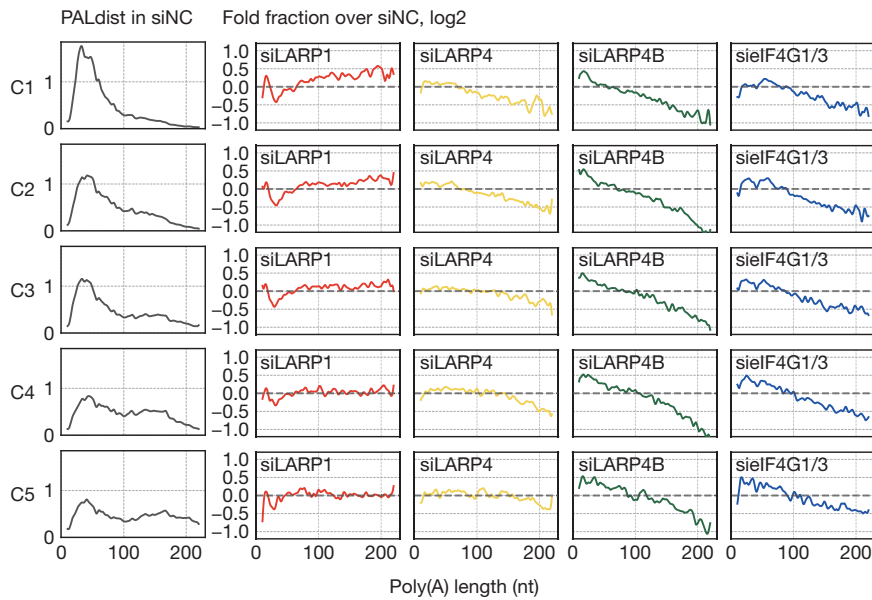
**Figure 3.12** The cluster poly(A) length distribution in the control, and the poly(A) length changes at a subgroup level. The clusters were defined in Figure 2.9, and C0 is omitted as it exclusively contains only mitochondrial mRNAs.
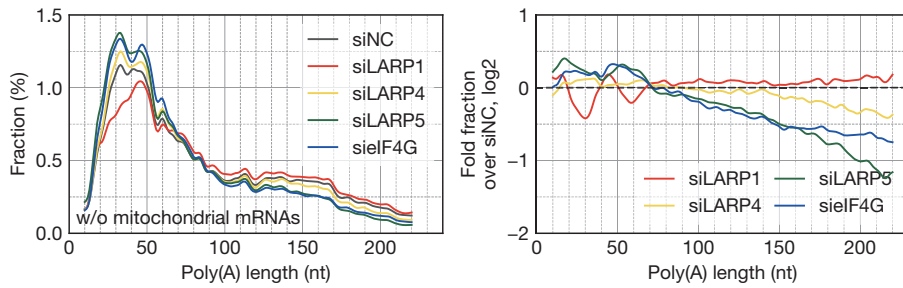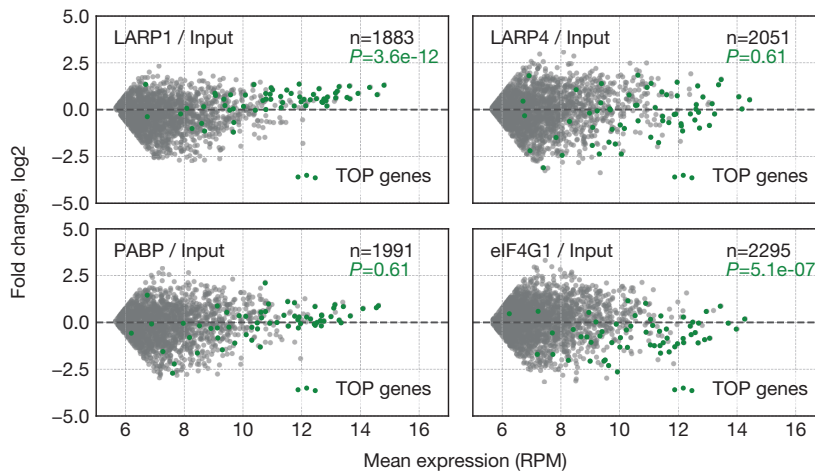
**Figure 3.13** MA plots for the barricade candidate RNA IP samples. The abundance were estimated from the mTAIL-seq read counts normalised by read-per-million (RPM). The list of human TOP genes was adopted from Yamashita et al. (Yamashita et al., 2008). n: number of genes, *P*: *p*-values calculated from the two-tailed Fisher's exact test for the TOP genes.

## 3.5 LARP1 binds to poly(A) and inhibits deadenylation *in vitro*

As LARP1 emerged as the most prominent candidate that exhibits the position-specific deadenylation retardation on short poly(A) tails, I further tested whether LARP1 has inhibitory function on deadenylation. To this end, a series of *in vitro* experiments were performed with a full length purified recombinant LARP1 protein.[5] In this study, among the several variants of human LARP1, the 1,063 a.a. variant form which contains all the major functional domains, including the La module and DM15,

---

[5]The experiments in this section were carried out by Dr. Kyungmin Baeg

was used in the assay. To prepare the recombinant LARP1 protein, 6Lys-MBP-6His-3C-tagged LARP1 was expressed in *E.coli* and purified (Figure 3.14).
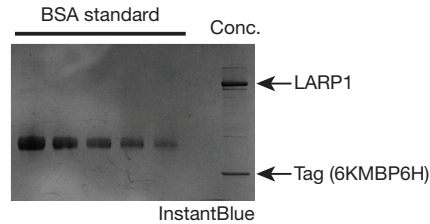


**Figure 3.14** Purified recombinant LARP1. The N-terminal 6Lys-MBP-6His-3C (6KMBP6H3C) tag was added to the full length 1,063 a.a. LARP1 for *E.coli* expression and purification.

To check the integrity of the purified recombinant LARP1, the electrophoretic mobility shift assay (EMSA) was performed with a chemically synthesised 20 nt CALM1 A50 RNA (Figure 3.15). Of note, although it was shown that LARP1 in cell lysate binds to the poly(A) RNA bait (Aoki et al., 2013), it has remained untested whether the purified full length LARP1 protein can bind to poly(A) in the *in vitro* experimental condition. In this assay, LARP1 bound to the RNAs with 50 adenosine residues, indicating that recombinant LARP1 retains the poly(A)-binding activity. Also, LARP1-poly(A) binding was sub-$\mu$M *Kd*, which is comparable to the *Kd* of N-terminal LARP4 fragment (111–303), which contains only the La module, to 15 nt poly(A) RNA (Yang et al., 2011).

Next, another EMSA was conducted to test the poly(A) binding of LARP1 in the presence or absence of PABPC1 (Figure 3.16). Under 100 nM PABPC1, most RNA molecules were shifted, meaning that most poly(A)s are bound by PABPC1 in this condition. Notably, A50 showed
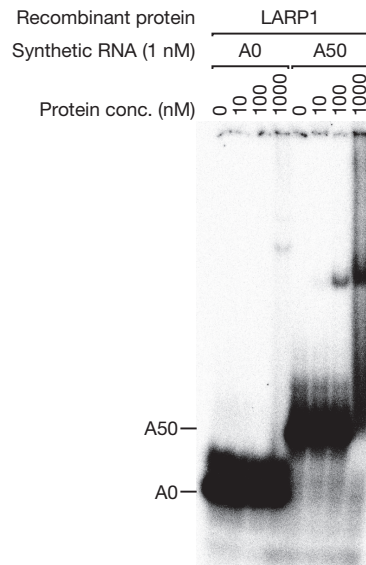
**Figure 3.15** Electrophoretic mobility shift assay with the recombinant LARP1. A0 and A50 are chemically synthesised RNA which carry 20 nt of CALM1 3′ UTR (A0) plus a 50-nt poly(A) sequence (A50). 6% polyacrylamide native PAGE gel is used for the assay.

the two bands which correspond to the ribonucleoprotein complexes that have one (lower band) and two (upper band) PABPC1 molecules, respectively. In this setup, the results showed that LARP1 also can bind to the PABPC1-coated poly(A)s as well as to the naked poly(A)s. This suggests that LARP1 may bind to the very end of poly(A) while PABPC1 is covering the poly(A) body. Notably, there are some evidence supporting the idea of end recognition in the previous studies; it was shown that LARP1 and LARP4 can discriminate the very end base and prefer an adenosine (Aoki et al., 2013; Mattijssen et al., 2017).

After confirming the condition that both LARP1 and PABPC1 can
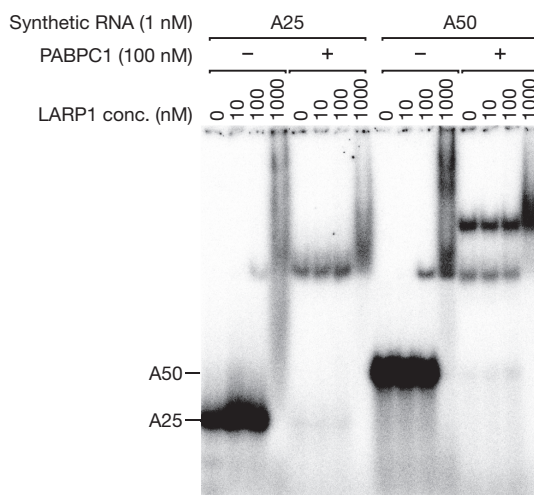
**Figure 3.16** Electrophoretic mobility shift assay with recombinant LARP1 and poly(A) in the presence or absence of PABPC1. A25 and A50 are chemically synthesised RNA which carry 20 nt of CALM1 3′ UTR plus 25 or 50 nt poly(A) sequence. 6% polyacrylamide native PAGE gel is used for the assay.

simultaneously bind to poly(A) by EMSA, the *in vitro* deadenylation assay was performed by immunopurified FLAG-CAF1a and CCR4b to check whether LARP1 has inhibitory effect on deadenylation.[6] The poly(A) substrates were incubated for 30 minutes with varying amounts of LARP1, in the presence or absence of PABPC1 (Figure 3.17). Strikingly, deadenylation by CCR4-CAF1 was clearly reduced when LARP1 exists, and the degree of the retardation was proportional to the amount of added LARP1, establishing LARP1 as a genuine deadenylation inhibitor.

---

[6]The immunopurified FLAG-CAF1a-CCR4b was prepared by Hyerim Yi
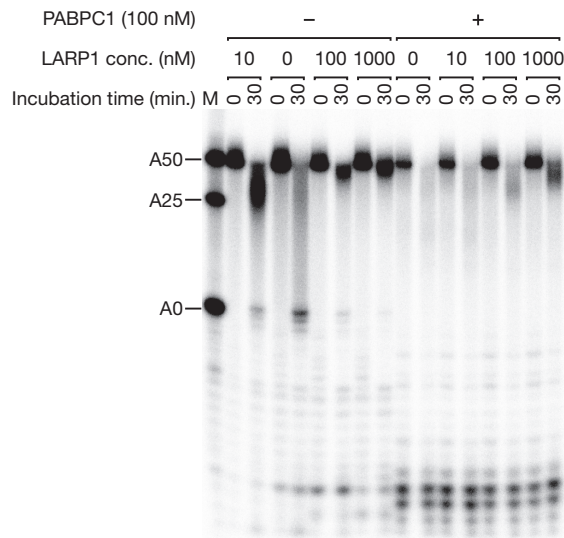
**Figure 3.17** The *in vitro* deadenylation assay by immunopurified CCR4-CAF1, with PABPC1 and varying amount of LARP1. Recombinant PABPC1 and LARP1 were pre-incubated before the CCR4-CAF1 reaction. The chemically synthesised A50 RNA comprises 20 nt CALM1 3′ UTR followed by 50 adenosines was used as a poly(A) substrate.

## 3.6 Model of the poly(A) barricade

Based on the results from the series of cellular and *in vitro* experiments, I unveil the poly(A) barricade which blocks deadenylation to protect poly(A) tails, and propose a model for its roles in post-transcriptional regulation of mRNAs (Figure 3.18). There are two types of the barricade which contains different LARPs: LARP4/4B and LARP1. When poly(A) tails are ~70–190 nt in length, LARP4/4B binds and protects the tail from stochastic processive deadenylation while promoting translation, which ensures protein production and blocks premature decay of the mRNAs. Translation-enhancing function of LARP4/4B is mediated by interacting

with RACK1, which is a constituent of ribosomal small subunit (Yang et al., 2011; Schäffler et al., 2010), and LARP4/4B inhibits deadenylation by capping the 3′ end of poly(A) tail. Later, when the tails are trimmed to have shorter poly(A)s about ~30–60 nt, which correspond to the length in which one or two PABPs can lie, LARP1 comes in and replaces LARP4/4B to more strongly block poly(A) degradation. Binding of LARP1 causes the position-specific short poly(A) accumulation that spawns the phased poly(A) length distribution. Of note, although LARP1 can inhibit deadenylation even in the absence of PABP *in vitro*, the tight association between LARP1 and PABP together with the 3′ end of poly(A) tail might reinforce the LARP1 binding, resulting in the poly(A) binding profile that resembles PABP footprints. Also, given that LARP1 can bind to the 5′ cap, LARP1 may can locate the both extremities in close proximity and facilitate the formation of a closed head-to-tail configuration, which might enhance the protective effect of LARP1. Particularly, ribosomal protein-coding genes, which have TOP motif at their 5′ head, may be more avidly bound to LARP1 so can be more strongly stabilised.

## 3.7 Discussion

In this chapter, I investigate deadenylation dynamics by looking at the steady-state poly(A) length distribution which can serve as a proxy for estimating the changes in deadenylation rates. As a result, I find that deadenylation slows down at ~30 nt position and discover the poly(A) barricade which protects the 3′ end of poly(A) tail. I suggest LARP1 as the prominent barricade component that is responsible for the position-specific deadenylation pauses, based on the evidence that 1) it shows a periodic poly(A) binding that resembles PABP footprints particularly
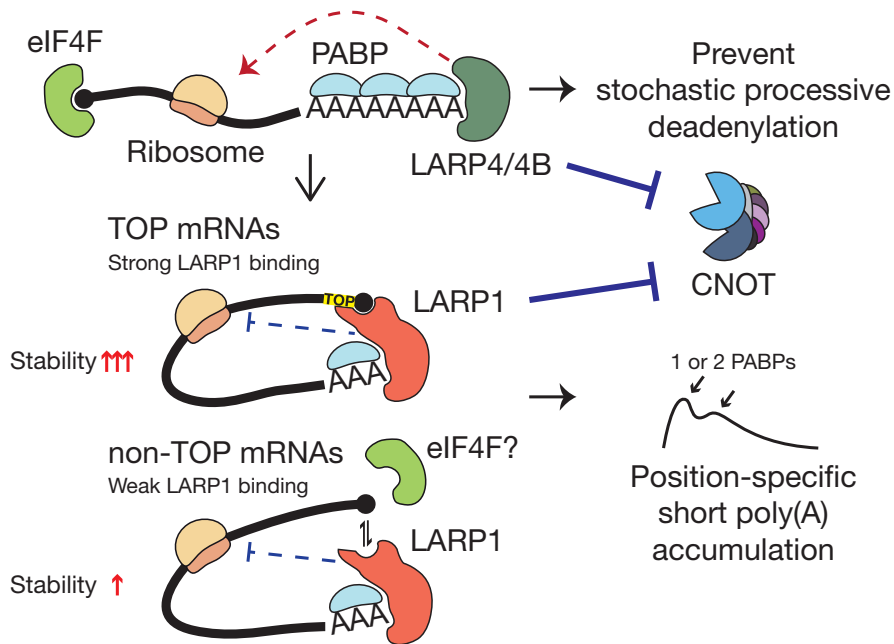
**Figure 3.18** Model of the poly(A) barricade. Together with PABPC1 (and possibly also with other components), LARP4/4B and LARP1 serve as the barricade on the poly(A) tail that inhibits deadenylation to stabilise mRNAs. TOP: 5′ terminal oligopyrimidine motif. CNOT: the CCR4-NOT complex

around ∼30–60 nt region, 2) knockdown of LARP1 results in the reduction of ∼30 or 60 nt heaps, and 3) recombinant LARP1 can directly interfere CNOT deadenylation *in vitro*. With its length-specific poly(A) binding and deadenylation-interfering activities, it may uncouple deadenylation from the following decay, leaving short-tailed but stable mRNAs. Moreover, specifically strong binding of LARP1 to the 5′ TOP motif offers the mechanistic insight on how highly expressed, house-keeping mRNAs can have short poly(A) tails, resolving the paradoxical relationship between mRNA stability and poly(A) length.

One immediate but barely addressed question is how the length-specific, and phased poly(A) binding of LARP1 can be achieved. Since the LARP1 poly(A) binding profile presents a periodic pattern with an interval of ~30 nt, which is reminiscent of PABP footprints, the strength of LARP1 binding may be somehow related to the PABP packing on the poly(A) tail. If LARP1 can bind more strongly to the end of poly(A) tail when PABPC1 is packed so it is more feasible to make the ternary complex, the phased pattern can be explained. Also, if this is true, short poly(A) preference of LARP1 may also be apprehensible because PABPC1 is more likely to be packed when the poly(A) tail is short. More in-depth biochemical and structural studies regarding poly(A)-end recognition of LARP1 would provide the answer in the future.

In the physiological aspect, LARP1 is deeply associated with protein synthesis. The 5′ TOP mRNAs, LARP1's favourite binding substrates, are mostly ribosomal protein coding (Yamashita et al., 2008). Also, LARP1 is known to be under regulation of mTOR signalling pathway, which is a master regulator of cellular metabolism, and represses translation of the TOP mRNAs downstream of mTORC1 (Fonseca et al., 2015). A more recent study further showed that LARP1 is phosphorylated by mTORC1 and Akt/S6K1, switching on and off LARP1 binding to the TOP mRNAs; the phosphorylation decreases the LARP1 affinity toward the 5′ TOP or the TOP-like motif. Phosphorylated LARP1 releases the head of the bound 5′ TOP mRNAs to turn translation on (Hong et al., 2017). Thus, when cells are in starvation, LARP1 protects, and may sequester the ribosomal protein-coding mRNAs in a translationally inactive dormant state. However, when the circumstance changes and mTOR signalling becomes active, LARP1 gets phosphorylated and being released for starting transla-

tion, thereby readily supplements translation machinery to promote over-all protein production. In this way, cells may be able to exploit resources more quickly and efficiently.

Another interesting point worth attention is the molecular function of LARP4/4B and its relationship to LARP1. It seems that both LARP1 and 4/4B can protect poly(A) tails considering that ectopic expression of those LARPs results in mRNA stabilisation and poly(A) lengthening (Mattijssen et al., 2017). However, differently from LARP1, LARP4/4B binds to the long poly(A) tails, indicating that LARP4/4B protects the mRNAs in the earlier stage of their metabolism. Intriguingly, LARP4/4B is known to recruit RACK1 (Schäffler et al., 2010; Yang et al., 2011), which is reported to be essential for full translation of capped mRNAs and efficient eIF4E recruitment (Gallo et al., 2018). The positive regulation of LARP4/4B for translation and mRNA stability may be devised to ensure efficient protein production for the messages in the early stage. As mRNA tails are trimmed, LARP4/4B is likely to be replaced by LARP1 by an unknown mechanism. There may be an active player which switches the LARPs depending on poly(A) tail length, or the compositional change in mRNP along the progression of mRNA life cycle might expel LARP4/4B and accept LARP1 without explicit measurement of the poly(A) length. In either way, revealing a more comprehensive list of the tail-binding RBPs would help understanding the poly(A) length regulatory mechanisms.

Other than LARPs, there is another candidate, MKRN1, that might play a role in poly(A) tail regulation. MKRN1 appears in the PABPC1 co-immunoprecipitates in our datasets, and is implicated in the mRNP interactome in mouse embryonic stem cell (Kwon et al., 2013). Also, a

mouse homolog of MKRN1 is co-IPed with mouse LARPs, implying that it may form a complex with LARPs to construct the poly(A) barricade (Cassar et al., 2015). Additionally, as MKRN1 is originally known for its E3 ubiquitin ligase activity, it would be interesting if ubiquitination activity of MKRN1 introduces and regulates the conformational change of the poly(A) barricade.

# 4. Conclusion

Regulation of the poly(A) tail is of central importance in post-transcriptional regulation as it serves as the first and often a rate-limiting step in mRNA degradation. Also, a great number of RNA-binding proteins associate with poly(A)-binding proteins to regulate mRNA translation and stability. However, the fundamental mechanisms of deadenylation has remained largely unknown; about how the deadenylase complexes select their substrates and to what extent they differ in contribution on shaping the poly(A) tail landscape.

The first part of this dissertation is to dissect the role and specificity of the human major deadenylases, the PAN2-PAN3 and the CCR4-NOT complexes. By performing global poly(A) tail profiling after depletion of each deadenylases, I revisit and revise the 'biphasic deadenylation' model which has been widely accepted but not thoroughly examined for a long time, particularly in a genome-wide scale. Here, I confirm that the size-specific deadenylation by PAN2/3 and CNOT can be applied to most endogenous mRNAs. Notably, however, I also find that PAN2/3 trimming does not necessarily precede the CNOT deadenylation. Even in the absence of PAN2/3, CNOT can deadenylate most part of the poly(A) tail so minimise the effect of PAN2/3 depletion. Furthermore, the statistical analysis reveals that nearly all mRNA tails are elongated in the CNOT-depleted cells, and the effect sizes were proportional to the cytosolic enrichment of mRNAs, establishing CNOT as a predominant and non-

specific enzyme that performs bulk deadenylation in the cytosol. As the poly(A) tail is a common feature of most mRNAs, exhibiting the specificity should be mediated by *cis*-acting elements. Therefore, discovering those regulatory elements and revealing the players that recruit deadenylases would be of interest in future studies.

The discovery of the poly(A) barricade offers the mechanistic explanation to the paradoxical phenomenon, that is the negative correlation between the mean poly(A) length and mRNA stability. The barricade impedes progressive and complete poly(A) degradation, thereby unlinks deadenylation from the subsequent decay processes, which results in the accumulation of short-tailed but stable poly(A)s. I reveal LARP1 as the most prominent position-specific deadenylation blocker on the poly(A), which provides the essential functionality to the poly(A) barricade. The strong affinity of LARP1 to the 5′ TOP motif supports that highly expressed, stable mRNAs like 5′ TOP mRNAs may have shorter poly(A)s in the steady state as they are more strongly protected by LARP1, contributing to the negative correlation between poly(A) length and mRNA stability.

Although this study expands our knowledge on the fundamentals of poly(A) length regulation, it also raises a number of interesting questions as well. Firstly, a complete list of the poly(A) barricade components should be revealed. Currently, it is not certain if there are players other than PABP and LARPs. Discovering other constituents or regulators would help understanding how the poly(A) barricade is regulated. Secondly, temporal binding of LARP1 and LARP4/4B should be examined to confirm the stage-specific formation of the different types of barri-

cades. The LARP1-bound and LARP4/4B-bound mRNAs have clearly distinct poly(A) profile apart from TOP binding specificity of LARP1, suggesting that those LARPs may act on the mRNAs in different stages of metabolic life cycle. Also, it remains unknown how the switching between LARP4/4B to LARP1 barricades can occur. Lastly, it should be tested how the closed-loop configuration that can be formed by the head-to-tail binding of LARP1 affects the actions of deadenylases or other decay factors. Thus, extensive follow-up studies are required to more comprehensively understand the dynamics of poly(A) length regulation.

# Methods and Materials

## Cell culture

HeLa and HEK293T cells were cultured in Dulbecco's modified Eagle medium supplemented with 9.1% fetal bovine serum. HCT116 cells were maintained in McCoy's 5A medium supplemented with 9.1% fetal bovine serum. All cell lines were authenticated using short tandem repeat (STR) profiling (PowerPlex 1.2; Promega) and results were compared with reference STR profiles available through the ATCC.

## siRNA transfection

For knockdown, cells were transfected twice with 50 nM of siRNAs for 3 days using Lipofectamine 3000 Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. We purchased siRNAs against PAN3, CCR4a/b, CAF1a/b, and PARN from Dharmacon (ON-TARGETplus SMARTpool). siRNA sequence for PAN2 knockdown is listed in Table S1. AccuTarget Negative Contol siRNA (Bioneer) was used for negative control siRNA. For plasmid transfection, HEK293T cells were transfected with 5 $\mu$g of DNA using Lipofectamine 3000 Reagent according to the manufacturer's instructions for 2 days.

## Quantitative RT-PCR

Total RNA was extracted using TRIzol (Thermo Fisher Scientific), treated with DNase I (TaKaRa), and purified using the RNeasy MinElute Cleanup Kit (QIAGEN). Purified RNA was reverse-transcribed by RevertAid Reverse Transcriptase (Thermo Fisher Scientific) using Random Primers (Thermo Fisher Scientific). Quantitative PCR was performed using the Power SYBR Green PCR Master Mix (Thermo Fisher Scientific). Primer sequences are listed in Table S1.

## Western blotting

Cells were lysed in 0.5% NP-40 lysis buffer (25 mM Tris pH 7.4, 100 mM KCl, 1 mM EDTA pH 8.0, and 0.5% NP-40) and sonicated for 5 min by Bioruptor (COSMO BIO). Protein samples (50 $\mu$g) were separated on Novex WedgeWell 10% Tris-Glycine Mini Gels (Thermo Fisher Scientific) and transferred to polyvinylidene fluoride membrane (GE Healthcare). Western blotting was performed as previously described (Lim et al., 2014). For western blotting, the following antibodies were used at 1:1,000 dilution in PBS containing 1% skim milk and 0.1% Tween 20 (USB): anti-GAPDH (Santa Cruz), anti-PAN2 (Sigma-Aldrich), anti-PARN (Cell Signaling), anti-CCR4a (Cell Signaling), anti-CAF1a (Abnova), anti-CAF1b (Lifespan Biosciences), anti-FLAG (Sigma-Aldrich), and anti-Tubulin (Abcam).

## Bulk poly(A) assay (BPA)

BPA was performed with slight modifications from a previous study (Schwede et al., 2008). Total RNA was extracted by TRIzol, treated with DNase I,

and purified using the RNeasy MinElute Cleanup Kit. Purified RNA (2 $\mu$g) was labelled with pCp bisphosphate 5$'$-$^{32}$P at the 3$'$ end by T4 RNA ligase 1 (NEB) for 1.5 h at 37$°$C and overnight at 4$°$C. Labeled RNA was purified by phenol extraction. Half of the purified RNA along with 20 $\mu$g of cold yeast tRNA (Roche) was treated with 40 $\mu$g RNase A (Sigma-Aldrich) and 5 U RNase T1 (Thermo Fisher Scientific) for 5 min at 37$°$C in RNase A/T1 buffer (10 mM Tris pH 7.5, 2 mM MgCl$_2$, 2 mM DTT, 0.2 mM ATP, 2% DMSO, 300 mM NaCl, and 5 mM EDTA). RNase A/T1-treated RNA was purified by phenol extraction. The RNA pellet after precipitation was eluted with 2xTBE-urea sample buffer (Biorad). Half of the RNA sample was loaded on a 10% polyacrylamide gel with 7 M urea. Radioactive signal intensities were obtained as described in IVD assay. To find the marker positions, the signal intensities from the marker lane were smoothened using the Savitzky-Golay filter. The window size and the order were subject to change depending on the data until the marker peaks were properly identified. After locating the markers, the signal intensities from the samples were also processed using the same filter (window size = 43, order = 3), and normalised by the total signal intensity in each lane, and multi- plied by 1000 to set to the arbitrary unit.

## Immunopurification

HEK293T cells grown on 100 mm dishes were collected 48 h after transfection of FLAG-tagged CCR4b and CAF1a expression plasmids. Cells were lysed in IVD lysis buffer (100 mM NaCl and 50 mM Tris pH 8.0) and sonicated for 5 min by Bioruptor followed by centrifugation for 15 min at 4$°$C. The supernatant was incubated with ANTI-FLAG M2 Affinity Gel (Sigma-Aldrich) with rotation for 1.5 h at 4$°$C (5 $\mu$L of FLAG beads

per 1 mg lysate). Incubated beads were washed 5 times with high salt wash buffer (800 mM NaCl and 50 mM Tris pH 8.0) and 5 times with TBS buffer (Sigma-Aldrich) (150 mM NaCl and 50 mM Tris pH 8.0). Supernatants were removed completely using a 30 and 1/2 gauge needle (BD Bioscience) and the beads were incubated with 2 mM 3x FLAG Peptide (Sigma-Aldrich) for 1 h at 4°C for elution. Eluted CNOT subcomplexes were confirmed by Coomassie blue staining (InstantBlue, Expedeon).

## High resolution poly(A) tail assay (Hire-PAT)

The same RNA samples used for mTAIL-Seq and bulk poly(A) assay were used. Hire-PAT and signal processing was performed as described previously (Chang et al., 2014). Sanger sequencing was done in parallel to identify the poly(A) site of each gene. Gene-specific forward PCR primer sequences are provided in Table S1.

## mTAIL-seq

mTAIL-Seq libraries were generated as described previously (Lim et al., 2016). In brief, 2 $\mu$g of total RNAs were ligated to the 3′ hairpin adaptor using T4 RNA Ligase 2 (NEB) and partially digested by RNase T1. 5′ end phosphorylation and endonucleolytic cleavage reaction using AP endonuclease 1 (NEB) were performed on the RNAs. RNAs were purified by Dynabeads M-280 Streptavidin, gel-purified for size fractionation (300–1000 nt), ligated to 5′ adaptor, reverse-transcribed, and amplified by PCR. The mTAIL-Seq libraries were sequenced with PhiX Control v3 (Illumina) and poly(A) spike-in mixture (Chang et al., 2014) by paired-end run (51+251 cycles) on Illumina MiSeq. The resulting sequencing data were preprocessed by Tailseeker v.3.1.6 (Chang & smaegol, 2017)

and subjected to in-depth analyses.

## N/C fractionation RNA-Seq

To infer the nucleus-cytoplasm abundance ratio, we used RNA-Seq datasets generated from ENCODE (ENCODE, 2012) experiments with HeLa S3 nuclear and cytoplasm fractions (ENCSR000CPP, ENCSR000CPQ). Transcripts were quantified and summarised as gene expression levels using RSEM v1.2.28 (Li & Dewey, 2011), along with Bowtie2 for alignment to the human genome (hg38). The gene expression levels of the biological replicates for each fraction were averaged by taking the means. For the genes that exceed 10 transcripts per million (TPM) were used for the analyses.

## Statistical test for elongated poly(A) tails

To identify genes with significantly lengthened poly(A) tails in the knockdown samples, we performed the one-tailed Mann-Whitney U test, which assesses how significantly the poly(A) length distributions of each gene changed toward elongation, with a pair of datasets, for the genes having more than 150 tags for 8 nt or longer poly(A)s in both samples. The *p*-values were adjusted using the Benjamini-Hochberg procedure. The effect sizes were estimated as described in (Fritz et al., 2012).

## *In vitro* deadenylation (IVD) assay

Synthesised RNA substrates (Lim et al., 2014) were labelled with radioisotope (ATP gamma-$^{32}$P) at the 5′ end and purified by phenol extraction. Labelled RNA substrates (0.45 nM) were pre-incubated with recombinant PABPC1 (OriGene) for 30 min at 37°C in IVD buffer (150

mM NaCl, 25 mM Tris pH 8.0, 3.2 mM $MgCl_2$, and 1 mM dithiothreitol (DTT)). Pre-incubated RNA-PABP samples were subsequently incubated with immunopurified CNOT subcomplexes at 37°C for designated time. The reaction solutions were mixed with 2xTBE-urea sample buffer (Bio-Rad) and treated with Proteinase K (Roche) for 15 min at 37°C and for 15 min at 50°C. Samples were boiled for 5 min at 95°C before RNA separation by 15% polyacrylamide gel electrophoresis with 7M urea. The gel was exposed to a phosphor imaging plate (Fujifilm) and radioactive signals were read using a Typhoon FLA 7000 (GE Healthcare). Intensity of RNA signals from the gel was quantified by Multi Gauge V3.0.

**Table S1** Oligonucleotide sequences

| Name | Sequence (5′-to-3′) |
| --- | --- |
| siPAN2 | GCTGCAGAATCACATACTA |
| siLARP1 | AGACUCAAGCCAGACAUCA |
| siLARP4-1 | AACAGAGGAAUCUUCUAUUAGAUCC |
| siLARP4-2 | CAUAAGCGUUGUAUUGUAA |
| siLARP5-1 | GCACAACAGGCUUACAAAU |
| siLARP5-2 | GCUAGUGACAUGUAUCUUA |
| CALM1-A0 | GCCUUUCAUCUCUAACUGCG |
| CALM1-A25 | GCCUUUCAUCUCUAACUGCG A25 |
| CALM1-A50 | GCCUUUCAUCUCUAACUGCG A50 |
| GAPDH-F (qRT-PCR) | CTCTCTGCTCCTCCTGTTCGAC |
| GAPDH-R (qRT-PCR) | TGAGCGATGTGGCTCGGCT |
| PARN-F (qRT-PCR) | TGTCCTGTCACGATTCCTGAG |
| PARN-R (qRT-PCR) | CCGGTACATGGCTCTAAATCCAA |
| PAN2-F (qRT-PCR) | CAGCAGCACTCTACTCGTTGG |
| PAN2-R (qRT-PCR) | GTGTGGCCGCAGAAGAAGAA |
| PAN3-F (qRT-PCR) | CAGCCCATGATCCTCTAACA |
| PAN3-R (qRT-PCR) | GCGATACTTCCTTGGCTTTC |
| CAF1a-F (qRT-PCR) | ATGCCAGCGGCAACTGTAG |
| CAF1a-R (qRT-PCR) | TCGGTGTCCATAGCAACGTAA |
| CAF1b-F (qRT-PCR) | GCTGACAGGAATGGCTTTCT |
| CAF1b-R (qRT-PCR) | GCCCACAGTACTTGGCATC |
| TMEM106C (Hire-PAT) | AACAGACCTAGTCAGGATATGAATTTGTTTC |
| RPSA (Hire-PAT) | GCCTATTCAGCAATTCCCTACTGAAG |
| SNRNP70 (Hire-PAT) | CCGGAGAATGGGTATTTGATGGAG |
| PABPC1 (Hire-PAT) | CCGAGCAAATGCCAGGTCTAGC |

# Bibliography

Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O., & Zenklusen, D. (2018). Spatial organization of single mrnps at different stages of the gene expression pathway. *Mol Cell*, 72, 1–12.

Aoki, K., Adachi, S., Homoto, M., Kusano, H., Koike, K., & Natsume, T. (2013). Larp1 specifically recognizes the 3' terminus of poly(a) mrna. *FEBS Lett*, 587(14), 2173–2178.

Aslam, A., Mittal, S., Koch, F., Andrau, J., & Winkler, G. (2009). The ccr4-not deadenylase subunits cnot7 and cnot8 have overlapping roles and modulate cell proliferation. *Mol Biol Cell*, 20(17), 3840–3850.

Bazzini, A., Lee, M., & Giraldez, A. (2012). Ribosome profiling shows that mir-430 reduces translation before causing mrna decay in zebrafish. *Science*, 336(6078), 233–237.

Bresson, S. & Conrad, N. (2013). The human nuclear poly(a)-binding protein promotes rna hyperadenylation and decay. *PLoS Genet*, 9(10), e1003893.

Cassar, P., Carpenedo, R., Samavarchi-Tehrani, P., Olsen, J., Park, C., Chang, W., Chen, Z., Choey, C., Delaney, S., Guo, H., Guo, H., Tanner, R., Perkins, T., Tenenbaum, S., Emili, A., Wrana, J., Gibbings, D., & Stanford, W. (2015). Integrative genomics positions mkrn1 as a novel ribonucleoprotein within the embryonic stem cell gene regulatory network. *EMBO Rep*, 16(10), 1334–1357.

Chang, H., Lim, J., Ha, M., & Kim, V. (2014). Tail-seq: genome-wide determination of poly(a) tail length and 3' end modifications. *Mol Cell*, 53(6), 1044–1052.

Chang, H. & smaegol (2017). *Tailseeker 3.1.7: the pipeline for high-throughput RNA poly(A) length and 3′ end modification measurement*.

Chen, C. & Shyu, A. (2011). Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA*, 2, 167–183.

Chen, C.-Y. A., Zhang, Y., Xiang, Y., Han, L., & Shyu, A.-B. (2017). Antagonistic actions of two human pan3 isoforms on global mrna turnover. *RNA*, 23(9), 1404–1418.

Christensen, A., Kahn, L., & Bourne, C. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. *Am J Anat*, 178(1), 1–10.

Copeland, P. & Wormington, M. (2001). The mechanism and regulation of deadenylation: identification and characterization of xenopus parn. *RNA*, 7(6), 875–886.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561–563.

Decker, C. & Parker, R. (1993). A turnover pathway for both stable and unstable mrnas in yeast: evidence for a requirement for deadenylation. *Genes Dev*, 7(8), 1632–1643.

Derry, M., Yanagiya, A., Martineau, Y., & Sonenberg, N. (2006). Regulation of poly(a)-binding protein through pabp-interacting proteins. *Cold Spring Harb Symp Quant Biol*, 71, 537–543.

ENCODE, P. C. (2012). An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414), 57–74.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., & Izaurralde, E. (2008). Deadenylation is a widespread effect of mirna regulation. *RNA*, 15(1), 21–32.

Fabian, M., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J., Svitkin, Y., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J., Chen, C., Shyu, A., Yates, J., Hannon, G., Filipowicz, W., Duchaine, T., & Sonenberg, N. (2009). Mammalian mirna risc recruits caf1 and pabp to affect pabp-dependent deadenylation. *Mol Cell*, 35(6), 868–880.

Fonseca, B., Zakaria, C., Jia, J., Graber, T., Svitkin, Y., Tahmasebi, S., Healy, D., Hoang, H., Jensen, J., Diao, I., Lussier, A., Dajadian, C., Padmanabhan, N., Wang, W., Matta-Camacho, E., Hearnden, J., Smith, E., Tsukumo, Y., Yanagiya, A., Morita, M., Petroulakis, E., González, J., Hernández, G., Alain, T., & Damgaard, C. (2015). La-related protein 1 (larp1) represses terminal oligopyrimidine (top) mrna translation downstream of mtor complex 1 (mtorc1). *J Biol Chem*, 290(26), 15996–16020.

Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen*, 141(1), 2–18.

Fukaya, T. & Tomari, Y. (2011). Pabp is not essential for microrna-mediated translational repression and deadenylation in vitro. *EMBO J*, 30(24), 4998–5009.

Funakoshi, Y., Doi, Y., Hosoda, N., Uchida, N., Osawa, M., Shimada, I., Tsujimoto, M., Suzuki, T., Katada, T., & Hoshino, S. (2007). Mechanism

of mrna deadenylation: evidence for a molecular interplay between translation termination factor erf3 and mrna deadenylases. *Genes Dev*, 21(23), 3135–3148.

Gallo, S., Ricciardi, S., Manfrini, N., Pesce, E., Oliveto, S., Calamita, P., Mancino, M., Maffioli, E., Moro, M., Crosti, M., Berno, V., Bombaci, M., Tedeschi, G., & Biffo, S. (2018). Rack1 specifically regulates translation through its binding to ribosomes. *Mol Cell Biol*.

Gentilella, A., Morón-Duran, F., Fuentes, P., Zweig-Rocha, G., Riaño-Canalias, F., Pelletier, J., Ruiz, M., Turón, G., Castaño, J., Tauler, A., Bueno, C., Menéndez, P., Kozma, S., & Thomas, G. (2017). Autogenous control of 5′top mrna stability by 40s ribosomes. *Mol Cell*, 67(1), 55–70.e4.

Goldstrohm, A. & Wickens, M. (2008). Multifunctional deadenylase complexes diversify mrna control. *Nat Rev Mol Cell Biol*, 9(4), 337–344.

Hong, S., Freeberg, M., Han, T., Kamath, A., Yao, Y., Fukuda, T., Suzuki, T., Kim, J., & Inoki, K. (2017). Larp1 functions as a molecular switch for mtorc1-mediated translation of an essential class of mrnas. *Elife*, 6.

Imataka, H., Gradi, A., & Sonenberg, N. (1998). A newly identified n-terminal amino acid sequence of human eif4g binds poly(a)-binding protein and functions in poly(a)-dependent translation. *EMBO J*, 17(24), 7480–7489.

Kahvejian, A., Roy, G., & Sonenberg, N. (2001). The mrna closed-loop model: the function of pabp and pabp-interacting proteins in mrna translation. *Cold Spring Harb Symp Quant Biol*, 66, 293–300.

Kahvejian, A., Svitkin, Y., Sukarieh, R., M'Boutchou, M., & Sonenberg, N.

(2005). Mammalian poly(a)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev*, 19(1), 104–113.

Kini, H., Silverman, I., Ji, X., Gregory, B., & Liebhaber, S. (2016). Cytoplasmic poly(a) binding protein-1 binds to genomically encoded sequences within mammalian mrnas. *RNA*, 22(1), 61–74.

Kircher, M., Stenzel, U., & Kelso, J. (2009). Improved base calling for the illumina genome analyzer using machine learning strategies. *Genome Biol*, 10(8), R83.

Klopfenstein, D., Zhang, L., Pedersen, B., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C., Yunes, J., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). Goatools: A python library for gene ontology analyses. *Sci Rep*, 8(1), 10872.

Kwon, S., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K., Castello, A., Krijgsveld, J., Hentze, M., & Kim, V. (2013). The rna-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol*, 20(9), 1122–1130.

Körner, C., Wormington, M., Muckenthaler, M., Schneider, S., Dehlin, E., & Wahle, E. (1998). The deadenylating nuclease (dan) is involved in poly(a) tail removal during the meiotic maturation of xenopus oocytes. *EMBO J*, 17(18), 5427–5437.

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., & Wahle, E. (2009). Poly(a) tail length is controlled by the nuclear poly(a)-binding protein regulating the interaction between poly(a) polymerase and the

cleavage and polyadenylation specificity factor. *J Biol Chem*, 284(34), 22803–22814.

Lahr, R., Fonseca, B., Ciotti, G., Al-Ashtal, H., Jia, J., Niklaus, M., Blagden, S., Alain, T., & Berman, A. (2017). La-related protein 1 (larp1) binds the mrna cap, blocking eif4f assembly on top mrnas. *Elife*, 6.

Lee, J., Lee, J., Trembly, J., Wilusz, J., Tian, B., & Wilusz, C. (2012). The parn deadenylase targets a discrete set of mrnas for decay and regulates cell motility in mouse myoblasts. *PLoS Genet*, 8, e1002901.

Li, B. & Dewey, C. (2011). Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323.

Lim, J., Ha, M., Chang, H., Kwon, S., Simanshu, D., Patel, D., & Kim, V. (2014). Uridylation by tut4 and tut7 marks mrna for degradation. *Cell*, 159(6), 1365–1376.

Lim, J., Kim, D., Lee, Y., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., & Kim, V. (2018). Mixed tailing by tent4a and tent4b shields mrna from rapid deadenylation. *Science*, 361(6403), 701–704.

Lim, J., Lee, M., Son, A., Chang, H., & Kim, V. (2016). mtail-seq reveals dynamic poly(a) tail regulation in oocyte-to-embryo development. *Genes Dev*, 30(14), 1671–1682.

Lima, S., Chipman, L., Nicholson, A., Chen, Y., Yee, B., Yeo, G., Coller, J., & Pasquinelli, A. (2017). Short poly(a) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol*, 24(12), 1057–1063.

Maraia, R., Mattijssen, S., Cruz-Gallardo, I., & Conte, M. (2017). The

la and related rna-binding proteins (larps): structures, functions, and evolving perspectives. *Wiley Interdiscip Rev RNA*, 8(6).

Mattijssen, S., Arimbasseri, A., Iben, J., Gaidamakov, S., Lee, J., Hafner, M., & Maraia, R. (2017). Larp4 mrna codon-trna match contributes to larp4 activity for ribosomal protein mrna poly(a) tail length protection. *Elife*, 6.

Miroci, H., Schob, C., Kindler, S., Ölschläger Schütt, J., Fehr, S., Jungenitz, T., Schwarzacher, S., Bagni, C., & Mohr, E. (2012). Makorin ring zinc finger protein 1 (mkrn1), a novel poly(a)-binding protein-interacting protein, stimulates translation in nerve cells. *J Biol Chem*, 287(2), 1322–1334.

Mittal, S., Aslam, A., Doidge, R., Medica, R., & Winkler, G. S. (2011). The ccr4a (cnot6) and ccr4b (cnot6l) deadenylase subunits of the human ccr4–not complex contribute to the prevention of cell death and senescence. *Molecular Biology of the Cell*, 22(6), 748–758.

Nousch, M., Techritz, N., Hampel, D., Millonigg, S., & Eckmann, C. (2013). The ccr4-not deadenylase complex constitutes the main poly(a) removal activity in c. elegans. *J Cell Sci*, 126(Pt 18), 4274–4285.

Park, J., Yi, H., Kim, Y., Chang, H., & Kim, V. (2016). Regulation of poly(a) tail and translation during the somatic cell cycle. *Mol Cell*, 62(3), 462–471.

Philippe, L., Vasseur, J., Debart, F., & Thoreen, C. (2018). La-related protein 1 (larp1) repression of top mrna translation is mediated through its

cap-binding domain and controlled by an adjacent regulatory region. *Nucleic Acids Res*, 46(3), 1457–1469.

Piao, X., Zhang, X., Wu, L., & Belasco, J. (2010). Ccr4-not deadenylates mrna associated with rna-induced silencing complexes in human cells. *Mol Cell Biol*, 30(6), 1486–1494.

Presnyak, V., Alhusaini, N., Chen, Y., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K., Graveley, B., & Coller, J. (2015). Codon optimality is a major determinant of mrna stability. *Cell*, 160(6), 1111–1124.

Roy, B. & Jacobson, A. (2013). The intimate relationships of mrna decay and translation. *Trends Genet*, 29(12), 691–699.

Sachs, A. & Davis, R. (1989). The poly(a) binding protein is required for poly(a) shortening and 60s ribosomal subunit-dependent translation initiation. *Cell*, 58(5), 857–867.

Schwede, A., Ellis, L., Luther, J., Carrington, M., Stoecklin, G., & Clayton, C. (2008). A role for caf1 in mrna deadenylation and decay in trypanosomes and human cells. *Nucleic Acids Res*, 36(10), 3374–3388.

Schäffler, K., Schulz, K., Hirmer, A., Wiesner, J., Grimm, M., Sickmann, A., & Fischer, U. (2010). A stimulatory role for the la-related protein 4b in translation. *RNA*, 16(8), 1488–1499.

Simón, E. & Séraphin, B. (2007). A specific role for the c-terminal region of the poly(a)-binding protein in mrna decay. *Nucleic Acids Res*, 35(18), 6017–6028.

Smith, B., Gallie, D., Le, H., & Hansma, P. (1997). Visualization of

poly(a)-binding protein complex formation with poly(a) rna using atomic force microscopy. *J Struct Biol*, 119(2), 109–117.

Subtelny, A., Eichhorn, S., Chen, G., Sive, H., & Bartel, D. (2014). Poly(a)-tail profiling reveals an embryonic switch in translational control. *Nature*, 508(7494), 66–71.

Tani, H., Mizutani, R., Salam, K., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., & Akimitsu, N. (2012). Genome-wide determination of rna stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res*, 22(5), 947–956.

Tarun, S. & Sachs, A. (1996). Association of the yeast poly(a) tail binding protein with translation initiation factor eif-4g. *EMBO J*, 15(24), 7168–7177.

Temme, C., Zaessinger, S., Meyer, S., Simonelig, M., & Wahle, E. (2004). A complex containing the ccr4 and caf1 proteins is involved in mrna deadenylation in drosophila. *EMBO J*, 23(14), 2862–2871.

Thomas, M., Liu, X., Whangbo, J., McCrossan, G., Sanborn, K., Basar, E., Walch, M., & Lieberman, J. (2015). Apoptosis triggers specific, rapid, and global mrna decay with 3' uridylated intermediates degraded by dis3l2. *Cell Rep*, 11(7), 1079–1089.

Tucker, M., Staples, R., Valencia-Sanchez, M., Muhlrad, D., & Parker, R. (2002). Ccr4p is the catalytic subunit of a ccr4p/pop2p/notp mrna deadenylase complex in saccharomyces cerevisiae. *EMBO J*, 21(6), 1427–1436.

Tucker, M., Valencia-Sanchez, M., Staples, R., Chen, J., Denis, C., & Parker, R. (2001). The transcription factor associated ccr4 and caf1

proteins are components of the major cytoplasmic mrna deadenylase in saccharomyces cerevisiae. *Cell*, 104(3), 377–386.

Viswanathan, P., Ohn, T., Chiang, Y., Chen, J., & Denis, C. (2004). Mouse caf1 can function as a processive deadenylase/3'-5'-exonuclease in vitro but in yeast the deadenylase function of caf1 is not required for mrna poly(a) removal. *J Biol Chem*, 279(23), 23988–23995.

Wang, Z., Day, N., Trifillis, P., & Kiledjian, M. (1999). An mrna stability complex functions with poly(a)-binding protein to stabilize mrna in vitro. *Mol Cell Biol*, 19(7), 4552–4560.

Wells, S., Hillner, P., Vale, R., & Sachs, A. (1998). Circularization of mrna by eukaryotic translation initiation factors. *Mol Cell*, 2(1), 135–140.

Wigington, C., Williams, K., Meers, M., Bassell, G., & Corbett, A. (2014). Poly(a) rna-binding proteins and polyadenosine rna: new members and novel functions. *Wiley Interdiscip Rev RNA*, 5, 601–622.

Winata, C. & Korzh, V. (2018). The translational regulation of maternal mrnas in time and space. *FEBS Lett*, 592(17), 3007–3023.

Yamashita, A., Chang, T., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C., & Shyu, A. (2005). Concerted action of poly(a) nucleases and decapping enzyme in mammalian mrna turnover. *Nat Struct Mol Biol*, 12, 1054–1063.

Yamashita, R., Suzuki, Y., Takeuchi, N., Wakaguri, H., Ueda, T., Sugano, S., & Nakai, K. (2008). Comprehensive detection of human terminal oligo-pyrimidine (top) genes and analysis of their characteristics. *Nucleic Acids Res*, 36(11), 3707–3715.

Yang, R., Gaidamakov, S., Xie, J., Lee, J., Martino, L., Kozlov, G., Craw-

ford, A., Russo, A., Conte, M., Gehring, K., & Maraia, R. (2011). La-related protein 4 binds poly(a), interacts with the poly(a)-binding protein mlle domain via a variant pam2w motif, and can promote mrna stability. *Mol Cell Biol*, 31(3), 542–556.

Yao, G., Chiang, Y., Zhang, C., Lee, D., Laue, T., & Denis, C. (2007). Pab1 self-association precludes its binding to poly(a), thereby accelerating ccr4 deadenylation in vivo. *Mol Cell Biol*, 27(17), 6243–6253.

# 국문초록

아데닌중합체형성은 거의 모든 진핵생물의 mRNA의 3′ 말단에서 일어난다. 그 결과 만들어지는 poly(A) 꼬리는 poly(A)결합 단백질들과 단단히 결합하여 mRNA의 부적절한 분해를 막고 번역을 촉진하는 등 mRNA 기능 수행을 위한 필수적인 요소로서 작용한다. 때문에 mRNA를 분해하기 위해서는 poly(A) 꼬리가 탈아데닌화를 통해 가장 먼저 분해되어야 한다. 이 때, 탈아데닌화는 일반적인 mRNA 분해 과정에서 종종 속도 결정 단계로서 작용하기 때문에, poly(A) 꼬리는 microRNA를 통한 유전자 발현억제, AU-rich 인자를 통한 mRNA 분해유도, 모계-접합체 전이 등을 포함한 다양한 전사 후 조절 기전에서 핵심적인 중추로서 기능한다.

그러나 유전자 발현 조절에서의 중요성에도 불구하고, 탈아데닌화와 관련된 생물학적 발견들은 기술적 한계, 특히 전사체 수준에서 poly(A) 서열을 분석하기 위한 기술의 부재로 인해 지체되어 왔다. 이에, 여기서 나는 RNA간섭 기술과 최근 개발된 전사체 수준의 poly(A) 서열 분석 기술들을 접목하여 인간 탈아데닌화 효소들의 역할과 기질 특이성에 대한 연구를 수행하였다. 첫째로, 나는 널리 받아들여져 왔던 탈아데닌화 모델인 두 단계 탈아데닌화 모델(the 'biphasic deadenylation' model)이 세포 내 거의 모든 mRNA에 적용될 수 있음을 확인하였다. 즉, PAN2-PAN3 (PAN2/3) 복합체가 먼저 아주 긴 poly(A) 꼬리를 다듬고 CCR4-NOT (CNOT) 복합체가 뒤를 이어 탈아데닌화를 마무리 한다. 그러나, 기존에 제시되었던 모델과는 달리, PAN2/3 복합체의 양을 현저히 낮췄을 때에도 전사체엔 별 영향이 없는 것으로 보아 CNOT 복합체가 PAN2/3 역할을 상당 부분 보완할 수 있으며, 따라서 PAN2/3 복합체의 작용이 CNOT 복합체가 작용하기 위한 선행 조건

은 아님을 추가로 밝혀냈다. 또한, 통계적 분석을 통해 CNOT 복합체가 없을 경우 거의 모든 mRNA의 꼬리가 길어지며, 길어지는 정도는 mRNA가 얼마만큼 세포질에 위치하는 지와 양의 상관관계가 있음을 알았다. 따라서, 나는 CNOT 복합체가 가장 지배적이고 비특이적인 세포질 탈아데닌화 효소임을 확고히 하였다.

다음으로, 나는 평형 상태에서의 poly(A) 길이 분포에 대하여 더 자세히 연구하였다. 나는 평형 상태의 poly(A) 길이 분포를 이용한 클러스터 분석을 통해 유전자들이 몇 개의 특징적인 집단으로 나뉠 수 있음을 알아냈다. 평형 상태 poly(A) 길이는 mRNA의 발현량, 안정성, 3′ 미번역 부위의 길이, 핵 대 세포질에서의 양 비율, 번역 효율 등 mRNA의 여러 특징들과 강하게 연관되어 있었다. 반직관적이게도, 안정한 mRNA일수록 더 짧은 poly(A) 꼬리를 갖는 현상을 발견하였는데, 이는 안정한 mRNA의 탈아데닌화가 이후에 수반되는 분해과정들과 분리되어 있음을 암시한다. 나는 이 역설적인 관계를 설명하기 위하여 완전한 탈아데닌화가 일어나는 것을 막아 결과적으로 꼬리가 짧고 안정한 mRNA를 만들어 내는 poly(A) barricade가 존재함을 상정하였다.

Poly(A) barricade를 찾아내기 위하여, 나는 PABPC1이 poly(A) barricade의 구조적 기반을 제공한다는 가정 하에 이와 상호작용하는 단백질들을 우선 찾아보았다. PABPC1과 함께 면역침강하는 단백질들을 액체 크로마토그래피 및 질량 분석법(LC-MS/MS)을 이용하여 분석함으로써 PABPC1과 상호작용하는 단백질들을 찾아내었다. 이들 중에 LARP1, 그리고 LARP4/4B가 가장 두드러진 poly(A) barricade 구성 요소 후보로 등장하였다. 각 후보 단백질에 결합하는 RNA들, 그리고 각 후보를 RNA간섭을 이용하여 저해하였을 때 poly(A)길이가 바뀌는 RNA들을 전사체 수준에서 분석하였는데, 이를 통해 LARP1은 약 30–60 nt 길이의 poly(A) 꼬리에 붙어 이들을 보호하며 PABPC1의 poly(A) 결합 양상과 닮은 주기적 결합 패턴을 형성하는

것을 확인하였다. 또한, LARP1은 5′ 말단 올리고피리미딘 모티프 (5′ termi-nal oligopyrimidine motif)를 가지는 리보솜 단백질 mRNA들에 대한 강한 선호가 있음을 확인하였으며, 이는 리보솜 단백질 mRNA를 위시한 house-keeping mRNA들이 왜 더 안정하지만 더 짧은 poly(A) 꼬리를 갖는지에 대한 부분적인 설명을 제공한다. 이에 더하여, 나는 일련의 *in vitro* 실험을 통해 LARP1 단백질이 poly(A) 결합 능력과 deadenylation을 저해하는 활성을 가짐을 확인하였으며, 이는 LARP1이 poly(A) barricade의 핵심 기능을 담당하는 구성 요소임을 보여준다. 한편, LARP4/4B는 길이가 약 70–190 nt에 이르는 긴 poly(A) 꼬리와 비특이적으로 결합하며, 따라서 LARP1과는 다른 대사과정 단계에 있는 mRNA에 작용한다 생각된다.

나는 이 학위논문에서 전사체 수준의 분석을 통한 poly(A) 길이 조절에 관하여 연구하였다. 이 연구는 두 단계 탈아데닌화 모델을 검증하고 수정함과 동시에 CNOT 복합체가 세포질에서의 비특이적이고 광범위한 탈아데닌화를 담당하는 효소임을 확고히 하였다. 또한, 안정한 mRNA가 더 짧은 poly(A) 꼬리를 갖는 역설적 현상을 설명하기 위해, 탈아데닌화를 저해함으로써 탈아데닌화와 그 이후 수반되는 RNA 분해 과정을 분리시키는 poly(A) barricade의 존재를 가정하고, LARP1이 짧은 poly(A) 특이적인 탈아데닌화를 저해하는 poly(A) barricade의 핵심 구성 요소임을 밝혔다. Poly(A) bar-ricade의 발견은 새로운 poly(A) 길이 조절 인자의 기능과 역할을 규명함으로써 유전자 발현 조절에 대한 이해의 폭을 근본적으로 넓히는 데 일조한다.