

GLOBAL INVESTIGATION OF POLY(A) TAIL LENGTH DYNAMICS DURING
MACROPHAGE ACTIVATION

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Yeonui Kwak

May 2022

© 2022 Yeonui Kwak

GLOBAL INVESTIGATION OF POLY(A) TAIL LENGTH DYNAMICS DURING MACROPHAGE ACTIVATION

Yeonui Kwak, Ph. D.

Cornell University 2022

Post-transcriptional regulation plays important roles in spatial-temporal dynamics of gene expression by controlling mRNA stability, translation efficiency, and mRNA localization. In metazoans, poly(A) tail length control plays crucial roles in almost every aspect of post-transcriptional mRNA regulation, and underlies development, normal homeostasis, and diseases. Most systematic, genome-wide investigations of poly(A) tail length control have been limited to specific biological contexts, such as oocyte fertilization. The absence of zygotic transcription makes oocytes a tractable system to examine changes in poly(A) tail lengths without the confounding influence of new transcripts. However, somatic systems are more challenging to monitor post-transcriptional poly(A) tail length regulation, since new transcripts with longer tails continuously enter the mRNA pool. Therefore, most examples of poly(A) tail length regulation in non-developmental systems have only been shown with a handful of genes in some specific biological contexts. The complexity, relevance and widespread nature of poly(A) tail dynamics are largely unknown for post-embryonic cellular processes.

In this thesis, I integrated multiple transcriptomic approaches for exploring post-transcriptional poly(A) tail dynamics in post-embryonic systems. By examining mRNA abundance, nascent transcription, and poly(A) tail length across a time course of macrophage activation, a period of widespread and dynamic changes in the gene

expression program, I found that a large fraction of the transcriptome underwent changes in poly(A) tail length, including transient increases for pro-inflammatory genes, with distinct patterns of changes in other sets of genes. Increases in tail length correlated with an increase in mRNA levels regardless of transcriptional activity, and many mRNAs that underwent tail extension encode proteins necessary for immune function and post-transcriptional regulation. Our analyses indicate that many mRNAs undergoing tail lengthening are, in turn, degraded by elevated levels of ZFP36, constituting a post-transcriptional feedback loop that ensures transient regulation of transcripts integral to macrophage activation. Collectively, my thesis work introduces an analytic framework to study post-transcriptional control of poly(A) tail length in transcriptionally active, cellular processes and provides evidence that readenylation can be widely used, exerting a profound effect on gene expression in a non-developmental context.

BIOGRAPHICAL SKETCH

Yeonui Kwak grew up in Seoul, the capital of South Korea. She earned a B.S. degree in Life Science and Biotechnology at Korea University, Seoul. After graduating from Korea University, she entered graduate school in the same university to expose herself to research in the field of biochemistry and molecular biology, where she characterized the selectivity and efficacy of small-molecule compounds in drug-resistant human cancers from many different tissues of origin under the guidance of Dr. Taebo Sim. In the late stage of her master's degree, she had a chance to perform transcriptome profiling in human endometrial cancer cells. She was impressed by its ability to examine gene expression across thousands of genes at once and took a great interest in exploring the transcriptomic data for a new biological discovery, which was the moment that she decided to deepen her knowledge in genomics and RNA biology. She came to Cornell University to pursue her Ph.D. in Genetics, Genomics, and Development in 2016. During her graduate studies, she has developed her skills in next-generation sequencing techniques and expanded her knowledge in RNA biology under the guidance of Dr. Andrew Grimson. Especially, she found herself enjoying interpreting large-scale genomic data, and beautifully visualizing the analytic results, thus turning herself into a computational biologist. She hopes to apply what she learned during her Ph.D. in uncovering the molecular mechanisms underlying human diseases. Now, she looks forward to embarking on a new chapter of her research career as a computational biologist.

To the loves of my life:

The folks who raised and supported me, and who are my reasons for being –
my family

ACKNOWLEDGMENTS

First and foremost, I gratefully acknowledge my thesis supervisor, Dr. Andrew Grimson. I would like to thank him for guiding me through the years and creating such a wonderful environment in the lab. He has been such a great mentor to me that I couldn't ask for more. Also, I would like to thank him for opening many doors for me every step of my Ph.D. journey. I truly think that I wouldn't have been able to make it this far without his guidance and optimism.

I would also like to thank my thesis advisory committee members, including Dr. Hojoong Kwak, Dr. Charles Danko, and Dr. Praveen Sethupathy. I owe a lot of my progress to them giving me good advice and encouragements for my thesis project.

Jessica is my amazing colleague, true friend and mini advisor that I cannot thank enough. I learned a lot from her about 3'UTR-mediated regulation. She also helped me find comfort in the lab. Without her on my side, I wouldn't have been able to make it through this far. She was also a huge help in writing a paper and preparing a research talk. I was lucky that I could meet such a good friend and brilliant colleague during my Ph.D. study.

I would like to thank Ciaran for always giving me great advice and support when addressing major revisions. He was one of the folks that deeply understood my research. Without his professional help, I could not have gotten the work published to the world. Also, He gave many good suggestions to me in writing manuscripts and giving a research talk.

I also want to thank Elizabeth Fogarty. Liz has taken care of all kinds of stuffs around the lab. Also, she was such a professional help when we were doing a reporter assay in response to a reviewer's request during the revision. Without her effort and time, I seriously think that my thesis work would not have made through the publication. Liz also related to my personal circumstances, comforting me.

I would like to thank all the members of the Grimson lab, including Divya, Hongya, Ya, David, Alexis, Louis, Carter, and Jason, They are amazing, smart, and friendly, making the lab fun and intellectually pleasing. I also thank all my collaborators including Sibylle at Memorial Sloan Kettering Cancer Center, and all the CFS project folks at Cornell: Jen, Faraz, and Paul.

Also, I would like to thank my GGD cohort friends: Dashiell, Dawn, Hui, Miwa, and Sylvia. For the past 6 years, close connection with these people would hold me to the ground so that I wouldn't run away being scared off. I am grateful that I could meet such good friends in my Ph.D. time.

Lastly, I want to thank my friends who would lift me up outside of the program: HeeJin, Dahlm, Jahyoung, Eunhee, Sarah, Hyunseung, Byeongdoo, Woohyung, Jiwon, Junho, Yooshin, Jaeun, and others.

I would like to thank Cornell, the Department of Molecular Biology and Genetics, and the Field of Genetics, Genomics, and Development for giving me the opportunity of being a part of them, which was such a great honor and fortune.

Lastly, and most importantly, I can't thank enough my family for supporting me

through the dramatic 6 years. Everything that I've achieved is only because of the love and support that you gave me. I also deeply thank my grandmother who always has been my inspiration and embodiment of eternal and infinite love.

TABLE OF CONTENTS

| | |
|--|-------------|
| BIOGRAPHICAL SKETCH | iii |
| ACKNOWLEDGEMENTS | v |
| TABLE OF CONTENTS | viii |
| LIST OF FIGURES | xi |
| CHAPTER 1: Introduction | 1 |
| 1.1 Polyadenylation in nucleus..... | 1 |
| 1.2 Factors regulating polyadenylation dynamics..... | 3 |
| 1.3 Role of poly(A) tail in the cytoplasm..... | 4 |
| 1.4 Significance of poly(A) tail length control in post-transcriptional regulation..... | 5 |
| 1.4.1 mRNA stability..... | 5 |
| 1.4.2 Translation..... | 6 |
| 1.4.3 mRNA localization and local translation..... | 7 |
| 1.5 Factors regulating poly(A) tail length in the cytoplasm..... | 8 |
| 1.6 Mechanisms modulating poly(A) tail length regulation during a cellular response..... | 10 |
| 1.7 Relevance of poly(A) tail length regulation in human diseases..... | 13 |
| 1.8 Existing tools to study poly(A) tail length..... | 15 |
| 1.9 Current methodology and its limitation..... | 17 |
| 1.10 Regulation of poly(A) tail in immunity..... | 19 |
| FIGURES | 20 |
| REFERENCES | 24 |
| CHAPTER 2: Dynamic and Widespread Control of Poly(A) Tail Length during Macrophage Activation | 32 |
| 2.1 Abstract..... | 32 |
| 2.2 Introduction..... | 32 |

| | |
|--|------------|
| 2.3 Results..... | 36 |
| 2.3.1 Determination of isoform-specific poly(A) tail lengths in THP-1 cells..... | 36 |
| 2.3.2 PAL Dynamics during Macrophage Activation..... | 39 |
| 2.3.3 3'UTR sequence features associated with changes in PAL..... | 41 |
| 2.3.4 Poly(A) tail length correlate with post-transcriptional changes..... | 44 |
| 2.3.5 Profiling readenylation during initiation of the macrophage immune response..... | 48 |
| 2.3.6 Concomitant readenylation of ZFP36 and its target mRNAs upon early activation..... | 51 |
| 2.4 Discussion..... | 55 |
| 2.4.1 Poly(A) tail length dynamics upon macrophage activation..... | 55 |
| 2.4.2 Basis for rapid and transient poly(A) tail lengthening..... | 57 |
| 2.5 Materials and Methods..... | 61 |
| FIGURES..... | 82 |
| REFERENCES..... | 123 |
| CHAPTER 3: Genome-Wide Identification of Polyadenylation Dynamics with TED-Seq..... | 132 |
| 3.1 Abstract..... | 132 |
| 3.2 Introduction..... | 132 |
| 3.3 Materials..... | 136 |
| 3.4 Methods..... | 141 |
| 3.5 Notes..... | 148 |
| FIGURES..... | 153 |
| REFERENCES..... | 158 |
| CHAPTER 4: Conclusions and Future Directions..... | 161 |

| | |
|--|------------|
| 4.1 Conclusions..... | 161 |
| 4.2 Future Directions..... | 164 |
| REFERENCES..... | 173 |
| APPENDIX I: Relationship between Poly(A) Tail Length and 3'UTR Isoform Identity and Dynamic Regulation of Alternative 3'UTR Isoform Expression during Macrophage Activation | 176 |

LIST OF FIGURES

CHAPTER 1

Figure 1.1 Two major types of mRNA isoforms generated using alternative polyadenylation sites.

Figure 1.2 Major decay pathways of eukaryotic mRNAs in the cytoplasm.

Figure 1.3 Poly(A) tail length control for TNF mRNAs during macrophage activation.

CHAPTER 2

Figure 2.1 Determination of PAL with isoform specificity.

Figure 2.2 PAL dynamics during macrophage activation.

Figure 2.3 RNA features contributing to PAL control during the macrophage immune response.

Figure 2.4 Association between changes in PAL and mRNA abundance regardless of transcriptional changes.

Figure 2.5 Widespread readenylation during macrophage activation.

Figure 2.6 ZFP36 mRNAs are readenylated through poly(U)-containing 3'UTR elements upon macrophage activation.

Figure 2.7 Concomitant readenylation of ZFP36 and its target mRNAs upon macrophage activation.

Figure S2.1 Reproducibility of TED-seq and 3'-seq, and association of poly(A) tail length with 3'UTR isoform usage; related to Figure 2.1.

Figure S2.2 Analysis of 3'UTR isoform switching and measurements of poly(A) tail length change; related to Figure 2.2.

Figure S2.3 Analysis of sequence elements associated with poly(A) tail length changes; related to Figure 2.2 and Figure 2.3.

Figure S2.4 Association between transcriptional changes and changes in poly(A) tail

lengths; related to Figure 2.3 and Figure 2.4.

Figure S2.5 Association between changes in poly(A) tail lengths and changes in mRNA abundance; related to Figure 2.4.

Figure S2.6 Analysis of readenylation upon LPS stimulation; related to Figure 2.5.

Figure S2.7 Analysis of readenylation upon LPS stimulation; related to Figure 2.5 and Figure 2.6.

CHAPTER 3

Figure 3.1 Schematics of poly(A) tail length calculation in TED-seq.

Figure 3.2 Overview of TED-seq experimental procedures shows all the experimental steps in Subheading 3.4.

Figure 3.3 Electrophoresis for the size selection. Shown is a post-excision polyacrylamide gel.

Figure 3.4 DNA elution from a polyacrylamide gel.

Figure 3.5 Example of a TED-seq result in a genome browser.

CHAPTER 1

INTRODUCTION

Although transcription is an essential first step, and certainly the most studied area in gene expression regulation, additional regulations occur post-transcriptionally throughout mRNA lifecycle, contributing to fine-tuning of gene expression (Carpenter et al. 2014). Following transcription, pre-mRNA intronic sequences are removed by splicing, and a poly(A) tail is added to the 3' end of transcripts (Passmore and Collier 2022; Edmonds and Abrams 1960). The poly(A) tail has instrumental roles in almost every aspect of post-transcriptional mRNA regulation, influencing nuclear export, translation, mRNA localization, and mRNA decay (Jalkanen et al. 2014). Thus, a full appreciation of the polyadenylation process and dynamic regulation of poly(A) tail length is paramount for our understanding of normal physiological processes and diseases. This chapter summarizes various mechanistic aspects of poly(A) tail, including its biogenesis, function, and regulation, as well as its biological importance.

1.1 Polyadenylation in the nucleus

Poly(A) tail is a non-templated addition of multiple adenosines at the 3' end of almost every eukaryotic mRNA, with the only known exception being some mammalian histone transcripts. Poly(A) tails are added co-transcriptionally upon 3' cleavage of nascent mRNA and are required for the export of mature mRNAs to the cytoplasm, and play important roles in mRNA translation and stability. Importantly, the majority of human genes have multiple cleavage and polyadenylation sites, which can be alternatively used, generating multiple mRNA isoforms of different protein-coding or regulatory potential (Elkon et al. 2013; Di Giammartino et al. 2011; Tian and Manley 2017). For example, alternative polyadenylation (APA) within the last exon can

generate mRNA isoforms with different 3'UTR lengths, while APA occurring in alternative last exons can lead to mRNA isoform with different protein-coding sequences along with entirely distinct 3'UTRs (**Figure 1.1**). Particularly, 3'UTRs serve as interaction hubs for a variety of RNA-binding regulatory factors (e.g. microRNAs), and thus the differences in 3'UTR length can make a substantial difference in mRNA fates encompassing translation, mRNA stability, and subcellular mRNA transport. Therefore, alternative polyadenylation is a crucial regulatory mechanism that dramatically modulate gene expression.

The process of cleavage/polyadenylation is regulated by specific sequences in the 3' end of unprocessed transcripts. The most important *cis* element is a poly(A) signal (PAS) hexamer located 10-30 nucleotides (nt) upstream of the poly(A) cleavage site (Zhao et al. 1999). Although the canonical hexamers AAUAAA and AUUAAA are most frequently used in vertebrates, multiple sequencing-based analyses identified 16 other variant poly(A) signal sequences (Beaudoing et al. 2000). The trans-acting factors necessary for cleavage and polyadenylation in mammals include CPSF (cleavage and polyadenylation specificity factor), CstF (cleavage stimulatory factor), CFIm and CFIIm (mammalian cleavage factors I and II, respectively), and PAP (polyA polymerase). CPSF binds the PAS hexamer, while the auxiliary motif UGUA, frequently found upstream of the poly(A) site, is preferentially bound by cleavage factor I complex consisting of CFIm proteins (Di Giammartino et al. 2011). The cleavage stimulation factors (CstF proteins) bind to a U- or GU-rich downstream sequence (DGE) of the cleavage site. These auxiliary elements work in concert with the PAS hexamer to determine the precise site of cleavage, and contribute to the strength of poly(A) sites, the efficiency in which a poly(A) site is recognized by 3' end processing machinery for cleavage and polyadenylation. For example, the canonical hexamers have a stronger affinity for the common cleavage factors than the non-

canonical hexamers. Therefore, when a canonical and a non-canonical PAS hexamers are located within the same transcription unit, the canonical hexamer is predominantly used for mRNA synthesis. Finally, endonucleolytic cleavage at the 3' ends is followed by the addition of poly(A) tail at the cleaved site by Poly(A) polymerase (PAP) with an average length of ~200 nt in mammals and ~70 nt in yeast. In mammals, PAP adds an initial tail (11~14 nt) to the cleaved 3' end of mRNAs, which is occupied by nuclear poly(A)-binding protein (PABPN1) to allow rapid addition of adenosine residues until the tail is about 200-250 nt in length (Eckmann et al. 2011). The resulting poly(A) tail facilitates the export of mature mRNAs to the cytoplasm through the interaction with PABPN1, a shuttling protein that moves between the nucleus and the cytosol, although the underlying mechanism remains elusive.

1.2 Factors regulating polyadenylation dynamics

Several polyadenylation studies have shown that not all transcripts are 'fully' polyadenylated to ~250 adenosines in all tissues. For example, sequence elements that limit the initial length of the poly(A) tail on nascent mRNAs were identified with the most notable example being a poly(A) limiting element (PLE) (Gu et al. 1999). The PLE element was first discovered in albumin mRNA from *Xenopus*. In contrast to most eukaryotic mRNAs, *Xenopus* albumin transcripts are added with very short poly(A) tails (less than 20 nt) during nuclear polyadenylation, and they do not show a notable difference in translation and mRNA stability compared to typical mRNAs. The PLE is composed of a pyrimidine-rich region followed by an AG dinucleotide located in the last exon of *Xenopus* albumin, and also found in many other transcripts across species to confer a short initial tail, although the molecular mechanisms underlying this process is unclear. Several trans-acting factors, including PABPN1 and CPSF, were found to be involved in controlling the length of poly(A) tail added during nuclear polyadenylation.

In addition, other additional factors, including NPM1, ZC3H14, ZFP36, and Hu proteins, were identified to engage in determining initial poly(A) tail length using a knockdown system. There are likely numerous other cis-acting elements and RNA-binding factors that influence polyadenylation dynamics that remain to be uncovered.

1.3 Role of poly(A) tail in the cytoplasm

Once in the cytoplasm, the poly(A) tail is predominantly coated with the cytoplasmic poly(A) binding protein (PABPC) (Baer and Kornberg 1983; Blobel 1973). Protein composition on a newly synthesized transcript is quite different from that of an actively translating mRNA in the cytoplasm and remodeling of many proteins must take place; for the trade-off between PABPN and PABPC, the first round of translation called a pioneer round of translation seems to promote this transformation (Sato and Maquat 2009). The exchange between PABPN and PABPC could also be influenced by nuclear export or through passive remodeling in the cytoplasm.

In the cytoplasm, the poly(A) tail enhances mRNA translation and protects mRNA from degradation through binding to PABPC. Specifically, the poly(A) tail can function synergistically with the 7-methylguanosine (m7G) cap on the 5' end of the mRNA; PABP binds to the poly(A) tail while eIF4G binds to the 5' cap structure, and the interaction of PABP and eIF4G brings the cap and the tail into close proximity to form a “closed loop” (Vicens et al. 2018), which is thought to facilitate translation through ribosome recycling and protect mRNA from degradation by preventing the poly(A) tail and 5' cap from being accessed by mRNA decay machinery (Gallie 1991). Previous studies showed that the minimum poly(A) tail length required to confer stability and translational increase is about 30 nt, which is equivalent to the footprint of PABPC (the area occupied by PABPC), and that mRNA with the poly(A) tail shorter than 16 nt cannot be translated, suggesting that PABPC binding is required

to mediate the poly(A) tail effect in the cytosol. However, recent studies showed that the role of PABPC is not definitive, since PABPC can also recruit deadenylase complexes PAN2/3 and CCR4-NOT complex, which shorten poly(A) tail (Uchida et al. 2004). In general, shortening of poly(A) tail below a certain threshold tail length leads to dissociation of PABP from the tail, which subsequently causes disruption of the “closed loop” structure of the mRNA leading to mRNA decay and translational repression.

1.4 Significance of poly(A) tail length control in post-transcriptional regulation.

Although poly(A) tail length is initially determined during poly(A) tail synthesis in the nucleus, virtually every mRNA has its tail remodeled post-transcriptionally in the cytoplasm, which significantly affects gene expression at multiple aspects of a transcript’s life cycle: mRNA stability, translation, and mRNA localization.

1.4.1 mRNA stability

mRNA degradation, together with transcription, determines the cellular mRNA abundance. Recently, metabolic RNA labelling of new transcripts coupled with pulse-chase approach revealed that different transcripts in the same cell vary significantly in their stability up to ~1000 fold (Geisberg et al. 2014; Eisen et al. 2020b). In mammalian cells, individual mRNA molecules can survive from only a few minutes to even several days. Importantly, after nuclear export, nearly all mRNAs undergo tail shortening (deadenylation) at the 3’ends as the first and rate-limiting step of mRNA decay with CCR4-NOT complex as the predominant deadenylase acting on all mRNAs in all biological systems. Deadenylation is followed by one of two alternate degradation pathways (**Figure 1.2**): 3’→5’ degradation by the cytoplasmic exosome, or removal of the 5’ cap (decapping) by Dcp1-Dcp2 complex and the subsequent

5'→3' degradation by the Xrn1 exonuclease (Schoenberg and Maquat 2012).

Deadenylation rates are differently controlled across the transcriptome, contributing to the regulation of the steady-state level of mRNAs and as a consequence protein outputs.

1.4.2 Translation

While mRNA abundance can explain the majority of variation in protein level, more than 30% of the variation are not explained by mRNA abundance alone, indicating significance of the regulation at the translation level. Translation is predominantly regulated at the initiation step. Especially, the poly(A) tail play an important role in enhancing translation in cooperation with a 5' cap structure of mRNA during translation initiation (Jacobson and Favreau 1983; Subtelny et al. 2014; Gallie 1991). Many biological responses often require rapid and dynamic changes in gene expression programs. Especially, when transcription is inactive, or transcriptional modulation is not fast enough to cause immediate changes in protein output, translational control on pre-existing mRNAs can make a rapid change in gene expression. Although deadenylation is a major tail length regulation occurring in the cytoplasm as a part of mRNA decay pathway, in some rare instances, deadenylation was shown to cause translational repression without triggering mRNA decay, where these mRNAs can be recycled into active translation through the tail extension during a cellular response. Such poly(A) tail-dependent control of translation was well described in oocyte maturation (Xiang and Bartel 2021; Eichhorn et al. 2016), early embryogenesis (Eichhorn et al. 2016; Subtelny et al. 2014), and other cellular response requiring immediate change, such as innate immune responses (Crawford et al. 1997) or synaptic stimulation (Goldstrohm and Wickens 2008). For example, in oocytes of various species, including frog (Xiang and Bartel 2021), *Drosophila* (Lim et al. 2016),

and sea urchin (Wilt 1973), many maternal mRNAs have been shown to exist with short poly(A) tails, and translationally repressed. During oocyte maturation or shortly upon fertilization, poly(A) tail lengthening occurs for the stored maternal mRNAs, which leads to translational entry for the transcripts (Lim et al. 2016; Slater et al. 1972; Wilt 1973). Although translational regulation is essential for the rapid control of protein output in various biological responses, the involvement of poly(A) tail control has only been shown in early developmental contexts on a global scale, and has been largely uncharacterized in non-developmental contexts with only a handful of genes validated.

1.4.3 mRNA localization and local translation.

Post-transcriptional mechanisms also affect location of protein synthesis. mRNA localization is especially important in polarized cells like neurons, where translation must be restricted to discrete subcellular locations. In neuronal cells, mRNA transport to dendrites and the subsequent local protein synthesis are required for appropriate synaptic plasticity and long-term memory formation. Interestingly, dynamic tail length control is believed to be essential for ensuring localized protein synthesis. Analogous to storage and activation of maternal mRNAs in oocytes (Lim et al. 2016; Sheets et al. 1995), during the transport to dendrites, some transcripts are translationally silent with short poly(A) tails, which limits protein synthesis in unwanted subcellular regions (Wu et al. 1998). These dendritic mRNAs undergo readenylation and concomitant translational activation upon synaptic stimulation. The involvement of poly(A) tail control in local translation has only been validated for a handful of example genes in some rare contexts, and the widespread usage in neuronal cells or broader cellular systems remains largely unknown.

Taken together, poly(A) tail length regulation underlies multiple aspects of a transcript's life cycle and plays important roles in various physiological contexts. However, several questions remain elusive, including whether cytoplasmic polyadenylation is widely used in non-developmental systems as a means of post-transcriptional regulation of gene expression.

1.5 Factors regulating poly(A) tail length in the cytoplasm

Poly(A) tail length control is largely directed by specific 3'UTR sequences and the cognate trans-acting factors. Numerous sequence elements were discovered to cause tail shortening and subsequent mRNA degradation. Among them, AU-rich elements (AREs, composed of multiple AUUUA pentamers) and guanosine-uridine rich elements (GREs) are well studied in mammals (Bakheet et al. 2001; Barreau et al. 2005; Louis and Bohjanen 2011). These two classes of 3'UTR elements bind to multiple AU-binding proteins or GU-binding proteins, which recruit deadenylase machinery. For example, ZFP36 (also known as TTP) RBPs recognize AU-rich elements (AREs) in 3'UTRs, and subsequently, recruit deadenylases such as polyA ribonuclease (PARN) or the CCR4-NOT complexes (Fabian et al. 2013; Bulbrook et al. 2018; Lai et al. 2003). However, the impacts of AREs are not definitive and depend on which ARE-binding proteins they interact with. Some of the ARE-binding proteins can destabilize mRNA by recruiting the RNA deadenylation machinery whereas others could stabilize mRNA by competing with the destabilizing ARE binding proteins for ARE occupancy, thus preventing mRNA decay. CUG-binding protein 1 (CUGBP1 or CELF1), a member of the CELF RNA-binding protein family, can specifically target mRNA containing GREs to mediate poly(A) tail shortening via recruitment of PARN, leading to mRNA decay (Louis and Bohjanen 2011). In addition, PUF family proteins recognize specific 3'UTR sequences to recruit

deadenylases and repress translation, conserved from yeast to humans (Cooke et al. 2011; Goldstrohm et al. 2007; Van Etten et al. 2012). microRNAs are another pillar of 3'UTR regulatory elements that regulate translation and mRNA decay via affecting poly(A) tail length (Djuranovic et al. 2012; Eisen et al. 2020). They are non-coding RNAs of ~21 nucleotides that base-pair to partially complementary sequences in the 3' UTR of their target RNAs (Bartel 2009). microRNAs in the form of the miRNA-induced silencing complex (RISC) recognize microRNA-binding sites in the 3'UTR of their target mRNA, and cause deadenylation of the mRNAs leading to translational silencing and mRNA decay. Recently, many other RNA–deadenylase adapters have also been identified in recent years, including nanos (Bhandari et al. 2014), roquin (Schuetz et al. 2014), and YTHDF2 (Du et al. 2016). Many of these trans-acting factors were functionally validated by an artificial tethering system where the RBPs are physically tethered to a reporter transcript, resulting in increased deadenylation and mRNA decay (Du et al. 2016). As such, there are many examples of *trans* factors that bind to the 3'UTRs that mediate poly(A) tail shortening via directly or indirectly recruiting deadenylase complexes, leading to mRNA degradation, and it is likely there are many other unknown factors.

Almost every human mRNAs have multiple sequence elements in the 3'UTRs, some of which act in an interactive way, variably influencing poly(A) tail length. For example, RBPs have been increasingly shown to crosstalk with miRNAs in cooperative or competitive ways (Srikantan and Gorospe 2012; Srikantan et al. 2012). ELAVL1 can recruit miR-19 and the associated RISC complex to the mRNA transcript of small GTPase RhoB, facilitating miRNA-mediated translational repression (Glorian et al. 2011). Certain miRNAs were reported to interact with ARE-specific binding proteins (AUBPs) to have complementary or antagonistic effects on mRNA repression (Jing et al. 2005). To summarize, individual mRNAs contain

multiple sequence elements which can act in an interactive manner to fine-tune the efficiency of tail length regulation.

While there are clearly multiple cis-acting elements that control deadenylation, several sequence elements have also been identified to direct cytoplasmic polyadenylation for specific mRNAs in certain conditions such as oocytes (Eichhorn et al. 2016; Lim et al. 2016), embryonic cells (Eichhorn et al. 2016; Lim et al. 2016), and neurons (Wu et al. 1998; Weill et al. 2012). Cytoplasmic polyadenylation is a critical mechanism for controlling the timing of gene expression in cells that are no longer transcriptionally active, such as oocytes, or for controlling the localized translation seen at neuronal synapses. The cytoplasmic polyadenylation element (CPEs) has been best characterized in *Xenopus* oocytes, and are typically (but not limited to) a UA-rich sequence (UUUUA1-3U) or a U-rich sequence (up to 18 uracil residues) that are bound by CPE-binding proteins (CPEBs) (Fox et al. 1989; Hake and Richter 1994). CPEBs direct polyadenylation for maternal mRNAs during the maturation of oocytes or post-fertilization. Analogous to the multiple cis-acting elements governing conventional nuclear polyadenylation, additional sequences act to precisely regulate polyadenylation across the developmental stage, including MSI-binding element (MBE) and translational control sequence (TCS) (Charlesworth et al. 2013). Generally, as with the CPE, these cis-acting elements also require the PAS to induce polyadenylation in the cytoplasm. Since most of these cis- and trans acting factors involved in readenylation are characterized in oocytes, and early developmental system, there are likely many other unknown factors that mediate readenylation in post-embryonic systems.

1.6 Mechanisms modulating poly(A) tail length regulation during a cellular response

Same transcripts can undergo differential tail length control across tissue via tissue-specific modulation of the interaction between the 3'UTR sequences and the cognate trans-acting factors. Likewise, the interaction between 3'UTR sequences and the trans factors can be dynamically regulated in response to environmental cues, allowing for dynamic tail length control. There are several mechanisms that control the interaction between 3'UTR sequences and trans factors. First, the interaction can be controlled by modulating the abundance of available cognate RNA-binding proteins, or their RNA affinity via post-translational modifications as well. For example, the phosphorylation of ZFP36 in cells disrupts its binding to CCR4–NOT and stabilizes mRNAs of the inflammatory response (Tiedje et al. 2016; Ronkina et al. 2019). Another well-known example is ELAV-like protein1 (ELAVL1) modulation during DNA damage response. ELAVL1 is an RBP that stabilizes target mRNAs by recognizing U-rich signature motif (Peng et al. 1998). Ionizing radiation triggers phosphorylation of ELAVL1 protein at residues near the RNA-recognition domain, consequently increasing the association of ELAVL1 with the target transcripts instead of affecting ELAVL1 abundance in the cytoplasm (Grammatikakis et al. 2017). Also, the interaction between a given *trans* factor and the target transcripts can be altered by modulating the abundance of other trans factors competing on the same RNA-binding motif. For example, CELF1 and ELAVL1 compete for overlapping binding sites within *MYC* transcripts with opposing effects on mRNA translation (Liu et al. 2015). During inhibition of intestinal mucosal growth, CELF1 repress translation of *MYC* mRNA without affecting its mRNA level, by associating more with the 3'UTR of *MYC* mRNA instead of ELAVL1.

Additionally, poly(A) tail length regulation can be modulated by changing the protein-protein interaction that an RBP forms on its target mRNA. In oocytes, embryonic cells, and neurons, CPEBs were shown to have dual roles in poly(A) tail length

regulation. Depending on the phosphorylation status of CPEBs, CPEBs can positively or negatively impact poly(A) tail length. For example, early in oocyte maturation, CPEB, bound to its target mRNA, forms a complex with symplekin, CPSF, the poly(A) ribonuclease (PARN) deadenylase, and germ-line development factor 2 (Gld2), a poly(A) polymerase (Mendez and Richter 2001; Weill et al. 2012). PARN removes the adenosine residues added by Gld2 to maintain the short (A) tail. Once it reaches the desired developmental stage/subcellular location, CPEB gets phosphorylated, allowing PARN to be released from the complex and thus Gld2 to elongate the poly(A) tail of the target mRNA for translational activation. Lastly, poly(A) tail length regulation is often modulated by affecting the presence and accessibility of cis-acting elements within the 3'UTR of transcripts in vertebrates. For example, in genes with multiple poly(A) sites in the 3'UTR regions, switching to proximal p(A) sites generates shorter 3'UTR isoforms, which are no longer susceptible to a tail length regulation mediated by the *cis* elements present only in longer 3'UTRs. Early embryogenesis highlights the intersection of poly(A) site switching (in other words, APA) and poly(A) tail length control. Very early after fertilization of oocytes in vertebrates, transcription is silent, and maternally derived mRNAs with long 3'UTRs containing CPEs are positively regulated by CPEBs to drive early embryogenesis. Upon transcriptional activation of the zygotic genome, factors are expressed that bind to maternal mRNA 3'UTRs and mediate their degradation. In addition, the newly expressed genes generally have shorter 3'UTRs and thus lack these negative regulatory sites. In general, these genes are involved in the rapid cell divisions in the early embryo. While embryonic tissues tend to express shorter 3'UTRs than differentiated tissues, alternative 3'UTRs are usually expressed in a cell-type specific manner across differentiated tissues: brain and testis generally express mRNAs with long and short 3'UTRs, respectively.

APA occurs through activation of signaling pathways with the most notable example of T cell activation (Jurgens et al. 2021; Sandberg et al. 2008). In quiescent T cells, genes required for activation are expressed with long 3'UTR isoforms, many of which have miRNA binding sites in the 3'UTR regions; thus these mRNAs undergo rapid deadenylation linked to poor translation and rapid degradation and, as a consequence, low levels of protein production. Since the mRNAs encoding activation factors are still being transcribed, the T cell is “primed” for rapid activation. Upon stimulation of T cells, general shortening of 3'UTRs occurs for important regulatory genes through a shift toward proximal PAS usage. This 3'UTR shortening results in the removal of miRNA binding sites that negatively regulate translation and mRNA abundance, thus allowing higher levels of expression of the encoded proteins.

In summary, many modes of regulation, including poly(A) site choice, abundance and post-translational modification of trans-acting factors, and competition among trans-acting factors are involved in dynamic poly(A) tail length control and thus it is challenging to predict poly(A) tail length control upon a single aspect of them.

1.7 Relevance of poly(A) tail length regulation in human diseases

Over the past few years, several studies have shown the impact of poly(A) tail control in several diseases. Abnormalities in the 3' end processing mechanisms is a common feature of many oncological, and immunological disorders (Rehfeld et al. 2013).

Alteration of poly(A) signal (AAUAAA) severely affect the expression of the transcripts, responsible for numerous human diseases (Sheets et al. 1990; Garin et al. 2010). It has also been shown that increased PAP activity is associated with poor prognosis in certain cancers (Scorilas 2002), while PAP inhibition affects some genes important for a proinflammatory response (Kondrashov et al. 2012). Mutation in

PABPN1 was also reported in several neurological diseases including the dominant oculopharyngeal muscular dystrophy (OPMD) (Richard et al. 2015).

Abnormal cytoplasmic poly(A) tail length regulation has also been shown to have detrimental effects on human physiology. For example, dyskerotosis congenita (DC) is an inherited bone marrow failure disorder that is associated with mutation in PARN, though the physiological relevance of PARN-targeted transcripts has not been fully characterized. Altered CPEBs and subsequent alteration in transcriptome polyadenylation have been associated to the etiology of various diseases such as cancer (Ortiz-Zapater et al. 2011; Pérez-Guijarro et al. 2016), chronic liver disease (Calderone et al. 2016), epilepsy (Parras et al. 2020), autism (Alexandrov et al. 2012), and Huntington's disease (Picó et al. 2021), thus suggesting CPEB-dependent dysregulated pathways as new therapeutic targets for these diseases.

ZFP36 is a well-known RNA-binding protein that targets ARE-containing mRNA for deadenylation and subsequent mRNA decay and/or translational repression, playing crucial roles in immune gene regulation. Recent studies have shown that as an inflammation-related protein, ZFP36 plays a significant role in cancer onset and progression by modulating the tumor microenvironment (TME), which suggests ZFP36's potential mechanism in the transformation from inflammation to tumorigenesis (Zhang et al. 2021). Moreover, ZFP36 can be regarded as a safeguard against carcinoma due to its role in the TME. Surprisingly, the mechanism is involved in PD-1/PD-L1. PD-L1 mRNA degraded by ZFP36 at the post-transcriptional level has the potential in reducing tumor immune evasion (Coelho et al. 2017; Kim et al. 2020). There have three subtypes of ZFP36 been reported (Zhang et al. 2021), ZFP36*2 (A>G), ZFP36*8 (C>T) and ZFP36*10 (2bp deletion). ZFP36*2 has been proved to be a potential biomarker in Caucasian breast cancer patients while ZFP36*8 has been found high associated with HER2-positive-breast cancer (Griseri et al. 2011).

As such, poly(A) tail length control deeply underlies various human disease, and thus more efforts are needed to link the pathophysiology of diseases to SNPs in the 3'UTR and RBPs involved in poly(A) tail length control, which will lead to the identification of novel disease markers and therapeutical targets for various human diseases.

1.8 Existing tools to study poly(A) tail length

Until recently, poly(A) tail length was investigated through northern blot, RT/PCR-based techniques, or oligo(dT)-based affinity purification. Despite the remarkable advances in next-generation sequencing (NGS) approaches, a global investigation of poly(A) tail length has been challenging, due to technical difficulties intrinsic to the NGS sequencing technology, Illumina platform, in particular (Quail et al. 2012).

However, more recently, there has been significant progress in NGS-based poly(A) tail length measurements, leading to the development of various genome-wide tail length profiling methods: TAIL-seq (Chang et al. 2014) and PAL-seq (Subtelny et al. 2014).

These methods provide accurate estimates for poly(A) tail length transcriptome-wide by modifying the biochemical or computational operations of a short-read sequencing platform either by directly sequencing through the tail or quantifying chemiluminescent signal proportional to the tail length. Both methods have their own advantages and limitations in their ability. PAL-seq enables accurate estimation of a broad range of tail lengths, whereas the maximum tail length that TAIL-seq can measure is limited to ~230 nucleotides (nt). However, unlike PAL-seq, TAIL-seq can determine terminal modifications of poly(A) tails, which reportedly play crucial roles in controlling mRNA stability. Studies using TAIL-seq have revealed that uridylation and guanylation are widespread at the 3' termini of poly(A) tails of human mRNAs, affecting mRNA stability (Lim et al. 2018; Kim et al. 2020; Chang et al. 2014). They

found that uridylation decreases mRNA stability, while the mixed tailing with intermittent guanosine tends to increase mRNA stability. There are some limitations common to both methods. Despite the regulatory importance of 3'UTR isoform usage, neither method provides information about the 3'UTR isoforms to which the tails are attached, which makes it infeasible to study the relationship between the tail length and the mRNA isoform choice. Also, these methods are not broadly available to other researchers due to the requirement of modifying sequencing hardware or software. In an effort to resolve the disadvantages of previous methods, new poly(A) tail length profiling methods were more recently developed, including TED-seq (Woo et al. 2018), and FLAM-seq (Legnini et al. 2019). TED-seq is an affordable global poly(A) tail length profiling method that is compatible with the regular short-read sequencing platform (e.g., Illumina) and thus can be broadly accessible to other researchers. TED-seq still employs short-read sequencing, but by confining the library insert size to a narrow range (± 10 bp) of target size, and together with the 3' cleavage site information, it estimates the tail length accurately at 3'UTR isoform level. The experimental and computational process of TED-seq will be discussed in detail in my thesis. More recently, a new method, named FLAM-seq, was introduced to measure tail length at transcript isoform level by sequencing the entire sequence of poly(A)-containing mRNAs using PacBio Sequel System. By providing full-length mRNA sequence as well as directly reading through the poly(A) tail, FLAM-seq allows for measuring tail length and its internal nucleotide non(A) composition. It unveiled that poly(A) tails contain significant amount of internal non-A nucleotides, mostly cytosines, though its regulatory potential is not validated yet. FLAM-seq can examine the relevance of poly(A) tail length to other mRNA features such as transcription start site, alternative splicing and 3'UTR choice in a single experiment using the full-length mRNA sequence information. Very recently, Nanopore direct RNA sequencing

showed potential with the advantage of avoiding amplification biases (Garalde et al. 2018), although its accuracy and precision were not yet thoroughly validated, and Nanopore sequencing has currently an extremely high error rate. Also, it is concerning that direct RNA/cDNA sequencing would require significantly more input material to have reads of a good sequencing depth, limiting its application in many cases with limited biological samples.

1.9 Current methodology and its limitation

Many studies using these genome-wide methods provided a global description of poly(A) tail length in various biological samples. They profile poly(A) tail lengths for all transcripts in the cells, calculating the average poly(A) tail length for mRNAs derived from the same gene. Using this approach, previous global studies revealed that poly(A) tail length varies transcriptome-wide in mammals. Also, it was shown that different transcript isoforms generated from the same genes have different tail lengths. Poly(A) tail lengths *in vivo* and *in vitro* are globally phased in ~30 nt increments (Lima et al. 2017), which is consistent with the footprints of PABPC proteins on the poly(A) tails (Baer and Kornberg 1983).

It should be noted that average poly(A) tail length of the mRNA pool doesn't tell poly(A) tail length dynamics (or kinetics) across the mRNA lifespan. For examples, the average length cannot tell us how long poly(A) tails are added to new transcripts during nuclear polyadenylation, how poly(A) tail lengths change upon export to the cytoplasm. There are few cellular contexts that the average poly(A) tail lengths reasonably represent actual post-transcriptional tail controls: oocytes and early embryos. In these early developmental contexts, transcription is silent, and thus poly(A) tails are not affected by new transcripts. More importantly, it seems straightforward to monitor post-transcriptional poly(A) tail length control across the time-course of their cellular processes (e.g. oocyte maturation and early

embryogenesis), because transcription doesn't influence the mRNA pool in the cells. Most global tail length studies were done in these systems, showing that poly(A) tail length correlates well with translation efficiency and half-life in oocyte maturation (Lim et al. 2016), and early embryogenesis (Eichhorn et al. 2016).

In fact, averaging poly(A) tail lengths of mRNA pool is especially problematic in somatic cells. Since new transcripts continuously enter the mRNA pool in a somatic system, and transcription activity varies across the transcriptome. Poly(A) tail lengths of the mRNA pool are likely confounded by new transcripts. Indeed, the differences in mean poly(A) tail length doesn't correlate with mRNA abundance, mRNA stability, translation, ribosome occupancy, and/or PABPC occupancy in various somatic cells (Rissland et al. 2017; Subtelny et al. 2014). Notably, poly(A) tails of highly translated, stable mRNAs are relatively short (about 30 nt, a length that accommodate a single PABPC), whereas poorly translated mRNAs tend to have long poly(A) tails (Lima et al. 2017). This may suggest that one PABPC is sufficient to promote efficient translation and maintain mRNA stability, but also may indicate the possibility that poorly translated mRNA may have long poly(A) tails, because they undergo rapid deadenylation and mRNA decay, leading to depletion of the short-tailed mRNA in the mRNA pool while enriching new transcripts with initially added long poly(A) tails (Eisen et al. 2020a; Eisen et al. 2020b). Indeed, the latter possibility was validated using a Tet promoter-regulatable reporter system where deadenylation rate was measured for the pulse-generated reporter mRNAs with high and low translation efficiency (Lima et al. 2017; Webster et al. 2018; Hanson and Coller 2018). Very recently, Eisen et al. first examined the kinetics of deadenylation instead of steady-state tail length, showing a strong correlation between mRNA half-life and deadenylation rate (Eisen et al. 2020a; Eisen et al. 2020b).

Collectively, these previous results suggest that poly(A) tail length-dependent regulation play a significant role in gene regulation in somatic cells, but the information on poly(A) tail length regulation may not be reflected by the averaged poly(A) tail length for the pool of mRNAs. Therefore, it is important to come up with alternative approaches that can discern transcriptional and post-transcriptional poly(A) tail regulation (e.g. measuring actual poly(A) tail length kinetics/dynamics) for our correct understanding of poly(A) tail length regulation in transcriptionally active, somatic systems.

1.10 Regulation of poly(A) tail in immunity

Poly(A) tail length dynamics was well-studied through the TNF mRNA during macrophage activation (Crawford et al. 1997). In unstimulated macrophages, TNF mRNAs are constitutively expressed but exist with short poly(A) tails and translationally repressed. Following LPS stimulation, TNF transcripts gain poly(A) tails and translationally activated, allowing the rapid and abundant expression of TNF proteins (**Figure 1.3**). Similar poly(A) tail length dynamics was also reported in memory CD8 T cell activation. In the resting cells, constitutively expressed mRNA that encodes CC-chemokine ligand 5 (CCL5) lacks a poly(A) tail and so is translationally repressed until the T cell receptor is activated (Swanson et al. 2002). Upon activation, readenylation occurs for the pre-existing pool of CCL5 mRNA, which facilitates rapid translation and CCL5 protein secretion. It is therefore possible that, in addition to TNF, many other transcripts may be constitutively produced in resting macrophages and stored in a translationally silent state until LPS triggers their rapid readenylation and translation. To answer this question, in my thesis various alternative approaches were employed to globally measure post-transcriptional tail length changes in macrophage cells challenged by LPS.

FIGURES

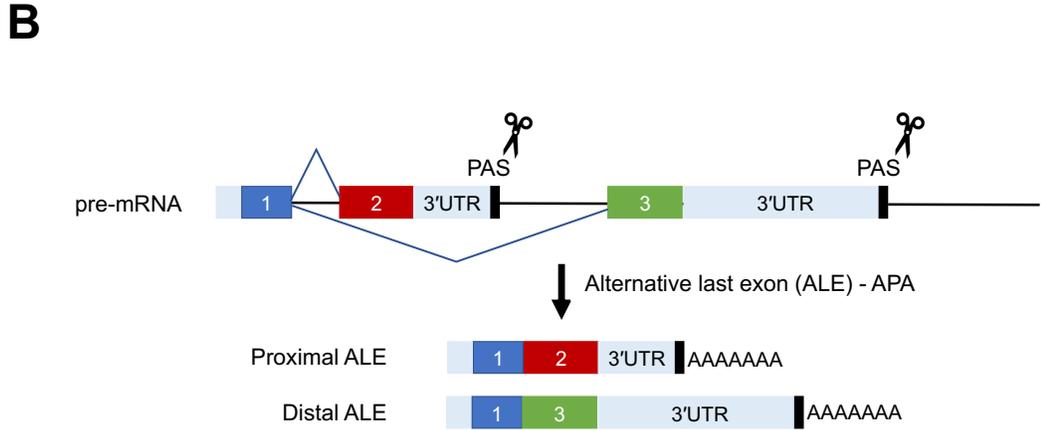
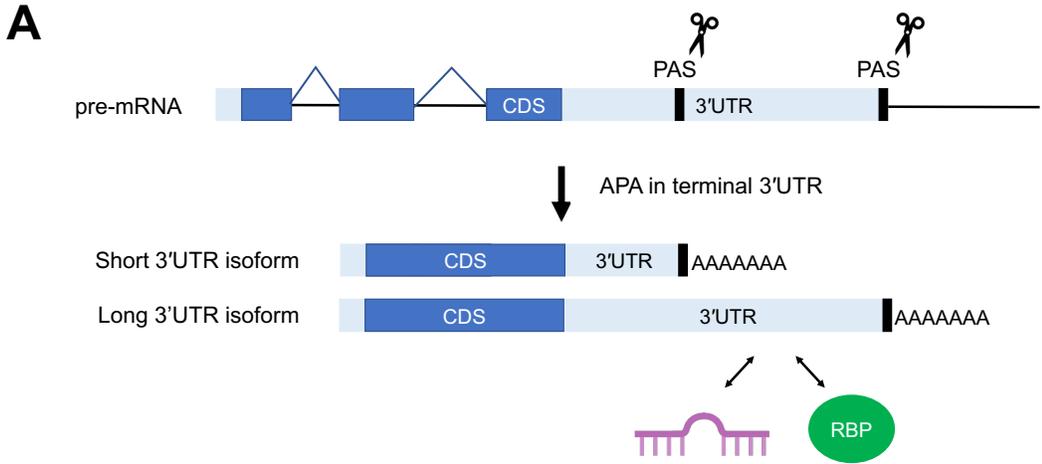


Figure 1.1. Two major types of mRNA isoforms generated using alternative polyadenylation sites.

(A) 3'UTR can harbor more than one polyadenylation site. The selection of proximal over distal polyadenylation sites, or reverse, generates 3'UTR isoforms of variable lengths (short 3'UTR isoform vs. long 3'UTR isoform), a process termed alternative polyadenylation (APA). Since long 3'UTR isoforms often have additional regulatory elements that can be recognized by *trans*-acting factors such as RNA-binding proteins (RBP, green), or microRNAs (miRNAs, pink), allowing for different regulation of the alternative mRNA isoforms.

(B) During alternative splicing of last exons (ALE), the selection of either exon 2 (red) or exon 3 (green) as the last exon produces two distinct isoforms with different protein-coding sequences along with entirely distinct 3'UTRs (proximal ALE vs distal ALE). The protein-coding, exon regions (CDS) are numerically denoted and color-coded: 1, 2, and 3.

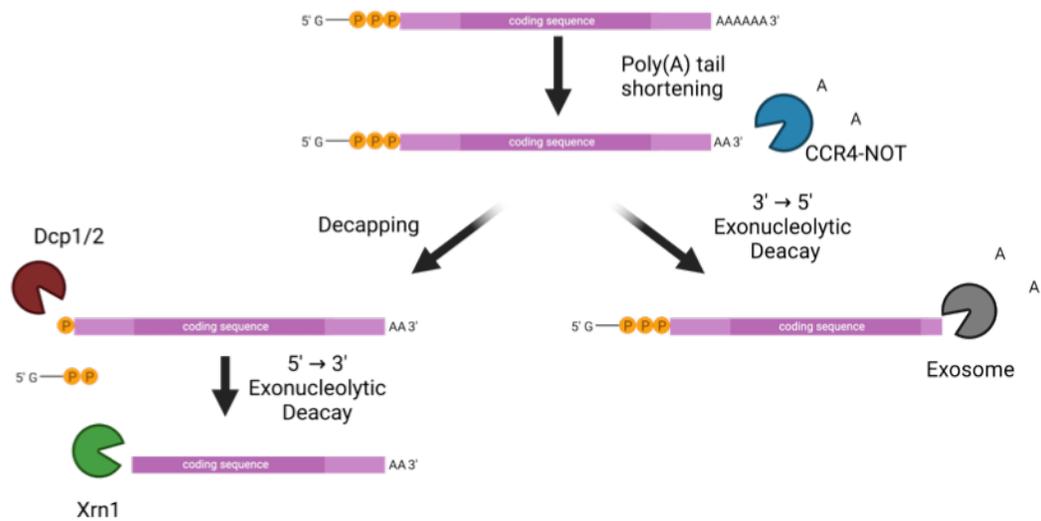


Figure 1.2. Major decay pathways of eukaryotic mRNAs in the cytoplasm.

The major pathways of cytoplasmic mRNA decay are triggered by poly(A) shortening (deadenylation). Deadenylation primarily leads to decapping by the Dcp1-Dcp2 complex at the 5' end and the subsequent 5' to 3' exonucleolytic digestion by Xrn1. Alternatively, after deadenylation, mRNA can undergo degradation from the 3' end by an exosome.

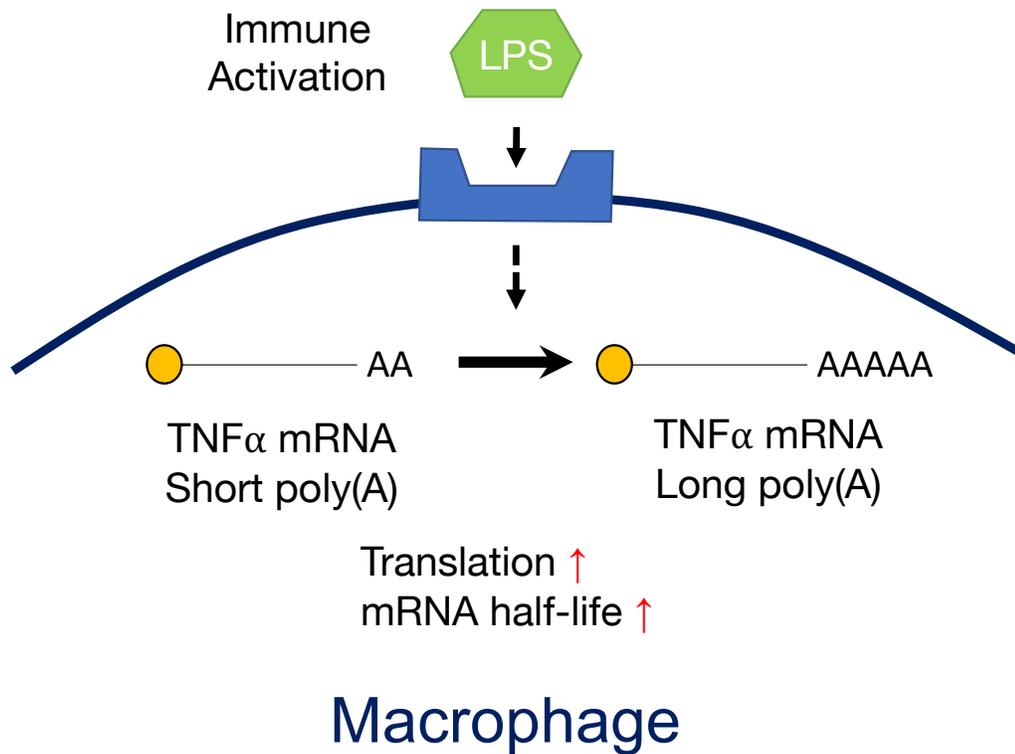


Figure 1.3. Poly(A) tail length control for TNF mRNAs during macrophage activation.

In the resting macrophages, TNF mRNAs are constitutively expressed but exist with short poly(A) tails and translationally repressed. Following LPS stimulation, TNF transcripts undergo readenylation leading to translational activation, consequently allowing the rapid induction of TNF proteins.

REFERENCES

- Alexandrov IM, Ivshina M, Jung DY, Friedline R, Ko HJ, Xu M, O'Sullivan-Murphy B, Bortell R, Huang Y-T, Urano F, et al. 2012. Cytoplasmic polyadenylation element binding protein deficiency stimulates PTEN and Stat3 mRNA translation and induces hepatic insulin resistance. *PLoS Genet* **8**: e1002457.
- Baer BW, Kornberg RD. 1983. The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *J Cell Biol* **96**: 717–721.
- Bakheet T, Frevel M, Williams BRG, Greer W, Khabar KSA. 2001. ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. *Nucleic Acids Res* **29**: 246–254.
- Barreau C, Paillard L, Osborne HB. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* **33**: 7138–7150.
- Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233.
- Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* **10**: 1001–1010.
- Bhandari D, Raisch T, Weichenrieder O, Jonas S, Izaurralde E. 2014. Structural basis for the Nanos-mediated recruitment of the CCR4-NOT complex and translational repression. *Genes Dev* **28**: 888–901.
- Blobel G. 1973. A protein of molecular weight 78,000 bound to the polyadenylate region of eukaryotic messenger RNAs. *Proc Natl Acad Sci U S A* **70**: 924–928.
- Bulbrook D, Brazier H, Mahajan P, Kliszczak M, Fedorov O, Marchese FP, Aubareda A, Chalk R, Picaud S, Strain-Damerell C, et al. 2018. Tryptophan-Mediated Interactions between Tristetraprolin and the CNOT9 Subunit Are Required for CCR4-NOT Deadenylase Complex Recruitment. *J Mol Biol* **430**: 722–736.
- Calderone V, Gallego J, Fernandez-Miranda G, Garcia-Pras E, Maillo C, Berzigotti A, Mejias M, Bava F-A, Angulo-Urarte A, Graupera M, et al. 2016. Sequential Functions of CPEB1 and CPEB4 Regulate Pathologic Expression of Vascular Endothelial Growth Factor and Angiogenesis in Chronic Liver Disease. *Gastroenterology* **150**: 982-997.e30.
- Carpenter S, Ricci EP, Mercier BC, Moore MJ, Fitzgerald KA. 2014. Post-transcriptional regulation of gene expression in innate immunity. *Nat Rev Immunol* **14**: 361–376.

- Chang H, Lim J, Ha M, Kim VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* **53**: 1044–1052.
- Charlesworth A, Meijer HA, de Moor CH. 2013. Specificity factors in cytoplasmic polyadenylation. *Wiley Interdiscip Rev RNA* **4**: 437–461.
- Coelho MA, de Carné Trécesson S, Rana S, Zecchin D, Moore C, Molina-Arcas M, East P, Spencer-Dene B, Nye E, Barnouin K, et al. 2017. Oncogenic RAS Signaling Promotes Tumor Immuno-resistance by Stabilizing PD-L1 mRNA. *Immunity* **47**: 1083-1099.e6.
- Cooke A, Prigge A, Opperman L, Wickens M. 2011. Targeted translational regulation using the PUF protein family scaffold. *Proceedings of the National Academy of Sciences* **108**: 15870–15875.
- Crawford EK, Ensor JE, Kalvakolanu I, Hasday JD. 1997. The Role of 3' Poly(A) Tail Metabolism in Tumor Necrosis Factor- α Regulation. *J Biol Chem* **272**: 21120–21127.
- Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**: 853–866.
- Djuranovic S, Nahvi A, Green R. 2012. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**: 237–240.
- Du H, Zhao Y, He J, Zhang Y, Xi H, Liu M, Ma J, Wu L. 2016. YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat Commun* **7**: 12626.
- Eckmann CR, Rammelt C, Wahle E. 2011. Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* **2**: 348–361.
- Edmonds M, Abrams R. 1960. Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *J Biol Chem* **235**: 1142–1149.
- Eichhorn SW, Subtelny AO, Kronja I, Kwasnieski JC, Orr-Weaver TL, Bartel DP. 2016. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos ed. E. Izaurralde. *eLife* **5**: e16955.
- Eisen TJ, Eichhorn SW, Subtelny AO, Bartel DP. 2020a. MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. *Molecular Cell* **77**: 775-785.e8.

- Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, Bartel DP. 2020b. The Dynamics of Cytoplasmic mRNA Metabolism. *Molecular Cell* **77**: 786–799.e10.
- Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics* **14**: 496–506.
- Fabian MR, Frank F, Rouya C, Siddiqui N, Lai WS, Karetnikov A, Blackshear PJ, Nagar B, Sonenberg N. 2013. Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat Struct Mol Biol* **20**: 735–739.
- Fox CA, Sheets MD, Wickens MP. 1989. Poly(A) addition during maturation of frog oocytes: distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes Dev* **3**: 2151–2162.
- Gallie DR. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* **5**: 2108–2116.
- Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, et al. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**: 201–206.
- Garin I, Edghill EL, Akerman I, Rubio-Cabezas O, Rica I, Locke JM, Maestro MA, Alshaikh A, Bundak R, del Castillo G, et al. 2010. Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. *Proceedings of the National Academy of Sciences* **107**: 3105–3110.
- Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**: 812–824.
- Glorian V, Maillot G, Polès S, Iacovoni JS, Favre G, Vagner S. 2011. HuR-dependent loading of miRNA RISC to the mRNA encoding the Ras-related small GTPase RhoB controls its translation during UV-induced apoptosis. *Cell Death Differ* **18**: 1692–1701.
- Goldstrohm AC, Seay DJ, Hook BA, Wickens M. 2007. PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J Biol Chem* **282**: 109–114.
- Goldstrohm AC, Wickens M. 2008. Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* **9**: 337–344.
- Grammatikakis I, Abdelmohsen K, Gorospe M. 2017. Posttranslational control of HuR function. *Wiley Interdiscip Rev RNA* **8**: 10.1002/wrna.1372.

- Griseri P, Bourcier C, Hieblot C, Essafi-Benkhadir K, Chamorey E, Touriol C, Pagès G. 2011. A synonymous polymorphism of the Tristetraprolin (TTP) gene, an AU-rich mRNA-binding protein, affects translation efficiency and response to Herceptin treatment in breast cancer patients. *Hum Mol Genet* **20**: 4556–4568.
- Gu H, Das Gupta J, Schoenberg DR. 1999. The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. *Proc Natl Acad Sci U S A* **96**: 8943–8948.
- Hake LE, Richter JD. 1994. CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* **79**: 617–627.
- Hanson G, Collier J. 2018. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* **19**: 20–30.
- Jacobson A, Favreau M. 1983. Possible involvement of poly(A) in protein synthesis. *Nucleic Acids Res* **11**: 6353–6368.
- Jalkanen AL, Coleman SJ, Wilusz J. 2014. Determinants and Implications of mRNA Poly(A) Tail Size - Does this Protein Make My Tail Look Big? *Semin Cell Dev Biol* **0**: 24–32.
- Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, Chen J, Di Padova F, Lin S-C, Gram H, Han J. 2005. Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* **120**: 623–634.
- Jurgens AP, Popović B, Wolkers MC. 2021. T cells at work: How post-transcriptional mechanisms control T cell homeostasis and activation. *Eur J Immunol* **51**: 2178–2187.
- Kim D, Lee Y-S, Jung S-J, Yeo J, Seo JJ, Lee Y-Y, Lim J, Chang H, Song J, Yang J, et al. 2020. Viral hijacking of the TENT4-ZCCHC14 complex protects viral RNAs via mixed tailing. *Nat Struct Mol Biol* **27**: 581–588.
- Kim DJ, Jang JH, Ham S-Y, Choi SH, Park SS, Jeong SY, Kim BC, Jeon DY, Lee BJ, Ko BK, et al. 2020. Doxorubicin inhibits PD-L1 expression by enhancing TTP-mediated decay of PD-L1 mRNA in cancer cells. *Biochem Biophys Res Commun* **522**: 402–407.
- Kondrashov A, Meijer HA, Barthet-Barateig A, Parker HN, Khurshid A, Tessier S, Sicard M, Knox AJ, Pang L, De Moor CH. 2012. Inhibition of polyadenylation reduces inflammatory gene induction. *RNA* **18**: 2236–2250.
- Lai WS, Kennington EA, Blackshear PJ. 2003. Tristetraprolin and its family members can promote the cell-free deadenylation of AU-rich element-containing mRNAs by poly(A) ribonuclease. *Mol Cell Biol* **23**: 3798–3812.

- Legnini I, Alles J, Karaiskos N, Ayoub S, Rajewsky N. 2019. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nature Methods* **16**: 879–886.
- Lim J, Kim D, Lee Y, Ha M, Lee M, Yeo J, Chang H, Song J, Ahn K, Kim VN. 2018. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science* **361**: 701–704.
- Lim J, Lee M, Son A, Chang H, Kim VN. 2016. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev.* <http://genesdev.cshlp.org/content/early/2016/07/14/gad.284802.116> (Accessed December 30, 2019).
- Lima SA, Chipman LB, Nicholson AL, Chen Y-H, Yee BA, Yeo GW, Collier J, Pasquinelli AE. 2017. Short Poly(A) Tails are a Conserved Feature of Highly Expressed Genes. *Nat Struct Mol Biol* **24**: 1057–1063.
- Liu L, Ouyang M, Rao JN, Zou T, Xiao L, Chung HK, Wu J, Donahue JM, Gorospe M, Wang J-Y. 2015. Competition between RNA-binding proteins CELF1 and HuR modulates MYC translation and intestinal epithelium renewal. *Mol Biol Cell* **26**: 1797–1810.
- Louis IV-St, Bohjanen PR. 2011. Coordinate Regulation of mRNA Decay Networks by GU-rich Elements and CELF1. *Curr Opin Genet Dev* **21**: 444–451.
- Mendez R, Richter JD. 2001. Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol* **2**: 521–529.
- Ortiz-Zapater E, Pineda D, Martínez-Bosch N, Fernández-Miranda G, Iglesias M, Alameda F, Moreno M, Eliscovich C, Eyrales E, Real FX, et al. 2011. Key contribution of CPEB4-mediated translational control to cancer progression. *Nat Med* **18**: 83–90.
- Parras A, de Diego-Garcia L, Alves M, Beamer E, Conte G, Jimenez-Mateos EM, Morgan J, Ollà I, Hernandez-Santana Y, Delanty N, et al. 2020. Polyadenylation of mRNA as a novel regulatory mechanism of gene expression in temporal lobe epilepsy. *Brain* **143**: 2139–2153.
- Passmore LA, Collier J. 2022. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol* **23**: 93–106.
- Peng SS, Chen CY, Xu N, Shyu AB. 1998. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J* **17**: 3461–3470.
- Pérez-Guijarro E, Karras P, Cifdaloz M, Martínez-Herranz R, Cañón E, Graña O, Horcajada-Reales C, Alonso-Curbelo D, Calvo TG, Gómez-López G, et al.

2016. Lineage-specific roles of the cytoplasmic polyadenylation factor CPEB4 in the regulation of melanoma drivers. *Nat Commun* **7**: 13418.
- Picó S, Parras A, Santos-Galindo M, Pose-Utrilla J, Castro M, Fraga E, Hernández IH, Elorza A, Anta H, Wang N, et al. 2021. CPEB alteration and aberrant transcriptome-polyadenylation lead to a treatable SLC19A3 deficiency in Huntington's disease. *Science Translational Medicine* **13**: eabe7104.
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**: 341.
- Rehfeld A, Plass M, Krogh A, Friis-Hansen L. 2013. Alterations in Polyadenylation and Its Implications for Endocrine Disease. *Frontiers in Endocrinology* **4**. <https://www.frontiersin.org/article/10.3389/fendo.2013.00053> (Accessed April 15, 2022).
- Richard P, Trollet C, Gidaro T, Demay L, Brochier G, Malfatti E, Tom FM, Fardeau M, Lafor P, Romero N, et al. 2015. PABPN1 (GCN)11 as a Dominant Allele in Oculopharyngeal Muscular Dystrophy –Consequences in Clinical Diagnosis and Genetic Counselling. *J Neuromuscul Dis* **2**: 175–180.
- Rissland OS, Subtelny AO, Wang M, Lugowski A, Nicholson B, Laver JD, Sidhu SS, Smibert CA, Lipshitz HD, Bartel DP. 2017. The influence of microRNAs and poly(A) tail length on endogenous mRNA–protein complexes. *Genome Biology* **18**: 211.
- Ronkina N, Shushakova N, Tiedje C, Yakovleva T, Tollenaere MAX, Scott A, Batth TS, Olsen JV, Helmke A, Bekker-Jensen SH, et al. 2019. The Role of TTP Phosphorylation in the Regulation of Inflammatory Cytokine Production by MK2/3. *The Journal of Immunology*. <https://www.jimmunol.org/content/early/2019/09/13/jimmunol.1801221> (Accessed April 15, 2022).
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**: 1643–1647.
- Sato H, Maquat LE. 2009. Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes Dev* **23**: 2537–2550.
- Schoenberg DR, Maquat LE. 2012. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* **13**: 246–259.

- Schuetz A, Murakawa Y, Rosenbaum E, Landthaler M, Heinemann U. 2014. Roquin binding to target mRNAs involves a winged helix-turn-helix motif. *Nat Commun* **5**: 5701.
- Scorilas A. 2002. Polyadenylate polymerase (PAP) and 3' end pre-mRNA processing: function, assays, and association with disease. *Crit Rev Clin Lab Sci* **39**: 193–224.
- Sheets MD, Ogg SC, Wickens MP. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Research* **18**: 5799–5805.
- Sheets MD, Wu M, Wickens M. 1995. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature* **374**: 511–516.
- Slater DW, Slater I, Gillespie D. 1972. Post-fertilization synthesis of polyadenylic acid in sea urchin embryos. *Nature* **240**: 333–337.
- Srikantan S, Gorospe M. 2012. HuR function in disease. *Frontiers in bioscience (Landmark edition)* **17**: 189.
- Srikantan S, Tominaga K, Gorospe M. 2012. Functional interplay between RNA-binding protein HuR and microRNAs. *Curr Protein Pept Sci* **13**: 372–379.
- Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**: 66–71.
- Swanson BJ, Murakami M, Mitchell TC, Kappler J, Marrack P. 2002. RANTES production by memory phenotype T cells is controlled by a posttranscriptional, TCR-dependent process. *Immunity* **17**: 605–615.
- Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* **18**: 18–30.
- Tiedje C, Diaz-Muñoz MD, Trulley P, Ahlfors H, Laaß K, Blackshear PJ, Turner M, Gaestel M. 2016. The RNA-binding protein TTP is a global post-transcriptional regulator of feedback control in inflammation. *Nucleic Acids Res* **44**: 7418–7440.
- Uchida N, Hoshino S-I, Katada T. 2004. Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *J Biol Chem* **279**: 1383–1391.
- Van Etten J, Schagat TL, Hrit J, Weidmann CA, Brumbaugh J, Coon JJ, Goldstrohm AC. 2012. Human Pumilio Proteins Recruit Multiple Deadenylation Factors to Efficiently Repress Messenger RNAs. *J Biol Chem* **287**: 36370–36383.

- Vicens Q, Kieft JS, Rissland OS. 2018. Revisiting the Closed-Loop Model and the Nature of mRNA 5'-3' Communication. *Mol Cell* **72**: 805–812.
- Webster MW, Chen Y-H, Stowell JAW, Alhusaini N, Sweet T, Graveley BR, Collier J, Passmore LA. 2018. mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol Cell* **70**: 1089-1100.e8.
- Weill L, Belloc E, Bava F-A, Méndez R. 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol* **19**: 577–585.
- Wilt FH. 1973. Polyadenylation of maternal RNA of sea urchin eggs after fertilization. *Proc Natl Acad Sci U S A* **70**: 2345–2349.
- Woo YM, Kwak Y, Namkoong S, Kristjánisdóttir K, Lee SH, Lee JH, Kwak H. 2018. TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Reports* **24**: 3630-3641.e7.
- Wu L, Wells D, Tay J, Mendis D, Abbott MA, Barnitt A, Quinlan E, Heynen A, Fallon JR, Richter JD. 1998. CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. *Neuron* **21**: 1129–1139.
- Xiang K, Bartel DP. 2021. The molecular basis of coupling between poly(A)-tail length and translational efficiency. *Elife* **10**: e66493.
- Zhang D, Zhou Z, Yang R, Zhang S, Zhang B, Tan Y, Chen L, Li T, Tu J. 2021. Tristetraprolin, a Potential Safeguard Against Carcinoma: Role in the Tumor Microenvironment. *Front Oncol* **11**: 632189.
- Zhao J, Hyman L, Moore C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63**: 405–445.

CHAPTER 2¹

Dynamic and Widespread Control of Poly(A) Tail Length during Macrophage Activation

2.1 Abstract

The poly(A) tail enhances translation and transcript stability, and tail length is under dynamic control during cell state transitions. Tail regulation plays essential roles in translational timing and fertilization in early development, but poly(A) tail dynamics have not been fully explored in post-embryonic systems. Here, we examined the landscape and impact of tail length control during macrophage activation. Upon activation, more than 1,500 mRNAs, including pro-inflammatory genes, underwent distinctive changes in tail lengths. Increases in tail length correlated with mRNA levels regardless of transcriptional activity, and many mRNAs that underwent tail extension encode proteins necessary for immune function and post-transcriptional regulation. Strikingly, we found that ZFP36, whose protein product destabilizes target transcripts, undergoes tail extension. Our analyses indicate that many mRNAs undergoing tail lengthening are, in turn, degraded by elevated levels of ZFP36,

¹ The majority of this work is adapted from a manuscript of the same name accepted on March 21th, 2022 in *RNA Journal*. The authors of this manuscript were Yeonui Kwak (Y.K), Ciarán W.P. Daly (C.D), Elizabeth A Fogarty (E.A.F), Andrew Grimson (A.G), and Hojoong Kwak (H.K). Y.K and H.K conceptualized the study; all authors contributed to the study design and methodological approaches. Experiments were performed by Y.K, and analysis of TED-seq and PRO-seq data was performed by Y.K and H.K. The TED-seq, PRO-seq, 3'-seq and transcription inhibition data were analyzed by Y.K. under the supervision of A.G and H.K. Y.K and E.A.F performed 3' UTR reporter assays with contributions from C.D. The manuscript was written by Y.K, A.G, and H.K.

constituting a post-transcriptional feedback loop that ensures transient regulation of transcripts integral to macrophage activation. Taken together, this study establishes the complexity, relevance and widespread nature of poly(A) tail dynamics, and the resulting post-transcriptional regulation during macrophage activation.

2.2 Introduction

Polyadenylation refers to the 3' extension of mRNAs with adenosines and occurs after nascent transcript cleavage (Millevoi and Vagner 2010). The newly synthesized poly(A) tail is up to 250 nucleotides long and serves as a binding site for poly(A)-binding proteins (PABPs), which control multiple events throughout the RNA lifecycle, including nuclear export, translation and mRNA stability (Gallie 1991; Jalkanen et al. 2014). The length of the tail changes throughout the mRNA lifecycle, and many of these changes are mediated by interactions between 3'UTR regulatory sequences and RNA binding proteins or microRNAs (miRNAs). 3'UTR-trans factors often recruit deadenylases to shorten the tail but can also recruit cytoplasmic polyadenylases (Braun et al. 2011; Weill et al. 2012). Deadenylation is associated with mRNA decay, translational repression and altered localization, and deadenylation impacts most mRNAs (Eichhorn et al. 2016; Park et al. 2016; Zheng et al. 2008). However, exceptions exist: shortened poly(A) tails can be re-elongated in the cytoplasm to stabilize mRNAs and promote translation. Maternal mRNAs in early embryogenesis and viral RNAs in host cells undergo poly(A) tail extension, mediated by 3'UTR sequence or secondary structures that recruit non-canonical poly(A) polymerases (Kim et al. 2020; Lim et al. 2016). While these examples show the significance of poly(A) tail regulation (Lim et al. 2016; Wells et al. 2001; Wu et al. 1998; Weill et al. 2012), the extent and the importance of cytoplasmic polyadenylation in somatic cells has been unclear.

Multiple genome-wide poly(A) tail profiling methods exist, including PAL-seq (Subtelny et al. 2014), TAIL-seq (Chang et al. 2014), TED-seq (Woo et al. 2018) and FLAM-seq (Legnini et al. 2019). Using these techniques, many studies found large variation of tail lengths in steady-state post-embryonic transcriptome. However, in contrast to early embryos, there were only weak associations of tail size to translation efficiency, mRNA stability, abundance and PABP binding (Subtelny et al. 2014; Lima et al. 2017; Rissland et al. 2017). For example, while miRNAs increase deadenylation rates of target mRNAs (Eisen et al. 2020a; Giraldez et al. 2006; Wu et al. 2006), tail length changes were only captured by pre-steady-state measurements (Eisen et al. 2020a). In steady-state post-embryonic systems, tail changes were masked as a consequence of the rapid decay of deadenylated intermediates (Eisen et al. 2020a, 2020b). These studies highlight the complexity of steady-state poly(A) tail lengths in post-embryonic systems, and suggest that understanding poly(A) tail length regulation requires discrete pre-steady-state measurements across a gene regulatory response. However, most systematic, genome-wide investigations of transient-state poly(A) tail length control had been limited to specific biological contexts, such as oocyte fertilization (Lim et al. 2016). Pre-fertilization maternal mRNAs are stored as an unadenylated or partially adenylated form, and thus repressed in translation. Upon fertilization, they are polyadenylated in the cytoplasm, and their translation initiates globally (Lim et al. 2016). The absence of zygotic transcription makes oocytes a tractable system to study post-transcriptional regulation, as it becomes possible to examine changes in poly(A) tail lengths without the confounding influence of new transcripts. The extent and significance of poly(A) tail length control upon a developmental cue in transcriptionally active somatic cells has been difficult to examine.

Immune responses often require rapid and adaptable gene regulation, features suited to post-transcriptional control (Carpenter et al. 2014; Corbett 2018). Exposure of macrophages to lipopolysaccharides (LPS) induces rapid expression of inflammatory cytokines, such as tumor necrosis factor- α (TNF; Kontoyiannis et al. 1999, Parameswaran et al. 2010). Upon LPS stimulation, TNF transcripts are stabilized and their translation is substantially enhanced; this post-transcriptional switch is associated with TNF tail lengthening (Crawford et al. 1997). Importantly, cytoplasmic adenylation has been implicated as the mechanism of TNF tail length control (Crawford et al. 1997), suggesting that a transcript-specific cytoplasmic poly(A) polymerase may engage in the regulation of macrophage activation. Intriguingly, the LPS-induced acute immune response is marked by rapid, short-term expression of inflammatory cytokines, followed by their rapid inhibition. This rapid shutdown of pro-inflammatory genes is critical for the prevention of chronic inflammation, and post-transcriptional inhibition by the RNA-binding protein ZFP36 is one component of this phenomenon (Tiedje et al. 2016; Mukherjee et al. 2014). ZFP36 guides pro-inflammatory mRNAs for degradation and translational repression, by recruiting deadenylation complexes to the poly(A) tail of target mRNAs (Sandler et al. 2011; Brooks and Blackshear 2013). These studies suggest that macrophage activation can be a model system to examine poly(A) tail length dynamics in a post-embryonic context. However, TNF has been the only example of poly(A) tail length control during the macrophage immune response, and several questions remain unsolved: are there other transcripts regulated by poly(A) tail control, and what are the implications and consequences of poly(A) tail length control during the macrophage immune response? Finally, and perhaps most interestingly, what is the extent and the importance of readenylation during macrophage activation?

In this study, we characterize poly(A) tail dynamics across a time-course using a cell-line model of human macrophage activation. We apply a combination of transcriptome-wide methods to profile nascent RNA synthesis (Kwak et al. 2013), poly(A) tail lengths (Woo et al. 2018), mRNA levels and 3'UTR isoform preferences (Fu et al. 2011). These methods generate accurate profiles of poly(A) tail length dynamics with 3' isoform resolution, allowing a comprehensive understanding of post-transcriptional regulation in macrophages during a time-resolved immune response. We discover widespread and complex patterns of regulation mediated, in part, by changes in poly(A) tail length. We find evidence of extensive poly(A) tail lengthening, which is most pronounced in immune-related genes and factors involved in post-transcriptional regulation itself. Notably, our data suggests that readenylation of ZFP36, along with other mRNAs bound by ZFP36 protein, is an important early event during macrophage activation. Interestingly, these ZFP36-interacting transcripts show rapid tail shortening later in the response, likely as a consequence of elevated ZFP36. Thus, readenylation first stabilizes a set of genes implicated in macrophage function, and then results in their inhibition. Taken together, we show that macrophage activation entails extensive post-transcriptional regulation involving poly(A) tail length control.

2.3 Results

2.3.1 Determination of isoform-specific poly(A) tail lengths in THP-1 cells

To study poly(A) tail length (PAL) control upon macrophage activation, we stimulated differentiated human THP-1 cells with LPS, and followed up the resulting acute inflammatory stage of the immune response over a 4-hour time course (**Figure 2.1A**). Induction of *TNF* and *IL1B* confirmed proper activation (**Figure 2.1B**). We used Tail End Displacement sequencing (TED-seq) to measure PAL transcriptome-wide prior to

stimulation (0 hours), and at three subsequent time-points (1, 2 and 4 hours) with biological replicates. TED-seq estimates PAL by an accurate size selection of the sequencing libraries, which include the poly(A) tail region (**Figure 2.1C** top, **Figure S2.1A**; Woo et al. 2018). We sized libraries at 300 nucleotides (nt), thus, PAL is derived by subtracting the distance between the 5' end of TED-seq reads and the 3' cleavage and polyadenylation site (PAS) from 300 nt. The 3'UTR isoform-specific poly(A) tail lengths is intuitively visualized: for transcripts with longer poly(A) tails, TED-seq reads map closer to the PAS, whereas those with shorter tails map further from the PAS and into the 3'UTR. The collection of reads derived from each 3'UTR isoform generates a clustered distribution of poly(A) tail lengths for that isoform (**Figure 2.1C** bottom). For example, *SPSB1*, an interferon-stimulated gene expressed preferentially in macrophages, has two annotated 3'UTR isoforms. In unstimulated THP-1 cells, TED-seq reads mapped to *SPSB1* indicate distinct distributions of poly(A) tails for both 3'UTR isoforms, with mean poly(A) tail lengths of 91 nt and 94 nt, respectively (**Figure 2.1D**). TED-seq accuracy and precision was validated by four spike-in standards with different poly(A) tail lengths (40, 80, 120 and 160 nt), displaying sharp PAL distributions with expected median sizes (**Figure 2.1E**, and **S2.1B**).

It is recognized that reference PAS annotations in mammals are incomplete, and PAS usage is highly cell-type specific (MacDonald and McMahon 2010; Smibert et al. 2012; Zhang et al. 2020). Since the accuracy of TED-seq depends on correct PAS annotation, we experimentally determined PAS in the differentiated THP-1 cells. We performed 3'-seq (Fu et al. 2011) in two biological replicates using the same LPS stimulation time points examined by TED-seq (**Figure 2.1A**, **Figure S2.1C** and **S2.1D**). 3'-seq identifies PAS by initiating reverse-transcription at the start of the poly(A) tail, and the resulting read counts correspond to the abundance of the 3'UTR

isoform ending at the given PAS. After filtering out nonspecific, internally priming-derived reads, we identified 47,986 PASs, 95% of which (n=44,791) are located within annotated genes (n=12,336), and only 5% mapped to intergenic sequences (**Figure S2.1E**). Most (64%) intragenic PASs mapped to annotated 3'UTRs or within 1,000 nt downstream (**Table S1**). The remaining sites (**Figure S2.1E**) mapped to CDS (12%), intron (18%) or 5'UTR (1%) regions, proportions equivalent to those found in other contexts (Jia et al. 2017). 63% of genes with at least one site exhibited multiple PASs (**Figure 2.1F**). 33% of *de novo* PASs mapped within 10 nt of annotated PASs (**Figure 2.1G** and **S2.1F**), and 67% PASs are discrepant with annotated sites (33% within 10 to 300 nt, 35% more than 300 nt away from any annotated site; **Figure 2.1G**). These fractions indicate the prevalence of novel isoforms, which are similar to 3'-seq studies in other cell types (Zhang et al. 2020, Katsanou et al. 2005). For example, *ANTXR1* uses *de novo* PAS at 160 nt upstream of the reference PAS (**Figure 2.1H**), whereas *CD83* exhibited two novel tandem 3'UTR isoforms instead of the single annotated PAS (**Figure 2.1I**). These results demonstrate the need to establish cell-type specific PAS usage for studies where comprehensive 3'UTR isoform annotations are required, such as TED-seq. Therefore, we used our experimentally determined PAS sites in THP-1 cells instead of annotated PAS in TED-seq analysis. We generated a customized annotation of 30,141 3'UTR isoforms in 10,589 genes (**Table S2**), which were used to calculate isoform specific PAL profiles. Biological replicates of TED-seq at each time point correlated well (**Figure S2.1G**; Pearson correlation coefficient, R=0.99), and we used mean poly(A) tail lengths from two replicates for subsequent analyses.

We used our 3'-seq data (**Table S3**) to assess whether 3'UTR usage changed across the activation time-course (**Figure S2.2A** and **S2.2B**). Genes often contain multiple PAS in their 3'UTR regions, generating alternative 3'UTR isoforms with different 3'UTR

lengths through alternative polyadenylation (APA; Tian et al. 2007, Mayr et al. 2016, Mayr et al. 2009). The degree of APA isoform usage between any two time points was assessed using the 3'UTR switch index (USI; Harrow et al. 2012). Shift toward distal APA (longer 3'UTR) isoforms result in positive USI values (USI >0.1) and proximal APA (shorter 3'UTR) isoforms for negative USI values (USI < -0.1, **Figure S2.2B**). Our data show that LPS stimulation induces extensive changes in isoform usage, for both distal (n=566, FDR<0.1) and proximal (n=464, FDR<0.1) switches, with a gradual increase in the proportion of proximal switching (**Figure S2.1D** and **S2.2C** left) over the time course. Notably, the functions of genes that exhibit 3'UTR isoform switching are enriched in immune responses, metabolic processes and protein transport/localization (**Figure S2.2C** right). Thus, macrophage activation involves extensive changes in 3'UTR isoform usage, which are potentially relevant to physiological changes during macrophage activation.

2.3.2 PAL Dynamics during Macrophage Activation

Global poly(A) tail profiling studies revealed that many human mRNAs have mean PALs between 50-100 nt, shorter than thought previously (Jalkanen et al. 2014; Subtelny et al. 2014; Chang et al. 2014). In addition, modest tail length changes of 10-20 nt can impact RNA fates and thus be consequential (Eisen et al. 2020a; Jalkanen et al. 2014). Therefore, we strove to ensure that our PAL calculations from TED-seq data were accurate and high resolution. Shifts in APA isoform preferences occurring within the 300 nt library-sizing window (referred to as local PAS switch hereafter) complicate determinations of which PAS the TED-seq reads derive from. To exclude such potential errors we stringently removed transcript isoforms that showed LPS-induced local PAS switches within a tandem PAS cluster (multiple PAS isoforms within 300 nt). This leaves us with 6,269 major isoforms in 5,079 genes, which

corresponded to ~70% of TED-seq reads that can be unambiguously assigned to PASs (**Figure S2.2D** and **S2.2E**). We calculated PALs for these major isoforms and the changes in PALs between end points, linear time-points, as well as any two time point comparisons (0 h to 4 h, 0 h to 1 h, 1 h to 2 h, 2 h to 4 h, 0 h to 2 h, and 1 h to 4 h intervals) to comprehensively identify isoforms with tail length changes, and to resolve transient changes (**Table S4**). This approach was necessary to capture transient PAL changes (*e.g.* between 0 h and 2 h). Thus, we generated an inclusive view of PAL dynamics during macrophage activation, identifying 1,520 transcript isoforms with PAL changes (length differences ≥ 10 nt) in at least one interval comparison (**Figures 2.2A** and **S2.2F**, illustrating the 0h to 1h comparison and all other comparisons, respectively). Transcripts from 237 genes exhibited PAL increases during the time-course, although the majority ($n=1,286$) of transcripts with significant changes (K-S test, $FDR < 0.1$) exhibited PAL decreases. Notably, the tail length changes were independent of initial tail lengths (**Figure S2.3A**). To validate these results, we selected examples of transcripts with PAL increases (*CCL4*, *CCRL2*, and *ZFP36*) and no PAL changes (*HNRNPF* and *ACTB*), and used a PCR based poly(A) tail (PAT) assay to validate the results (**Figure 2.2B**).

TNF had been proposed to undergo readenylation upon stimulation in mouse macrophage cells, exhibiting a PAL increase one hour post activation (Crawford et al. 1997). Consistent with previous studies, LPS stimulation caused a shift in TED-seq reads upon stimulation (0-1h), indicating a PAL increase in *TNF* (**Figure 2.2C**).

However, our extended time-course data revealed that this increase is transient and followed by a rapid reduction in tail length after one hour. Moreover, this increase was not due to any local PAS switching, since 3'-seq shows only one single dominant 3'-seq peak (98% of reads from the major isoform) in *TNF* throughout the time course (**Figure 2.2C**; Proportion of a PAS isoform over all 5 isoforms [PPI] ≥ 0.98). *TNF*

inhibition after a transient activation is known to be integral to the macrophage response (Carballo et al. 1998). Our TED-seq data, generated at multiple time-points, revealed more complex changes in *TNF* PAL than previously appreciated.

To characterize temporal PAL dynamics in the transcriptome, we applied k-means clustering to identify coordinated PAL change patterns across genes (**Figure 2.2D**, left; n=1,520). The majority of genes are in clusters 1 to 3 (C1-C3), each of which show gradual PAL decreases. Other clusters exhibit PAL increases (C4-C6), revealing diverse and distinctive temporal patterns (**Figure 2.2D**, left). C4 and C5 undergo early PAL increases (within 2 h), followed by rapid or gradual decreases in PAL, respectively. In contrast, C6 exhibits persistent PAL lengthening later in macrophage activation. Notably, C7-C9 show early decreases in PAL, which is reversed to initial or longer lengths. These data reveal that PAL regulation during macrophage activation is more widespread and complex than previously recognized.

To gain insights into the biological relevance of PAL changes, we examined gene ontology (GO) enrichments in each cluster (**Figure 2.2D**, right). Genes in C1-C3, with gradual PAL decreases, are enriched in transmembrane proteins (C1), phagocytosis factors (C2) and oxidoreductases (C3). These GO terms may reflect reduced requirements for the corresponding products during the inflammatory response. More interestingly, genes in C4-C6, characterized by PAL increases upon activation (**Figure 2.2D-2.2F**), are enriched with immune-related terms such as cytokines, chemokines, and chemotaxis (**Figure 2.2D** right). These enrichments imply a role of PAL control during the inflammatory process. In particular, enrichment of pro-inflammatory genes in C4, characterized by early transient increases in PAL, is notable given that immediate expression of pro-inflammatory genes is integral to early macrophage activation (Carpenter et al. 2014; Corbett 2018). Within C4, we also observed enrichment of 3'UTR binding proteins, including ZFP36 and ELAVL1, factors known

to regulate *TNF* (Mukherjee et al. 2014, Tiedje et al. 2012, Katsanou et al. 2005; highlighted in red, **Figure 2.2E**). Moreover, RNA binding proteins (RBPs) are also enriched in C8 and C9. Thus, mRNAs with extensive PAL changes during activation encode genes important to immune function and *trans*-factors related to the poly(A) tail itself.

2.3.3 3'UTR sequence features associated with changes in PAL

Poly(A) tail length control is often mediated by interactions between 3'UTR *cis* elements and *trans*-acting factors. AREs are well characterized 3'UTR *cis*-elements that mediate rapid decay of many short-lived mRNAs, in particular, of cytokine and chemokine transcripts during immune responses (Caput et al. 1986; Xu et al. 1997). ARE-binding RBPs such as ZFP36 recognize AREs and destabilize mRNAs by recruiting deadenylation and decay factors (Lai et al. 2003; Sanduja et al. 2011). To gain insights into possible relationships between AREs and PAL changes, we first examined the association between 3'UTR A/U content and LPS-induced PAL changes. We classified transcript isoforms into three groups: those with increases in PAL ($\Delta\text{PAL} \geq 10$ nt), those with decreases ($\Delta\text{PAL} \leq -10$ nt), and those with little or no change ($|\Delta\text{PAL}| \leq 5$ nt). We then compared 3'UTR AU content across the three ΔPAL groups (**Figure 2.3A**, left). 3'UTRs of PAL-decreased transcripts (0 h to 4 h) have lower AU content compared to PAL-unchanged and -increased transcripts ($P < 10^{-8}$; K-S test). This association between low AU content and tail shortening was observed for all time intervals except for the 1 h to 2 h interval (**Figure 2.3A** right and **Figure S2.3B and S2.3C**). We also examined the association between ΔPAL and other potentially relevant features, such as 3'UTR length and codon optimality, which revealed that 3'UTR AU content is the most strongly correlated feature with ΔPAL (**Figure 2.3A** right and **Figure S2.3D**). In parallel, we performed 6-mer enrichment

analysis to search for sequence motifs enriched in the 3'UTRs of PAL-increased transcripts, which revealed that AU-rich 6-mers are enriched in the 3'UTRs of PAL-increased genes (Student's t-test FDR<0.1) (**Figure S2.3E**).

There is growing evidence that 3'UTR *cis*-elements are functionally sensitive to their location within the 3'UTR (Geissler and Grimson 2016; Geissler et al. 2016; Grimson et al. 2007; Piqué et al. 2008; Dai et al. 2019). Therefore, we examined how the association of AU content with poly(A) tail length change depends on the relative location of AU content within the 3'UTR, which revealed that AU content is enriched near the 3' ends of 3'UTRs of PAL-increased transcripts compared to PAL-decreased ones (**Figure 2.3B** and **S2.3F**). These results suggested an association between changes in poly(A) tail length and AREs located towards the 3' terminus of the 3'UTR. To test this interpretation, we examined the association between individual 6-mers (n=4,096) and Δ PAL in the first and the last 500 nt segments of 3'UTRs, for 3'UTRs longer than 1 kb. For each 6-mer, we binned transcripts into 4 quartiles by the frequency of each 6-mer in each 500 nt segment, and then compared Δ PAL values between the top and bottom quartiles (Student's t-test, FDR <0.1). AU-rich 6-mers are more frequently associated with PAL changes when located in the last 500 nt compared to the first 500 nt of the 3'UTRs (**Figure 2.3C**). Moreover, for those 6-mers significantly associated with Δ PAL (Student's t-test, FDR <0.1; **Figure 2.3C**), we obtained $\Delta\Delta$ PAL, the difference in mean Δ PAL between the top and bottom quartiles. Positive $\Delta\Delta$ PALs indicate that the 6-mers promote PAL increases, and those with negative $\Delta\Delta$ PALs are associated with PAL decreases. The AU-rich 6-mers tend to have $\Delta\Delta$ PAL > 0 (**Figure 2.3D**). In addition, we repeated the 6-mer analysis using alternative smaller terminal segments, and obtained consistent results using the last 300 nt segments (**Figure S2.4A**). The last 100 nt segments showed less clear position effect, suggesting that PAL-controlling *cis* elements may work in a wider 3' terminal

region of 3'UTRs. Collectively, these results demonstrate that 3'UTR AU content is a major feature associated with LPS-induced PAL changes and this association manifests near the 3' end of 3'UTRs.

3'UTR regulatory elements often work in combination to mediate poly(A) tail length control (Dai et al. 2019; Piqué et al. 2008). Therefore, we examined combinations of 6-mer elements associated with PAL changes during macrophage activation. We identified 6-mer pairs that may be functioning combinatorially by comparing the Δ PAL of transcripts harboring both 6-mers to transcripts containing two instances of either of the 6-mers. For every pair of 6-mers sampled from the 6-mers identified as individually associated with Δ PAL in the last 500 nt regions, we assessed whether the Δ PAL of the mRNAs containing both 6-mers (once each) are significantly greater than Δ PALs of the mRNA containing one type of the 6-mers twice. This strategy controls for the total number of the tested 6-mers in one mRNA. We found a total of 138 6-mer pairs that may act in combination to mediate PAL control upon macrophage activation in any time-point comparison (**Table S5** and **S6**). The 6-mers were assigned to known RBP motifs based on position weight matrix scores (Ray et al. 2013). For example, mRNAs with the 3' terminal co-localization of specific pairs of AU-rich motifs tend to undergo greater tail length increase during early stage of macrophage activation (between 0h and 1 h, **Figure 2.3E**). The RBPs corresponding to these 6-mer pairs include many previously established poly(A)-tail-associated proteins (e.g., ELAVL1, PABPC1, CREBs, TIA1, and ZFP36), but also some relatively uncharacterized proteins (e.g., IGF2B2, HNRNPs, SYNCRIP, SART3, U2AF2, and RALY). This result indicates the complexity of poly(A) tail length control, and functionally associates various novel proteins, which had little connection to poly(A) tail previously, with poly(A) tail length control.

2.3.4 Poly(A) tail length correlate with post-transcriptional changes

In post-embryonic systems, due to the presence of active and dynamic transcriptional regulation, assessing the relative role of post-transcriptional events is challenging. In particular, multiple mechanisms could explain poly(A) tail length changes in somatic cells upon stimulation. Our observations of poly(A) tail length increases in 237 transcripts (**Figure 2.2A** and **S2.2F**) could result from rapid accumulation of new transcripts from LPS-induced transcriptional burst, increased co-transcriptional polyadenylation, decreases in deadenylation, and/or cytoplasmic readenylation (**Figure S2.4B**). These possible mechanisms may apply differently to different transcripts or transcript isoforms (Kondrashov et al. 2012). We first examined whether the poly(A) tail length increases we observed derived from increased transcript synthesis, which are expected to have longer tails initially. Thus, we measured transcriptional activity genome-wide using a nascent RNA profiling assay, Precise-Run-On sequencing (PRO-seq; Kwak et al. 2013). PRO-seq profiles transcription activity genome-wide by performing a nuclear-run-on reaction with biotin-labelled nucleotides (biotin-NTPs). Incorporation of biotin-NTPs occurs at the 3' end of nascent RNAs, providing a molecular handle with which to selectively purify nascent RNAs and construct libraries for sequencing. We performed PRO-seq across the macrophage immune response (0, 1, 2, and 4 h upon LPS; two biological replicates) (**Figure S2.4C** and **Table S7**), and found that transcriptionally upregulated genes exhibited increases in poly(A) tail length compared to those with transcriptional downregulation (**Figure 2.4A**, **S2.4D**, and **S2.4E**). Thus, distinguishing the impact of *de novo* transcription from post-transcriptional events is essential to determine the degree of post-transcriptional poly(A) tail length control.

To determine the extent of post-transcriptional poly(A) tail length changes and the influence of such changes have on transcript abundance, we integrated our PRO-seq

data with our mRNA tail and abundance profiling data (TED-seq and 3'-seq). First, we selected transcript isoforms with minimal transcriptional changes (PRO-seq, $|\log_2$ Fold change (FC)| < 0.5), and from this set of genes, identified transcript isoforms with $|\Delta\text{PAL}|$ greater than 10 nt (FDR < 0.01; K-S test) between two time points. We included only genes with minimal transcriptional changes for all pairwise time intervals, and still identified transcripts that exhibited significant shifts in the poly(A) tail length distribution (**Figure 2.4B** and **S2.4F**), implying that their poly(A) tail length changes are post-transcriptional. Notably, the association between low AU content and tail shortening was also observed in the set of genes that exhibited minimal transcriptional changes (**Figure S2.3C**). Taken together, these results suggest that transcription alone cannot explain the observed poly(A) tail length dynamics and post-transcriptional events contribute to tail length regulation.

Poly(A) tail length regulation and its association with mRNA fate has been observed for several genes in various physiological contexts, including neuronal cells (Wells et al. 2001; Weill et al. 2012). More recently, transcriptome-wide tail length study showed that PAL changes (ΔPAL) correlate with changes in transcript abundance (ΔRNA), stability and translation efficiency during the endoplasmic reticulum stress response (Woo et al. 2018). However, it is unclear whether these relationships apply in macrophage activation. To characterize the association of poly(A) tail length changes (ΔPAL) with changes in RNA abundance (ΔRNA) independent of transcriptional changes (ΔTXN), we made robust control sets of genes, only including the ones with minimal ΔTXN . We selected genes with minimal ΔTXN (PRO-seq, \log_2 FC < 0.5) through the time-course (0 h through 4 h), and splitted them into three groups (Down, No change, Up) based on ΔRNA (3'-seq, \log_2 FC [4 h/0 h] threshold = 1). We then applied stratified random sampling to the gene groups to normalize the ΔTXN distributions; we split the genes in each group into 10 bins based on ΔTXN (PRO-seq,

\log_2 FC [4 h/0 h]), and sampled the same number of genes from each Δ TXN bin across the three Δ RNA groups (**Figure S2.4G**). This analysis revealed a significant relationship between Δ PAL and Δ RNA after precisely controlling for Δ TXN (**Figure 2.4C**), which holds true for other time point comparisons (**Figure S2.5A**, 0 h and 1 h comparison shown as a representative). The association was also tested after grouping genes based on Δ PAL and examining Δ RNA (**Figure S2.5B**).

To consider these relationships further, we examined genes across the entire range of transcriptional changes. We divided all genes into three Δ RNA groups, as described above, and applied stratified sampling to maintain the same distribution of Δ TXN across the groups, by splitting the genes into 6 Δ TXN bins. We observed a significant positive association between Δ RNA and Δ PAL for all time point comparisons (0 h vs 1 h and 0 h vs 4 h comparison shown as a representative, **Figure 2.4D** and **S2.5C**).

This trend was also observed when genes were binned by Δ PAL and then Δ RNA assessed (**Figure S2.5D**). Collectively, these results indicate that during macrophage activation, post-transcriptional events couple changes in tail length to RNA abundance under conditions of extensive changes in transcriptional regulation.

To further dissect the post-transcriptional relationships between poly(A) tail dynamics and RNA abundance, we used the Δ RNA/ Δ TXN metric, which approximates mRNA stability (Patel et al. 2020; Woo et al. 2018; Blumberg et al. 2021), and explored the relationship between Δ PAL and the Δ RNA/ Δ TXN in our time-resolved data. This analysis revealed that Δ PAL has a strong positive association with Δ RNA/ Δ TXN, only when Δ PAL was from the preceding interval than Δ RNA/ Δ TXN (e.g. 0 h vs 2 h for Δ PAL and 2 h vs 4 h for Δ RNA/ Δ TXN; **Figure 2.4E** right), but not when both were from the same interval (e.g. 0 h vs 2h for Δ PAL and Δ RNA/ Δ TXN both; **Figure 2.4E** left). These results suggest that during macrophage activation, changes in mRNA stability is coupled to changes in poly(A) tail lengths in a temporally delayed manner.

Given the evidence that poly(A) tail length might control RNA abundance during macrophage activation, we investigated if tail control can affect opposite influences of transcription on mRNA dynamics. We selected genes that are transcriptionally up-regulated ($\log_2 \text{FC}([1 \text{ h}, 2 \text{ h and } 4 \text{ h}] / 0 \text{ h}) > 1$), and compared how the changes in poly(A) tail lengths (ΔPAL , 0 h to 2 h) are associated with the $\Delta\text{RNA}/\Delta\text{TXN}$ metric (**Figure 2.4F** and **S2.5E**). The transcriptionally up-regulated transcripts did not exhibit changes in the mRNA stability metric between 0 h and 2 h regardless of changes in tail length (**Figure 2.4F**, left panel). However, when we considered changes in the mRNA stability metric during the 2 to 4 h interval, changes in stability ($\Delta\text{RNA}/\Delta\text{TXN}$) occurred in the same direction as the ΔPAL from the earlier time interval (**Figure 2.4F**, right panel). In particular, mRNA levels at later time-points are reduced in genes with decreased tail length (0 h to 2 h) despite increases in transcript synthesis (**Figure S2.5E**, see mean (RNA) in PAL:DN). Collectively, these analyses demonstrate that change in mRNA abundance coupled to PAL shortening may override the influence of increased transcription, and post-transcriptional control is evident even in genes under active and opposing transcriptional control. Additionally, we also examined genes that were transcriptionally repressed during the time-course (**Figure 2.4G** and **S2.5F**). Genes that exhibited reductions in tail length have greater decreases in RNA stability ($\Delta\text{RNA}/\Delta\text{TXN}$) than those with no PAL change, and their difference manifested at later time interval than the same time interval (**Figure 2.4G** and **S2.5F**). Together, these results indicate that poly-(A) length control mediates significant and widespread impacts on transcript abundance during macrophage activation.

2.3.5 Profiling readenylation during initiation of the macrophage immune response

TNF is thought to be regulated through cytoplasmic polyadenylation in mouse macrophages (Crawford et al. 1997), although this readenylation hypothesis has not been examined in human macrophages. Moreover, this phenomenon has not been examined at a transcriptome-wide level during macrophage activation. The added complexity of dynamic transcriptional regulation during macrophage activation further complicates this question. Therefore, we repeated our TED-seq profiling after inhibiting transcription with actinomycin D (ActD), which was performed prior to LPS stimulation (**Figure 2.5**). The inhibition of transcription by ActD was confirmed by a bulk poly(A) tail length assay that revealed a global shortening of poly(A) tails after ActD treatment (**Figure S2.6A**; (Kojima and Green 2015)). In addition, for selected genes known to be induced by LPS stimulation, we verified by qRT PCR that ActD treatment was sufficient to negate induction (**Figure S2.6B**). Then we generated TED-seq libraries from ActD-treated cells, at 0, 1 and 2 hours post-LPS activation, constructing a pair of biological replicate libraries for each time point. The resulting poly(A) length profiles were well-correlated between replicates (**Figure S2.6C**). Quantitative analysis of the suppression of known LPS induced transcripts in the ActD TED-seq replicates also showed at least 95% suppression by ActD on average. Upon ActD-treatment, we expect PAL increases only for mRNAs targeted by cytoplasmic polyadenylation. This readenylation is necessary and sufficient to explain transcription independent PAL increases without *de novo* synthesis of longer PAL transcripts. PAL changes were quantified for the PAS isoforms (n=6,876 derived from 5,609 genes) that passed the cutoff of 50 TED-seq reads across all time points (0, 1, 2 h). Following ActD treatment, TED-seq identified tail length increases for *TNF*, with Δ PAL values of 8.6 nt (10.6 and 6.6 nt respectively in each of the replicates) (**Figure 2.5A**). This increase was validated with PAT assays under ActD treatment (**Figure 2.5B**), in which we included *in vitro* deadenylated controls (**Figure S2.6D**). As a

negative control for the PAT assay, we also included *IL1B*, which did not exhibit any change in poly(A) tail length in TED-seq (**Figure 2.5B** and **S2.6D**). In our normal LPS time-course TED-seq data, we observed significant overlap of PAL increase genes between biological replicates using a smaller cut-off of $\Delta\text{PAL} > 5$ ($p < 10^{-8}$, fisher exact tests), and therefore applied this threshold to the ActD-LPS data to identify readenylation targets with a higher sensitivity. Transcripts that exhibited $\Delta\text{PAL} > 5$ across both biological replicates (FDR < 0.2 ; K-S test) include *TNF* and 266 potential readenylation targets (n=61, 166 and 86 for 0 h vs 1 h, 0 h vs 2 h, and 1 h vs 2 h comparisons, respectively) out of 6,876 transcript isoforms considered upon macrophage activation, corresponding to 255 (of 5,609) genes (**Figure S2.6E**, **Table S8**). The majority of the potential readenylation targets have intermediate tail lengths, and only 6 transcripts had very short tails (< 25 nt) that may be the targets of oligouridylation (Chang et al. 2014; Lim et al. 2014) (**Figure S2.6F**). Moreover, the potential readenylation target transcripts exhibited tail length increases averaging 14 nt (**Figure S2.6G**), and there was no relationship apparent between tail increase and starting tail length (**Figure S2.6H** and **S2.6I**).

To examine the potential impact of readenylation, we investigated the association between readenylation and changes in RNA abundance (ΔRNA , \log_2 FC) upon LPS stimulation. Potential readenylation targets ($\Delta\text{PAL UP}$, ActD-treated) exhibited greater ΔRNA (3'-seq, ActD-untreated,) compared to non-target genes (CTRL) at 0-1 h and 1-2 h, but not at 0-2 h (**Figure 2.5C** and **S2.7A** right), after stratified random sampling to equalize transcription change (ΔTXN ; PRO seq, \log_2 FC; **Figure S2.6J** and **S2.7A** left). These results were also observed using the $\Delta\text{RNA}/\Delta\text{TXN}$ metric (\log_2 transformed [$3'$ -seq FC/ PRO-seq FC], **Figure 2.5D**). Additionally, we confirmed the association between readenylation and RNA abundance, even when restricting the analysis to the subset of genes with minimal changes in transcription (PRO-seq, $|\log_2$

$FC|<0.5$, **Figure S2.7B**). Collectively, these results implicate readenylation as a process responsible for stabilizing transcripts during macrophage activation. To identify regulatory sequences involved in readenylation, we examined the enrichment of any putative RBP motifs in the 3' terminal regions (500 nt) of 3'UTRs of the readenylation targets. To identify RBPs that mediate PAL increases during macrophage activation, we used position weight matrices (PWM; $n=202$) from the Cis-RBP database (Ray et al. 2013). For each RBP expressed in THP-1 cells ($n=86$), we calculated PWM scores in the last 500 nt of 3'UTRs of PAL increased transcripts, and assessed the occurrences over background. The top significant motifs include the sequences bound by *RALY*, *ZFP36*, *HNRNPC*, *CPEB2*, *ELAVL1* and *U2AF2*, all of which are characterized by poly(U) sequences (**Figure 2.5E**). Poly(U) sequence motifs were consistently enriched in the examination of the last 300 nt and the last 100 nt of 3'UTRs of the potential readenylatoin targets (**Figure S2.7C and S2.7D**), suggesting that poly(U) sequences are the readenylation control sequences upon LPS stimulation.

Additionally, to understand the role of post-transcriptional poly(A) tail length elongation (post-TXN Δ PAL UP), we performed gene ontology analysis (Huang et al. 2009) of potential readenylation targets, using all genes expressed at all time points as the background (**Figure 2.5F**). Target mRNAs of cytoplasmic polyadenylation were enriched with RNA binding terms, including AU-rich element binding, RNA binding, poly(A) RNA binding, and post translational modification (PTM) targets such as phosphoproteins, acetylation, and ubiquitin conjugation (Fisher's exact test, $FDR<0.1$). To a lesser significance (Fisher's exact test, $0.1\leq FDR<0.2$), immune-related terms such as viral process, NF-kappa B signaling pathway as well as cellular localization and intracellular transport were detected (**Figure 2.5F**). These observations suggest the RBP and PTM target proteins are prevalently regulated at a

post-transcriptional level through cytoplasmic polyadenylation at the early stage of macrophage activation, which is expected to exert broader secondary effects on their target mRNA/protein spanning their stability, translation, and possibly cellular localization.

2.3.6 Concomitant readenylation of ZFP36 and its target mRNAs upon early activation

ZFP36, also known as tristetraprolin (TTP), is an RBP that regulates mRNAs of proinflammatory genes to attenuate inflammation during macrophage activation (Brooks and Blackshear 2013). ZFP36 binding to AREs in target mRNAs, such as TNF, results in the recruitment and activation of deadenylase complexes and translational repression. However, the mechanisms underlying regulation of ZFP36 activity in early macrophage activation are not well understood, in particular at the post-transcriptional level. Intriguingly, in our TED-seq data, ZFP36 itself was identified as one of the potentially readenylated targets upon macrophage activation (**Figure 2.6A**). We validated ZFP36 readenylation by performing PAT assay in the ActD-treated condition (**Figure 2.6B**). Of note, comparing PAT assay results between with and without RNase H (lane 1, 2 and 4) indicates that *ZFP36* transcripts exist in very short poly(A) tail forms (near A0). Next, we tested whether these tail length increases are affected by mutating poly(U) stretch sequences in the 3'UTR of *ZFP36* mRNAs. We constructed eGFP reporter genes fused with human ZFP36 3'UTR sequences, either wild-type (WT) or with poly(U) regions mutated (MUT-DEL or MUT-GC), and expressed in the THP-1 cells (**Figure 2.6C**). The poly(U)-containing motifs in the 3'UTR of *ZFP36* mRNA were either deleted (MUT-DEL) or substituted with G and C nucleotides (MUT-GC). We used lentiviral transduction of these constructs in the THP-1 cells to avoid any notable immune induction, which was tested

using qRT-PCR (**Figure S7E**). The untransduced and 3 transduced THP-1 cells were PMA differentiated and LPS stimulated as described previously (**Figure 2.1A**). It should be noted that we performed PAT assays on the 4 THP-1 RNAs with and without LPS stimulation in the absence of ActD treatment based on the following reasons: first, we observed tail length increases near A0 in the LPS-treated condition, both with and without ActD treatment (lane 3 and 5, **Figure 2.6B**), indicating that *ZFP36* mRNA readenylation can be detected well enough even in the ActD-untreated condition. Second, given the potential side effects of ActD on mRNA regulations, validating readenylation in ActD-untreated condition provides a more natural, cellular context. Therefore, we performed PAT assays on the 4 THP-1 RNAs with and without LPS stimulation using a primer that can detect both endogenous and transgenic *ZFP36* poly(A) tails and focused on the tail length changes for short-tailed mRNAs. Tail length increased only in the endogenous and WT 3'UTR reporter mRNAs, but in none of the mutant 3'UTR reporter mRNAs upon activation (**Figure 2.6D** and **S2.7F**). Notably, the near-completely deadenylated products (A0, red asterisk band in **Figure 2.6D**) disappear in both endogenous and WT, but not in MUT samples. This indicates that even if the PAT assay is detecting both the endogenous and transgenic *ZFP36* poly(A) tail populations in the WT sample, the very short poly(A) tail forms from both populations disappeared and were converted to longer tail forms. The greater degree of the length increases in WT transgene compared to the endogenous sample in the gel quantification analysis also support this finding (**Figure 2.6D**, right panel). Taken together, these results demonstrate that *ZFP36* mRNAs undergo readenylation mediated by poly(U) readenylation control sequences in the 3'UTR during macrophage activation.

Intriguingly, It should be noted that both *TNF* and *ZFP36* were identified to undergo readenylation early upon activation in our data. These observations suggest that

readenylation rapidly induces TNF expression, but it becomes transient through readenylation of *ZFP36* mRNAs and the resulting induction of ZFP36 protein, which then negatively regulates *TNF*. In support of this model, we found that levels of ZFP36 exhibited a robust increase at both the total protein level and for dephosphorylated forms at 1 h upon LPS stimulation compared to the non-treated condition (**Figure 2.7A**). Dephosphorylated ZFP36 are known to act preferentially on their target mRNAs (Chrestensen et al. 2004; Stoecklin et al. 2004). Therefore, we examined whether this readenylation model explains early transient expression of a larger set of ZFP36-targeted mRNAs including *TNF*.

First, we asked whether there are other ZFP36-targeted mRNAs in our readenylation candidates. Our RBP binding motif enrichment analysis revealed enrichment of ZFP36 motifs in the 3'UTRs of readenylation target mRNAs (Fisher's exact test, FDR <0.2). To confirm this association, we turned to existing ZFP36 iCLIP data in bone marrow derived macrophage (BMDM) cells upon 1hr LPS treatment (Tiedje et al. 2016). Transcripts undergoing rapid (0 to 1 hour) PAL increases in our data are associated with *in vivo* ZFP36 binding in BMDM cells, compared to transcripts with no changes in PAL (K-S test P<0.01, **Figure 2.7B**). In addition, putative ZFP36 binding motifs (Fisher's exact test, FDR <0.2, **Figure 2.7C left**) and *in vivo* ZFP36 binding sites (K-S test P <10⁻⁴, **Figure 2.7C right**) are strongly enriched in the set of transcripts with LPS-induced PAL increases, implying that poly(A) tails of ZFP36 targeted mRNAs are elongated. In addition, transcripts with increased PAL tend to have higher ZFP36 motif densities (**Figure S2.7G**). Collectively, these analyses suggest that ZFP36 mRNA itself and ZFP36-targeted mRNAs undergo readenylation together at the early stage of macrophage activation.

Next, we investigated the tail length dynamics of the ZFP36-target mRNAs. Our readenylation model predicts that ZFP36-targeted mRNAs exhibit early and transient

tail lengthening during macrophage activation, due to early readenylation followed by ZFP36 mediated deadenylation. Indeed, using the ZFP36 iCLIP data from BMDM cells (Tiedje et al. 2016), we found that transcripts bound by ZFP36 were specifically enriched in a set of transcripts characterized by early transient increases in PAL and GO terms associated with proinflammation (Cluster 4 in **Figure 2.2D**, Wilcoxon test $P < 10^{-6}$; **Figure 2.7D**). This finding was corroborated further using ZFP36 bound mRNAs identified using HITS-CLIP data from activated CD4⁺ T cells (Wilcoxon test $P < 10^{-14}$, **Figure S2.7H**; Stoecklin et al. 2004). Altogether, these observations suggest that *ZFP36* readenylation upon macrophage activation plays a key role in shaping the transient expression of proinflammatory genes in macrophages through immediate deadenylation by the rapidly induced ZFP36 protein, thus crucial for preventing hyperinflammation (**Figure 2.7E**).

2.4 Discussion

In this study, we set out to examine the prevalence and consequences of poly(A) tail regulation in a somatic-cell context. We selected a model of macrophage activation, in which we examined transcription, RNA abundance and poly(A) tail length in unstimulated cells and across a time-course following LPS stimulation, enabling us to study tail dynamics in a complex regulatory environment. Importantly, our approach enabled us to profile the tail with 3'UTR isoform resolution. We found extensive regulation of transcript abundance associated with poly(A) tail control. In response to activation, many transcripts exhibited tail lengthening, associated with increased transcript abundance. These transcripts preferentially encoded proteins associated with immune function and trans-acting factors that function in post-transcriptional regulation.

2.4.1 Poly(A) tail length dynamics upon macrophage activation

Prior to our work, TNF has been the sole example of a transcript known to be regulated by poly(A) tail control during macrophage activation (Crawford et al. 1997). Our study revealed that in addition to TNF, more than a thousand transcripts undergo poly(A) tail length changes during the macrophage immune response. Many are likely to be regulated by deadenylation, a well-established mode of control. However, several hundred transcripts appear to be readenylated, greatly expanding the scope of such regulation from TNF alone, and implicating readenylation as a major mode of control during macrophage activation.

One challenge in studying poly(A) tails in transcriptionally active, non-steady state systems is the difficulty in discriminating tail changes mediated post-transcriptionally from those derived from nascent transcription upon cellular activation. Our approach was to use PRO-seq to quantify transcriptional changes, and thus discriminate between transcriptional and post-transcriptional inputs on tail length. This approach was accomplished by stratifying genes by their transcriptional state: first, we examined genes with stable transcription, and second, we examined if genes with changes in transcriptional status also exhibit changes in RNA abundance as a function of poly(A) tail status. Thus, by normalizing transcriptional inputs, our integrative analysis demonstrated that transcriptional change alone does not explain changes in poly(A) tail status. Most importantly, changes in tail length and RNA abundance are correlated regardless of transcriptional change, indicating that post-transcriptional regulation is a major component of overall gene regulatory changes during macrophage activation. This study was designed to examine the impact of poly(A) tail dynamics during a rapid cellular response in a differentiated cell, here, macrophage activation; the temporal relationships between changes in transcription, transcript abundance and the

status of the poly(A) tail revealed a global preference for tail length changes that preceded changes in RNA abundance. This observation implies that post-transcriptional tail changes influence mRNA stability, and most mRNAs do not decay immediately upon deadenylation, but undergo intervening rate-limiting steps. We further reveal that tail length control can override transcriptional influences on RNA abundance. For many transcripts, their abundance is correlated primarily to rapid and transient changes in tail length. These results support a view of post-transcriptional control as a major component of gene regulation, even for genes under extensive transcriptional control, a potentially important phenomenon during rapid transitions.

2.4.2 Basis for rapid and transient poly(A) tail lengthening

Notably, we report widespread transcription-independent poly(A) tail lengthening, including for TNF. The simplest and most likely explanation is that such transcripts undergo readenylation. This interpretation relies on our transcription inhibition experiments, yet we acknowledge that such experiments may have limitations deriving from secondary effects, and from technical limitations inherent to the complexity of the experiment, that is LPS stimulation concomitant with transcriptional inhibition. Future work using approaches such as RNA metabolic labeling will be needed to confirm these striking results. Nevertheless, the rapid and transient nature of macrophage activation and the timing of poly(A) tail changes allow us to reason that they are not the secondary effects of transcription inhibition. Moreover, the common 3'UTR features in these target transcripts suggest that the rapid and transient transcript readenylation has a post-transcriptional mechanistic basis. This observation is particularly meaningful because the extent of post-transcriptional readenylation has been less understood in somatic cells, and our transcription-independent experiments

addressed a major challenge in the field by distinguishing polyadenylation from reduced deadenylation of de novo transcripts.

Interestingly, transcripts encoding RNA binding proteins (RBPs) often exhibited changes in poly(A) tail length; indeed, many encode proteins involved in post-transcriptional regulation by 3'UTR cis elements. Thus, a regulatory network linked through post-transcriptional control may play a significant role in macrophage activation. We identified factors associated with changes in the poly(A) tail using 3'UTR RBP binding inferences and CLIP-seq, converging to AU and U rich sequences, and their corresponding trans-factors. We elucidated widespread readenylation during macrophage activation in ZFP36 bound transcripts. We also found pronounced enrichment of poly(U)-containing RBP motifs in the 3'UTRs of readenylated mRNAs, which we tested in CMV-eGFP-3'UTR reporter assays. In yeast, 3'UTR poly(U) sequences are known to protect mRNAs from deadenylation (Muhlrad and Parker 2005), which is functionally relevant to our findings in human cell. In human, CPEB1 mediates cytoplasmic polyadenylation by binding to poly(U)-containing cis elements called cytoplasmic polyadenylation elements (CPE) during oocyte maturation (Hake and Richter 1994), and host-viral mRNA readenylation in cytomegalovirus infection (Batra et al. 2016). However, CPEB1 is not expressed in THP-1 cells. Thus, other non-canonical factors likely engage in readenylation during macrophage activation. Notably, ELAVL1 is expressed in THP-1 cells, and its binding motif is among the most enriched elements in transcripts undergoing tail lengthening. ELAVL1 stabilizes mRNAs, but whether this is mediated through readenylation is unknown (Charlesworth et al. 2013). Further efforts will be needed to test this and the relative contribution of ZFP36 and ELAVL1 or other factors to poly(A) tail dynamics during macrophage activation.

The identity of the polymerase responsible for readenylation during macrophage activation is an important question arising herein. In humans, multiple non-canonical poly(A) polymerases (TENT1-TENT6) exist (Liudkovska and Dziembowski; Warkocki et al. 2018), all relatively uncharacterized but for TENT2, a well-known poly(A) polymerase responsible for maternal mRNA polyadenylation. Based on our RNA-seq data, only TENT2, TENT4A/B and TENT5A transcripts are expressed in THP-1 cells. As TENT4 recruitment is mediated by 3'UTR structure rather than U-rich elements (Warkocki et al. 2018, Kuchta et al. 2016), we propose TENT5A as a candidate responsible for readenylation, recruited by a probable interaction with ELAVL1 (bioGRID interactome; Stark et al. 2006). ELAVL1-mediated recruitment of TENT5A during macrophage activation is a testable model to explain the widespread changes in post-transcriptional poly(A) tail lengthening and ensuing consequences on transcript abundance.

In our study, we also tested the role of poly(U)-containing motifs in poly(A) tail length increase upon macrophage activation by performing CMV-eGFP-3'UTR reporter assays, where we found some aspects to be improved for a more robust experimental validation. It should be noted that we couldn't design a PAT forward primer distinguishing human ZFP36 WT 3'UTR reporter mRNA from the endogenous ones in the human macrophage THP-1 cells due to their complete sequence homology.

Nonetheless, comparing the PAT assay result between original and WT reporter THP-1 cell lines showed that tail length increase appeared greater in the WT reporter-expressing THP-1 cells than the original THP-1 cells, which we think is reliably indicative of the tail length increase for WT reporter mRNAs (biological replicates n=2). We also found that plasmid transfection itself, but not viral infection, caused considerable immune induction in the macrophage cells, which might have a confounding effect on the tail length changes and thus suggests viral integration over

transfection in expressing reporter mRNAs. At this point, we only performed this assay with CMV promoter, but this result can be corroborated using a Tet-off promoter system.

In summary, we integrated multiple techniques to provide an overview of signal-activated poly(A) tail length control in a post-embryonic system. Poly(A) tail length has impacts on diversifying RNA dynamics, and the underlying mechanisms, including readenylation, are significant modes of rapid modulation of gene expression. Furthermore, we shed light upon control mediated by RNP dynamics and the consequences of these dynamics on the poly(A) tail, in particular between ZFP36-ARE mediated mRNA deadenylation and decay, and poly(U)-mediated readenylation and stabilization. We propose that readenylation facilitates rapid induction of ZFP36 targeted mRNAs and ZFP36 mRNA itself, which is crucial for short term response of macrophage immune activation. We conclude that complex, widespread patterns of post-transcriptional poly(A) tail control underlie the rapid and transient macrophage immune response. Here we have provided genomic analysis-driven evidence for the mechanistic models underlying poly(A) tail length change and macrophage immune response. The underlying mechanisms will be experimentally determined in future studies.

2.5 Materials and Methods

Cell lines, Cell culture and Compound Treatment

THP-1 cells used in this study are an authenticated cell line purchased from ATCC (TIB-202; human male). THP-1 cells were cultured at 37°C in 5% CO₂ in RPMI1640 (Gibco, 11875093) supplemented with 10% FBS (VWR) and 1% antibiotics (Gibco, 15240062). THP-1 cells were differentiated to macrophage-like cells by incubating them overnight in complete media containing 200 ng/ml PMA (Sigma Aldrich, P1585-1MG), followed by 3 days incubation in fresh media without PMA. The resulting differentiated cells were stimulated with 200 ng/mL LPS (Sigma-Aldrich, L4391) and collected at four time points: 0 hours post-stimulation (no stimulation), and 1, 2 and 4 hours post-stimulation. Total RNA was extracted with TRIzol (Invitrogen, 15596018). To inhibit transcription, THP-1 cells were incubated in media with 10 ug/mL Actinomycin D (Sigma Aldrich, A9415) for 15 minutes prior to stimulation with LPS.

3'-sequencing library preparation

Total RNA was extracted using TRIzol from differentiated THP-1 cells throughout the LPS stimulation time-course (0, 1, 2 and 4 hours). For each sample, poly(A) RNA was isolated from 10 ug of total RNA (Dynabeads™ mRNA Purification Kit; Invitrogen, 61006) followed by RNA fragmentation with 0.1 N NaOH, 5' RNA phosphorylation (NEB, M0201S), and 5' RNA ligation (NEB, M0204L) to VRA5 (5'-CCUUGGCACCCGAGAAUCCA-3'). After heat denaturation at 65°C for 2 minutes, 5' adapter-ligated poly(A)-containing RNA fragments were reverse transcribed by superscript II enzyme (Invitrogen, 18064-014) using RT primer (CPS_RTP: 5'-GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNNT8VN-3') at 50°C for 1 hour. The 3'-terminal ten nucleotides of CPS_RTP were designed to anneal to the junction between the poly(A) tail and the site of cleavage and polyadenylation

within the transcript, and also contain an eight-nucleotide (nt) unique molecular index (UMI) barcode for PCR deduplication, with the remaining sequence designed for PCR amplification (NEB, M0530L). The resulting cDNA molecules were amplified by PCR for 14 cycles with RP-1 primer (See **Table S9**) and RPI-X primers (See **Table S9**) using Phusion High-Fidelity DNA Polymerase (NEB, M0530), followed by gel purification of 200 to 500 bp products on a 6% PAGE gel in TBE buffer. PCR products were eluted from the excised gel in TE-TW buffer overnight at 37°C, and then filtered through a DNase-free spin X column (Costar Spin-X centrifuge tube filters; Corning, CLS8160) and purified using Ampure XP beads (Beckman Coulter, A63881). The purified, barcoded libraries were quantified and pooled prior to Illumina sequencing on a Next500 platform (75 bp single-end reads). Unless otherwise stated, enzymatic reactions were performed as described in the manufacturer's protocols. The 3'-sequencing libraries were prepared in two independent biological replicates.

3'-seq data pre-processing and mapping

5' RNA adaptor sequence was removed from the 3' end of sequencing reads using Cutadapt (Martin 2011) with option `-e 0.10, --overlap 2, --minimum-length=10, --nextseq-trim 20`. After adaptor removal, low quality reads were removed (those with quality scores <20, at any position). The first 30 nt, containing the 8 nt UMI, were used to deduplicate the reads (FASTX-Toolkit v0.0.11, http://hannonlab.cshl.edu/fastx_toolkit/, 2018; PRINSEQ v0.20.4, (Schmieder and Edwards 2011); Seqtk v1.3-r106, <https://github.com/lh3/seqtk>, 2012). After trimming 16 nucleotides (8 nt UMI and 8 nt corresponding to the dT8 portion of the adapter oligonucleotide) from the 5' end of each read, reads with at least 10 nt remaining were mapped to the human genome (hg38; UCSC), using STAR aligner (Dobin et al. 2013) with the option `--sjdbGTFfile "gencode.v26.annotation.gtf" --alignSJDBoverhangmin`

3, --outFilterMultimapNmax 1. The aligned reads were represented by their 5' end mapping coordinate on the opposite strand, and converted to BedGraph format, where the mapping position and the corresponding read counts of a 3'-seq peak were used to determine the cleavage and polyadenylation site and mRNA abundance of a transcript isoform.

Determination of cleavage and polyadenylation sites (PAS)

To avoid potential contamination of 3'-seq reads by the annealing of oligo-dT primers to internal A-rich sequences, we filtered the 3'-seq reads to remove internally primed reads from A-rich internal regions, as previously described (Fu et al. 2011; Li et al. 2012). Briefly, we searched for consecutive A sequences (>5 consecutive A nt) downstream of 3'-seq peaks, filtering out these reads from our 3'-seq reads. Then the 3'-seq read counts were normalized by counts per million mapped reads (CPM). The 3'-seq peaks were collapsed across all samples (0, 1, 2 and 4 h) with the read count per position totaled. Next, each 3'-seq peak position was converted to a 10 nucleotide-wide window, and the overlapping windows of 3'-seq peaks within the window merged, totaling the merged 3'-seq peaks, retaining the midpoint of the merged window as the PAS coordinate. Merged windows with five or more normalized reads (final PAS window) were retained. For each 3'-seq timepoint data, the read counts of the 3'-seq peaks mapped in a final PAS window were summed to represent mRNA abundance of the PAS isoform expressed in the given sample. All final PAS located in the reference (GENCODE V26)-annotated 3'UTR(s) + 1kb downstream region, were considered as distinct 3'UTR isoforms expressed in THP-1 cells. Finally, a custom transcript isoform annotation (bed12) file was built by modifying the reference transcript isoforms to terminate at our experimentally determined 3'-seq PAS sites. Poly(A) tail lengths were estimated for this comprehensive set of experimentally

determined PAS isoforms. Additional PAS analysis was performed as follows. The identified PAS sites (n=47,986) were tested for their locations (within gene or inter-genic) within annotated genes (n=12,336) using bedtools intersect function. Similarly, we then tested the positions of the PAS sites (n=44,791) located within genes against the genomic coordinates of annotated 3'UTRs or within 1,000 nt downstream of the 3' terminus (n=10,589), CDS, intron or 5'UTR. Multiple PAS were defined based on the genes with at least one mapped PAS sites (n=12,336) containing multiple PAS sites (n=7,760), or a single site (n=4,576). Discrepant PAS were defined by the distance between the mapped PAS and the annotated PAS, and binned by the distance: 10 to 300 nt (33%) or greater than 300 nt (35%). Based on this classification, we generated a customized reference transcript annotation file representing 30,141 3'UTR isoforms from 10,589 genes for downstream usage. Related to Figure 2.1 and Figure S2.1.

TED-seq library preparation

Tail End Displacement sequencing (TED-seq) was applied to total RNA samples (5 - 10 ug) collected at multiple time points (0, 1, 2 and 4 h) after LPS treatment. TRIzol-Purified RNA was subjected to poly(A) RNA purification using the manufacturer's protocol (Dynabeads™ mRNA Purification Kit, Invitrogen), and ligated with the adaptor molecule (RA3; Table S9) to their 3' terminus. The products of the ligation reaction were purified using TRIzol, and then fragmented with 0.1 N NaOH. Fragmented RNAs were purified with a P-30 column (Bio-Rad, 732-6251), and poly(A)-containing fragments enriched using Dynabeads mRNA purification kit. T4 polynucleotide kinase (PNK; NEB, M0201S) was used to phosphorylate the 5' terminus of RNA fragments, enabling ligation of the 5' terminus to the adaptor oligonucleotide containing UMIs (RA5; Table S9). The resulting RNA libraries were reverse transcribed and PCR amplified using KAPA HiFi HotStart ReadyMix PCR Kit

(Kapa Biosystems, KR0370), using no more than 8 cycles of amplification. Prior to sequencing, 350-360 bp DNA molecules were purified using PAGE. Following PCR amplification, PAGE-mediated size selection was repeated on the amplified DNA. The resulting size-selected libraries were pooled and sequenced on an Illumina NextSeq500 (75 bp single-end reads). The TED-seq libraries were prepared in two independent biological replicates.

Synthesis of spike-in poly-(A) standards

Poly(A) spike-in RNAs of 40, 80, 120 and 160 nt were generated by *in vitro* transcription of a PCR amplified double stranded DNA template composed of T7 promoter sequence, unique sequences for alignment from plasmid vector backbones [pmRFP-C1 (Addgene, 54764) for A40; pEGFP-C1 (Clontech) at EGFP ORF for A80; pEGFP-C1 at NeoR/KanR ORF for A120; pGL4.23 (Promega, E8411) for A160], and poly(A) repeats of desired lengths (Table S9). To generate four distinct 700-bp backbone sequences, different coding sequences were targeted and PCR-amplified with the set of gene-specific primers (Table S7) using Phusion High-Fidelity DNA Polymerase (25 cycles of 10 sec at 98°C, 30 sec at 60°C, and 30 sec at 72°C, followed by 5 min extension at 72°C in final volume 100 ul; 70 ul H₂O, 20 ul 5x buffer, ~2.5 ng template plasmid, 300 nM forward and reverse primers, 250 uM dNTP and 1 ul Phusion polymerase). After PAGE purification, oligo-dT tails of distinct sizes were added to the 3' end of the corresponding backbone template, and amplified with the specific primers (Table S9) using Phusion Polymerase (2 PCR cycles of 10 sec at 98°C and 50 sec at 68°C, followed by 5 min extension at 72°C). After PAGE purification, the resulting four distinct 700 bp templates with different coding sequences and tail sizes were *in vitro* transcribed using MAXIscript™ T7

Transcription Kit (Invitrogen, AM1314). The RNA products were purified by denaturing polyacrylamide electrophoresis (Urea; Thermo Scientific, U15-500) and quantified using a Nanodrop. The spike-in RNAs were added to purified mRNAs from samples used for TED-seq library generation (1 ng of each spike-in RNA species per 100 ng of poly(A)-selected RNA). The entire sequences of individual spike-in poly(A) standards are provided in Table S9.

TED-seq data pre-processing and mapping

For sequencing reads ending with >10 A residues, consecutive (A) sequences were trimmed from the 3' end (PRINSEQ v0.20.4; 33). After poly(A) tail trimming, reads with a length ≥ 15 nt and mean quality score > 20 were retained for further analysis. PCR duplicates were removed using the first 15 nt of the trimmed reads, which includes an 8 nt UMI (FASTX-Toolkit v0.0.11, PRINSEQ v0.20.4, Seqtk v1.3-r106). Nucleotides corresponding to the UMI were then trimmed from the 5' end of the deduplicated reads, followed by the exclusion of trimmed reads shorter than 15 nt. The resulting reads were mapped to the human genome (hg38) using STAR (2.4.2a; 26) with the option `--sjdbGTFfile "gencode.v26.annotation.gtf"` `--alignSJDBoverhangmin 3`, `--outFilterMultimapNmax 1`. BWA (Li and Durbin 2009) was used to align the reads corresponding to the poly(A) spike-in standards. The relationship between the mapped TED-seq reads, the cleavage and polyadenylation site, and the library insert size enables 5' terminus mapping coordinate of the aligned TED-seq reads to shift the 3' tail ends upstream into the 3' UTR by the insert size of the library (300 nt): for transcripts with longer poly(A) tails, TED-seq reads map closer to the PAS, whereas those with shorter tails map further from the PAS and into the 3'UTR. Accordingly, 3'UTR isoform-specific poly(A) tail length distribution is reproduced immediately upstream of the corresponding PAS by 5' termini of the mapped TED-seq reads, and

visualized on a genome browser track with the IGV genome browser (Robinson et al. 2011). Related to Figure 2.1 and Figure S2.1.

Poly(A) tail length estimation and identifying significant changes in poly(A) tail length

From the library insert size (I) and the distance from 5' termini of TED-seq reads to PAS (D), poly(A) tail length (L) is derived as $L=I - D$. GENCODE V26 annotation (Frankish et al. 2019; Harrow et al. 2012) of the human transcriptome (bed12 format) was amended to reflect experimentally determined THP-1 cleavage and polyadenylation sites (PAS) identified using 3'-seq. Finally, a frequency table of TED-seq read 5' termini located within the 3' terminal 500 nt of 3' exons within the custom transcriptome annotation was constructed. Transcript isoforms with ≥ 50 mapped reads in the terminal 500 nt region were used to calculate the mean value of poly(A) lengths and to represent the distribution of poly(A) tail reads for that region. Due to the dependency of reliable poly(A) tail length calculations on precise PAS annotations, any shifts in APA isoform preferences occurring within the 300 nt window (referred to as local PAS switch hereafter), limit the ability of TED-seq to determine poly(A) tail length changes. To avoid any errors in poly(A) tail inferences caused by a shift in APA usage, we removed PAS isoforms subject to potential bias from APA by processing our 3'-seq data as follows. First, we defined tandem PAS clusters of size = 300 nt by clustering experimentally determined PAS isoforms within ≤ 300 nt (9,658 PAS in 5,128 genes clustered to 6,745 tandem clusters) and testing whether APA usage in a cluster was significantly altered upon LPS stimulation for all isoforms. Switching in the tandem poly(A) sites in a tandem PAS cluster was determined based on previously described approaches (Jia et al. 2017; Fu et al. 2011). The PPI (Proportion of an individual PAS isoform in each tandem cluster) index was

calculated for each time point across the activation time course. If the PPI within each cluster significantly differ between any pair of time points of the macrophage LPS activation time-course (Chi-squared test with FDR <0.1), we considered those as locally shifted (124 tandem clusters, 410 PAS), and removed these isoforms from further analysis. We only considered PAS clusters with consistent APA isoform usage across the time-course for further analysis. Additionally, polychoric correlation coefficients, a version of Pearson correlation coefficient with discrete UTR indices, were calculated to estimate the switching direction, and defined as USI, UTR switching index (described as TSI; 39). A positive USI value (USI>0.1) indicated a switch to the longer tandem 3' UTRs (distal), while a shift to short tandem 3'UTR (proximal) has a negative USI value (USI < -0.1). Next, if a PAS cluster contained multiple PAS, the PAS with the most read counts was defined as the major PAS isoform, whereas PAS isoforms with fewer read counts were considered as minor PAS isoforms (n=2,979). To avoid redundant use of a collection of TED-seq reads for the PAS in the same cluster, minor PAS isoforms were removed from the tail length analysis, leaving 6,269 isoforms (5,079 genes). For this set of transcript isoforms, differences in poly(A) tail length for a given transcript isoform between two biological conditions were compared using the Kolmogorov–Smirnov test and the p-value adjusted by FDR with the criteria of FDR<0.1, and $|\Delta\text{PAL}|$ (difference in mean poly(A) tail length) ≥ 10 nucleotides. Related to Figure 2.2 and S2.2D-F.

Identification of significant changes in tandem 3'UTR isoform expression

By analyzing our custom transcript isoform annotation, each gene was scrutinized for tandem 3'UTR expression. If a gene had multiple transcript isoforms that shared the same 3' terminal exon (5' splice site, thereof), but were cleaved and polyadenylated at different positions, these transcript isoforms were considered as tandem 3'UTR

isoforms. For analytic stringency, we only considered transcript isoforms with ≥ 50 TED-seq reads within the 300 nt region upstream of the PAS for further analysis. This approach allowed us to remove any potential decay intermediates and other artifacts derived from internal priming from our collection of PAS isoforms, and focus on a more confident set of transcript isoforms. In cases with multiple transcript isoforms located within ≤ 300 nt, we determined the isoform with the highest read counts as the major isoform, and the rest as minor isoforms. Finally, for the final set of major transcript isoforms, each transcript isoform was indexed based on their genomic location, as follows. In the case of a 3' terminal exon expressing multiple PAS isoforms, the PAS closest to the stop codon was indexed as 1, incrementing the index with increasing distance from the stop codon. For a tandem 3'UTR isoform, mRNA abundance was calculated as the read counts of the given isoform. As described in ***Poly(A) tail length estimation***, we determined the switching direction using USI, and a Chi-squared test was performed to determine the statistical significance of the 3'UTR switch. Under the criteria of the adjusted p-value (FDR) < 0.1 , a positive USI value (USI > 0.1) indicated a switch to the longer tandem 3'UTRs (distal APA isoforms), while a shift to short tandem 3'UTRs (proximal APA isoforms) has a negative USI value (USI < -0.1). Related to Figure S2.2A-C.

PRO-seq library preparation

PRO-seq libraries for differentiated THP-1 cells were produced essentially as described (Mahat et al. 2016). 1×10^7 cells were permeabilized in 100 μ l buffer D (50mM Tris-Cl pH 8.0, 25% glycerol, 5mM MgAc₂, 0.1mM EDTA, 5mM DTT) for each PRO-seq sample. 2X concentrated nuclear-run-on buffer (20 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 2 mM DTT, 600 mM KCl, 500 μ M ATP (ThermoFisher, R0481), 500 μ M GTP (ThermoFisher, R0481), 50 μ M biotin-11-CTP (Perkin Elmer,

NEL542001EA), 50 μ M biotin-11-UTP (Perkin Elmer; NEL543001EA), 1% sarkosyl (Fisher Scientific; AC612075000), 1ul SUPERase In (ThermoFisher, AM2694) was added to the permeabilized cell, followed by incubation for 3 mins at 37°C, followed by 3x volume addition of Trizol LS (ThermoFisher, 10296028) to stop the reaction. GlycoBlue (ThermoFisher, AM9515) was added to the run-on products, and subjected to ethanol precipitation. The resulting pellet was dissolved in DEPC-treated water, denatured at 65°C, and subject to base hydrolysis in 0.2N NaOH on ice for 15 minutes, stopped by the addition of the same volume of 1M Tris-HCl (pH 6.8). Biotinylated RNAs were isolated using Streptavidin-conjugated magnetic beads (ThermoFisher, #11205D). Subsequently, 3'RNA adaptor ligation was performed (NEB; M0204L), followed by another round of bead binding, 5' de-capping using RppH (NEB, M0356S), 5' end phosphorylation (NEB, M0201S), and then 5' adaptor ligation. After the 5' adaptor ligation, a final round of bead binding was performed; the purified products were reverse transcribed, and PCR amplified, followed by Ampure XP-mediated size selection. Adaptor dimers were removed from the sequencing library by PAGE-mediated size selection, retaining fragments \geq 175 nt, and the resulting libraries sequenced from their 3' ends using an Illumina NextSeq500 platform. PRO-seq libraries were prepared in two biological replicates.

PRO-seq data pre-processing, mapping, and quantifying transcription activity

PRO-seq reads were trimmed by removing the adaptor sequence TGGAATTCTCGGGTGCCAAGG using Cutadapt (Martin 2011) with the parameter of `--e 0.10 --overlap 2`. From the trimmed reads, PCR duplicates were removed by collapsing reads with the same UMI barcodes, followed by UMI trimming. The de-duplicated reads with lengths \geq 15 nt were aligned to the human genome (hg38) using BWA (Li and Durbin 2009). The resulting uniquely mapped reads were used for

further analysis. The 5' ends of the aligned reads represent the 3' termini of the nascent RNA, but are located on the opposite strand. Therefore, the bam file of uniquely mapped reads was sorted, converted to bed format (BedTools; 41), retaining the 5' position of the reads, and switched to the opposite strand to correct transcript orientation. Finally, the number of reads mapped to the gene body region, which spans from 500 nt downstream of the transcription start site (TSS + 500 bp) to 500 nt upstream of the poly(A) site (PAS - 500 bp) for transcript isoforms longer than 1 kb, was calculated based on hg38 GENCODE(V26) transcript annotation, and normalized to reads per kilobase per million mapped (RPKM).

Combined analysis of PRO-seq, 3'-seq, and TED-seq

Stratified Random Sampling (SRS) was performed as follows. Genes were grouped by Δ RNA [\log_2 Fold Change (3'-seq RPM)] or Δ PAL [Δ (mean poly(A) length)] gene groups (Down, No change, Up) then randomly sampled to maintain the same distribution of Δ TXN [\log_2 Fold Change (PRO-seq)] across each group. Genes in each Δ RNA (or Δ PAL) group were split into bins based on Δ TXN, with the same number of genes sampled from each Δ TXN bin across the three Δ RNA (or Δ PAL) groups. For the analysis of genes with minimal transcriptional changes, we selected genes with minimal Δ TXN [\log_2 Fold Change (PRO-seq) < 0.5] throughout the time-course (0 h through 4 h), splitting such genes into three groups (Down, No change, Up) based on Δ RNA (or Δ PAL), followed by SRS based on Δ TXN (4h/0h) for more stringent Δ TXN equalization. These analyses were used to test the relationship between Δ PAL and Δ RNA after controlling for Δ TXN. Related to Figure 2.4, 2.5, and 2.6 (Figure S2.4, S2.5 and S2.7).

qRT-PCR

To quantify RNA abundance, total RNA was reverse transcribed using Maxima H minus reverse transcriptase (Thermo Scientific, EP0751) with random hexamers at 50°C for 1 hour, followed by qPCR using SYBR green master mix (Bio-Rad, 172-5270), as described in the manufacturer's guide. The following primers were used to amplify individual target genes: human GAPDH (forward, 5'-CAGCAAGAGCACAAGAGGAA-3'; reverse, 5'-TGGTTGAGCACAGGGTACTTT-3'); GAPDH signal was used to normalize signal from other genes. The entire qPCR primer sets are detailed in Table S9. Reactions were performed in 10 ul volume per reaction in 4 technical replicates, which was repeated twice (two biological replicates). The data are presented as mean \pm SEM, and two different conditions were compared using a two-tailed Student's t-test.

PAT assay

Poly-(A) tails on specific mRNAs were analyzed with a Poly(A) Tail-Length Assay Kit (Thermo Fisher, 764551KT), as follows. Total RNA was isolated and denatured at 65°C for 3 minutes. 3' termini were extended by poly(A) polymerase, with guanosine and inosine nucleotides. The G/I tailed poly-(A) RNAs were reverse transcribed using a universal primer provided by the manufacturer, which was annealed to the G/I tail region, and then amplified by PCR with the following gene-specific custom forward primer and universal reverse primer; *TNF* F primer 5' - TGACCAACTGTCACTCATT-3', *IL1B* F primer 5' - GTGCTCTCTTTAAATCAAGTCCT-3'. Other gene-specific primers and oligos used in this study are listed in Table S9. To confirm the size of the deadenylated mRNAs for each mRNA species, poly(A) tails were digested by RNase H, as follows. 2 ug of total RNA was denatured in the presence of oligo d(T)₁₈ at 65°C for 5 min, followed by slowing cooling to 30°C. The poly(A) tails, hybridized to oligo dTs, were digested

with 5U of RNase H (NEB, M0297) at 37°C for 1 hour, followed by heat inactivation at 65°C for 20 min, and TRIzol extraction. The purified RNAs were subjected to 3' RNA ligation with an RNA adaptor sequence, followed by reverse transcription (Thermo Scientific, EP0441) as described in the manufacturer's protocol. PAT PCR was performed as described by the manufacturer. Finally, the size of the PAT PCR product was assessed by running half of the PCR reaction through 6% PAGE gel and staining the gel with SYBR gold (Invitrogen, S11494). PAT assay was repeated at least twice for each gene of interest, and one of the replicates was shown as a representative.

Association of AU content in the 3'UTR with poly(A) tail length changes

For each transcript isoform, AU content in the 3'UTR was calculated by the number of A and U nucleotides divided by the 3'UTR length, using hg38 reference sequence of the target transcript's 3'UTR. For Δ PAL association tests, we classified transcript isoforms with Δ PAL \geq 10 nt as PAL-increased, those with Δ PAL \leq -10 nt as PAL-decreased, and with $|\Delta$ PAL| \leq 5 nt as not undergoing changes in tail length. 3'UTR length was calculated by integrating our experimentally determined PAS into the reference 3'UTR annotations (GENCODE v26). For a given transcript isoform, codon optimality was calculated as the mean of the codon stabilization coefficients (CSC) of all codons constituting the corresponding coding sequence using CSC values previously determined in four human cell lines, HEK293T, HeLa, RPE, and K562 (Wu et al. 2019). Related to Figure 2.3A, 2.3B and S2.3.

Association of 6-mer frequency with poly(A) tail length changes

A matrix of 6-mer counts for 3'UTRs corresponding to our experimentally determined isoforms was created by counting the number of instances of each 6-mer (n=4,096) in

each 3'UTR, which were normalized by the 3'UTR length. Alternatively, first and last 500 nt of 3'UTR regions were used in genes with 3'UTR lengths longer than 1 kb. Associations of each 6-mer count with changes in poly-(A) tail length were tested by binning genes into four quartile groups based on the given 6-mer content; 0-25%, 25-50%, and 50-75% and 75-100%, and comparing Δ PAL between the top and bottom quartiles, and assessed by Student's t-test. After FDR correction, 6-mers were considered to be associated with poly-(A) tail length changes with FDR <0.1. The identified 6-mers (significant 6-mers) were grouped into 7 groups based on the number of A and U residues; 0, 1, 2, 3, 4, 5 and 6. To examine the effects of the 6-mer on poly(A) tail length change, the value $\Delta\Delta$ PAL was defined for each of the 6-mers significantly associated with Δ PAL, as follows:

$$\Delta\Delta\text{PAL} = \text{mean } \Delta\text{PAL}_{\text{top quantile group (75\% - 100\%)}} - \text{mean } \Delta\text{PAL}_{\text{bottom quantile group (0\% - 25\%)}}$$

For 6-mers with $\Delta\Delta$ PAL > 0, we considered the corresponding 6-mer as associated with poly-(A) tail lengthening in response to LPS stimulation, and vice versa for 6-mers with $\Delta\Delta$ PAL < 0, indicating association with a decrease in tail length. Related to Figure 2.3C and 2.3D.

Combinatorial codes analysis in the 3'UTR

To identify two different cis-elements whose co-presence is more associated with Δ PAL than that of a single type, we examined combinatorial effects of two different 6-mers (A and B, hereafter) in the association with Δ PAL using a set of the 6-mers whose frequency was individually identified as associated with Δ PAL in the last 500 nt 3'UTR region. Transcript isoforms were categorized into 4 groups based on the frequency of A and B in the last 500 nt 3'UTR regions; 1) none of A and B, 2) A only (n=2), 3) B only (n=2), and 4) both A and B (n=1 for each; total n=2), where the total number of the tested 6-mers were fixed to avoid the potential bias from the number of

motifs. This analysis was iterated for all different time point comparisons. To predict RBPs that are likely to bind to a given 6-mer, RBP motif scores were calculated for the 6-mer using a corresponding position weight matrix (pwm) of a given RBP, which was iterated across all RBPs (n=202) using CISBP-RNA database (Ray et al. 2013). If a RBP has a pwm with more than 6 nucleotide positions, the given 6-mer was tested for a match to the pwm in a nucleotide increment; for a RBP motif pwm containing 7 positions, two motif scores were generated for a given 6-mer, for example. Any RBP showing the RBP motif score ≥ 6 was considered as a RBP whose motif aligns to a given 6-mer.

RBP enrichment in the 3'UTR

To scan for sites corresponding to RBP (RNA binding proteins), we used the CISBP-RNA database, which consists of 202 position weight matrixes for human RBP recognition motifs (Ray et al. 2013). Only the RBPs that are expressed in THP-1 cell, as defined by the 3'-seq data, were included (n=86). Here, 3'UTR sequences of experimentally determined PAS isoforms were used in the analysis. To count instances of an RBP binding motif enriched in the 3'UTRs of interest (PAL increase) relative to the background (no change in PAL) or counterpart (PAL decrease), each 3'UTR sequence was searched for the given RBP binding site with a score ≥ 6 , calculated as the natural log transformation of the odds ratio of the occurrence of the given motif in the examined 3'UTR sequence versus a two-order Markov model background. Fisher's exact test was performed to calculate the statistical significance, and the corresponding p-value was adjusted by FDR with a threshold for significant RBP motifs set as FDR<0.2. Related to Figure 2.5E, 2.6E and S2.7D.

Analysis of ActD-pretreated LPS activation

Poly(A) tail length changes were quantified for the PAS isoforms that passed a cutoff of 50 TED-seq reads across all time points (0, 1 2 h) in ActD pre-treated sample (n=7,771 derived from 6,069 genes). The genes with mean PAL increase greater than 5 ($\Delta\text{PAL} \geq 5$ and $\text{FDR} < 0.2$) in both biological replicates were defined as cytoplasmic/post-transcriptional polyadenylation targets and those with no change or decrease ($\Delta\text{PAL} < 1$) defined as controls. Otherwise, data was processed as described in the section *Poly(A) tail length estimation and identifying significant changes in poly(A) tail length*. Related to Figure 2.5 and 2.6 (Figure S2.6 and S2.7).

Western blot

At different, indicated time points upon LPS stimulation, cells were lysed in ice-cold lysis buffer (10 mM Tris, 10 mM NaCl (pH 8.0), 1% NP-40, 2mM MnCl_2) containing 1x Protease inhibitor (Roche, 11873580001) on ice. After scraping the cells into a tube, samples were sonicated until the lysate solution cleared. For the phosphatase-untreated samples, SDS was added to a final concentration of 1%, followed by 10 min incubation at 95°C. For the control sample (phosphatase-treated), a 30 min incubation with Lambda phosphatase (NEB, P0753S) was performed at room temperature, followed by a 10 min incubation with Alkaline phosphatase (NEB, M0290). After addition of SDS to a final concentration of 1%, phosphatase-treated samples were boiled for 10 min at 95°C to inactivate the phosphatase enzymes. Protein concentration was determined by BCA assay, and then equal amounts of proteins (50 ug) were prepared, mixed with 2x Laemmli buffer containing 2.5% beta-mercaptoethanol, then boiled again at 95°C for 5 min before loading. After electrophoresis in 7% SDS-PAGE gel, and transfer to a nitrocellulose membrane, the blotted membrane was incubated with 3% BSA containing TBS-T, followed by overnight incubation with primary antibodies for ZFP36 (Cell signaling, #71632), and

vinculin (Sigma-Aldrich, V9131). After incubation with infrared fluorescent dye-conjugated secondary antibodies (IRDye 800CW goat anti-rabbit for ZFP36, IRDye 680RD goat anti-mouse for vinculin), proteins were detected on a LI-COR Odyssey. Phosphorylated and un-phosphorylated ZFP36 were discriminated from one another by their difference in molecular weight, and sensitivity to phosphatase treatment (Kuan-Ting Wang et al, 2015). This experiment was repeated three times, and one of the replicates was shown as a representative in the main figure.

Bulk poly(A) assay

3 ug of total RNA (extracted by TRIzol) was labelled with 33.3 uM pCp-Biotin (Jena Bioscience, NU-1706-BIO) with 20 U T4 ssRNA ligase (NEB, M0204S) in a 50 ul reaction overnight at 16°C, followed by RNA purification by RNA Clean and Concentrator (Zymo Research, R1013). The labelled RNAs were digested by RNase A/T1 mixture at a 1:1000 dilution (Thermo Scientific, EN0551) for 30 mins at 37°C. The resulting poly(A) tails, which are insensitive to RNase A/T1 enzymes, were purified by TRIzol, followed by electrophoresis on a denaturing 7% 7M Urea-PAGE gel, together with labelled RNA ladder (Invitrogen, AM7145). For northern blotting, the RNAs were transferred to nitrocellulose membrane on a semi-dry Transfer blotting system (Bio-Rad, 1703940). After membrane blocking and subsequent wash steps, the 3' biotin-labelled poly(A) tails were detected with streptavidin-conjugated horseradish peroxidase (HRP) and a 1-min incubation with the chemiluminescent substrate (Luminol/Enhancer Solution), followed by brief exposure to X-ray film. All the materials used post transfer, including Blocking buffer, wash buffer, streptavidin-HRP, and Luminol/Enhancer Solution, were components of the North2South™ Chemiluminescent Hybridization and Detection Kit (Thermo Scientific, #17097).

Cloning of human ZFP36 3'UTR reporter constructs

For transfection into RAW 264.7 cells, an EGFP coding region and the human ZFP36 3'UTR sequences, either wild type (WT) or mutant versions (MUT-DEL, MUT-GC), were cloned downstream of the CMV promoter of the vector pCMV-7.1 (Addgene #47948) by Gibson assembly. For transduction into THP-1 cells, CMV-EGFP-ZFP36 3'UTR fragments, with either wild type or mutant (MUT-DEL, MUT-GC) versions of the human ZFP36 3'UTR, were cloned into a modified lentiviral pGFP-3'UTR plasmid (Clontech) by Gibson assembly. The human ZFP36 3'UTR sequences (WT, MUT-DEL, MUT-GC) were ordered as gBlocks (IDT) with vector-overlapping sequences at the fragment ends to be compatible with Gibson assembly.

Designing mutant versions of human ZFP36 3'UTR sequence

Wild type 3'UTR sequence of human ZFP36 was scanned for known RBP motifs using CISBP-RNA database. Motif score was computed by summing up weights (PWM) at corresponding positions. Out of the RBP motifs with the binding score ≥ 5 , those, which have 1) the number of consecutive Us ≥ 3 (e.g., UUU and UUUU), and 2) the portion of U in the given motif > 0.5 , were regarded as poly(U)-containing RBP motifs, subject to sequence modification. In all chosen poly(U) motifs, the consecutive poly(U) sequences were deleted to build MUT-DEL version of ZFP36 3'UTR, or replaced with G and C of the same length as the U stretch. The 3'UTR sequences of WT, MUT-DEL, and MUT-GC versions are provided in Supplementary Table S9.

Viral packaging and transduction

Lentiviral packaging and lentivirus infection with the 3'UTR reporter constructs were performed following the protocols from Broad Institute GPP (The Genetic

Perturbation Platform) web portal

(<https://portals.broadinstitute.org/gpp/public/resources/protocols>). For lentiviral packaging, HEK293T cells were seeded to a density of 5×10^5 cells in 10 cm culture dishes. 24 h later, transfection was carried out using TransIT-LT1 transfection reagent (Mirus Bio, #MIR2304) to introduce packaging and lentiviral plasmids into HEK293T cells. After harvesting the media containing lentivirus, the virus supernatant was filtered using 0.45 μ M syringe filters (VWR; #514–4133) and stored at -80°C until the lentiviral transduction. For transduction of THP-1 cells, 20,000 cells were plated in six-well plates, followed by spinfection with lentivirus at $300\times g$ at 30°C for 1.5 h. After 24 h incubation, 10 mL fresh medium was added to the cells, and transferred to T25 flask. Transduction status of the 3'UTR reporter constructs was monitored by checking GFP intensity on microscope on a daily basis. 3 days post-infection, G418 (Gibco, #10131027) was added to the cells at a final concentration of 1 mg/mL, followed by 2-week selection for the virus-integrated cells with media change twice a week. After antibiotic selection, cells were differentiated with 200 ng/mL PMA overnight, followed by media change with regular fresh media (without PMA). 3 days later, the differentiated cells were treated with or without 200 ng/mL LPS for 1 h before total RNA extraction for PAT assay.

Quantification and Statistical Analysis

Data presented as mean \pm SD or mean \pm 95% CI (as indicated). Statistical significance was calculated with two-tailed Student's t-test, Kolmogorov–Smirnov test, or Wilcoxon signed-rank test with the significance denoted as follows, $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), unless noted otherwise. The type of statistical test, and the value of n, and statistical significance (P value or FDR) are described in the Figures, Legends and/or Results. All the experiments and the sequencing libraries

were performed and prepared in two biological replicates, unless noted otherwise. For the screen shot of sequencing reads or immunoblot/PAT assay results, a single library/experiment was shown as a representative of ≥ 2 biological replicates confirmed to show consistent results. All the downstream sequencing data analysis were performed using the values averaged from two biological replicates. Stratified Random Sampling were used to determine the sample size. All graphs and statistical tests were performed using R.

Data Availability

The code for processing TED-seq, PRO-seq and 3'-seq and the analyses reported in this paper is available at https://github.com/YeonuiKwak/Project_MacActivation.

Accession Numbers

Raw and processed data of TED-seq, PRO-seq, and 3'-seq are available at the GEO accession number GSE161188.

ACKNOWLEDGEMENT

We thank Y. Woo in the Kwak lab for providing synthetic poly(A) RNA standards, and thank the members of the Grimson and the Kwak labs for constructive discussion and the Cornell Institute of Biotechnology, Genomics facility for next-generation sequencing. We also thank the Graduate Field of Genetics, Genomics and Development, and the Department of Molecular Biology in Cornell University for student and research support.

Author Contributions: Y.K and H.K conceptualized the study; all authors contributed to the study design and methodological approaches. Experiments were performed by Y.K., and analysis of TED-seq and PRO-seq data was performed by Y.K. and H.K. The TED-seq, PRO-seq, 3'-seq and transcription inhibition data were analyzed by Y.K. under the supervision of A.G. and H.K. Y.K. and E.A.F performed 3' UTR reporter assays with contributions from C.D. The manuscript was written by Y.K., A.G., and H.K.

Funding: This work was supported by discretionary funds provided to A.G. and H.K.

Conflict of Interest: The authors declare no competing interests.

FIGURES

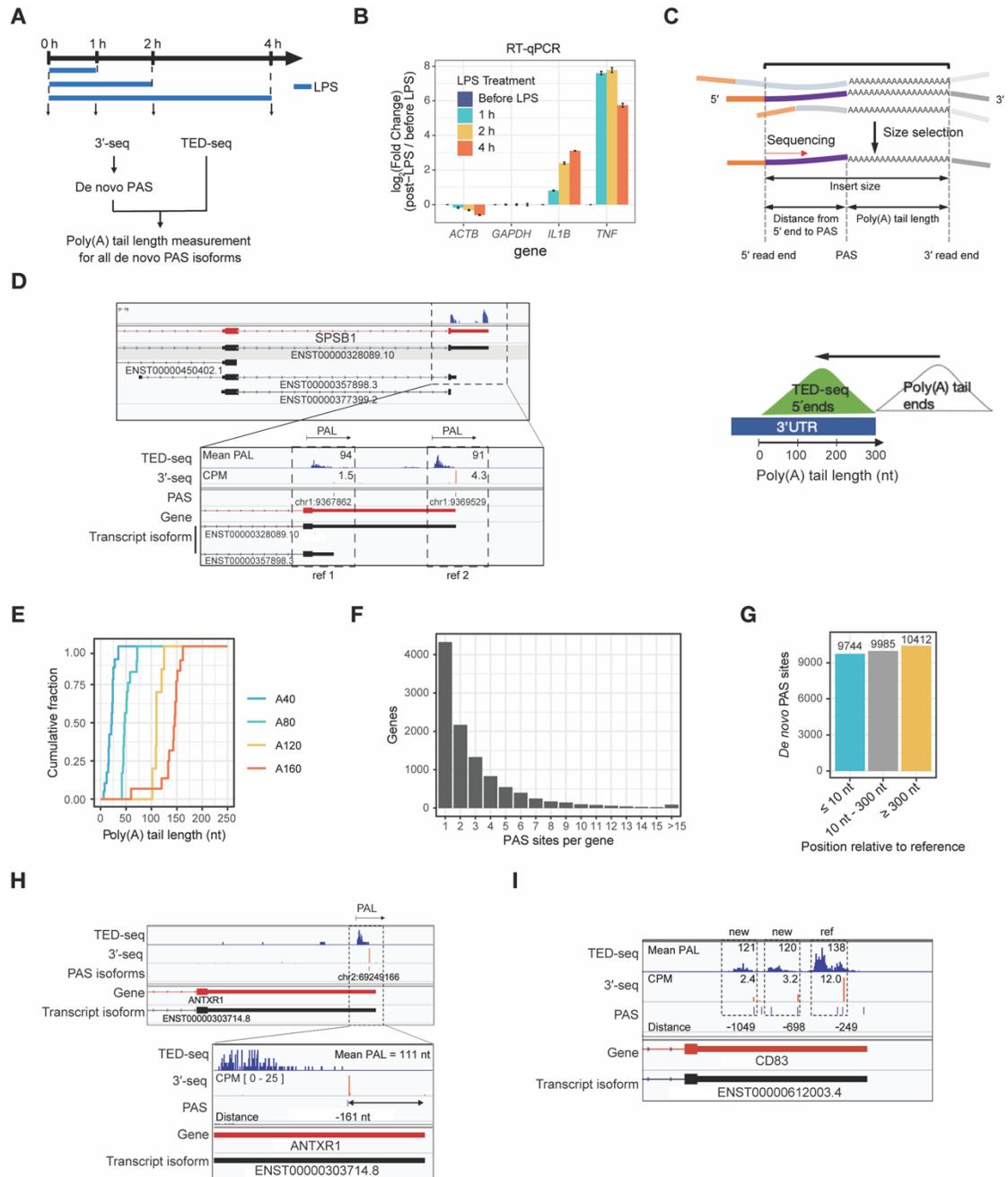


Figure 2.1. Determination of PAL with isoform specificity

(A) Schematic of activation time-course and sequencing strategy.

(B) Validation of macrophage activation using qRT-PCR. In each sample, expression values were normalized to GAPDH expression. For each gene, plotted is fold change of gene expression (post-stimulation/unstimulated). Data is representative of two biological replicates, each performed with 4 technical replicates.

(C) Structure of TED-seq libraries (left) and an illustration of PAL visualization by TED-seq on genome browser track (right).

(D) Genome browser tracks of TED-seq (5' terminus of reads) for *SPSB1*, indicating PAS previously annotated and determined by 3'-seq. Mean PAL (nt) and CPM values displayed for each 3'UTR isoform in TED-seq and 3'-seq tracks, respectively. Arrow marks on top of the genome browser track indicate PAL from reference point for each 3'UTR isoform.

(E) Cumulative distribution (CDF) of spike-in PALs (x-axis). Representative data from a single library. The results from all time points are shown in the supplemental figure S2.1B.

(F) Histogram of PAS counts per gene.

(G) PAS counts grouped by distance to reference PAS (nearest annotated PAS), for 3'UTR sites and those within 1 kb downstream (n=30,141).

(H) Genome browser tracks of TED-seq and 3'-seq on *ANTXR1* 3'UTR. Mean PAL (nt) and read count (CPM) displayed, with TED-seq read distributions magnified (inset box), and relative position of 3'-seq peaks to reference PAS shown under the PAS track (minus indicates upstream).

(I) Genome browser tracks of TED-seq and 3'-seq on *CD83* 3'UTR. *De novo* PAS isoforms track (PAS) shows the positions of PAS and their distances from annotated PAS, as in panel H. Mean PAL (nt) and read count (CPM) displayed as in panel C.

Two biological replicates of TED-seq and 3'-seq libraries were prepared at each time point. One of two 0 h biological replicates was shown for TED-seq and 3'-seq data as a representative on the genome browser for the panels D, G, and H.

See also Figure S2.1.

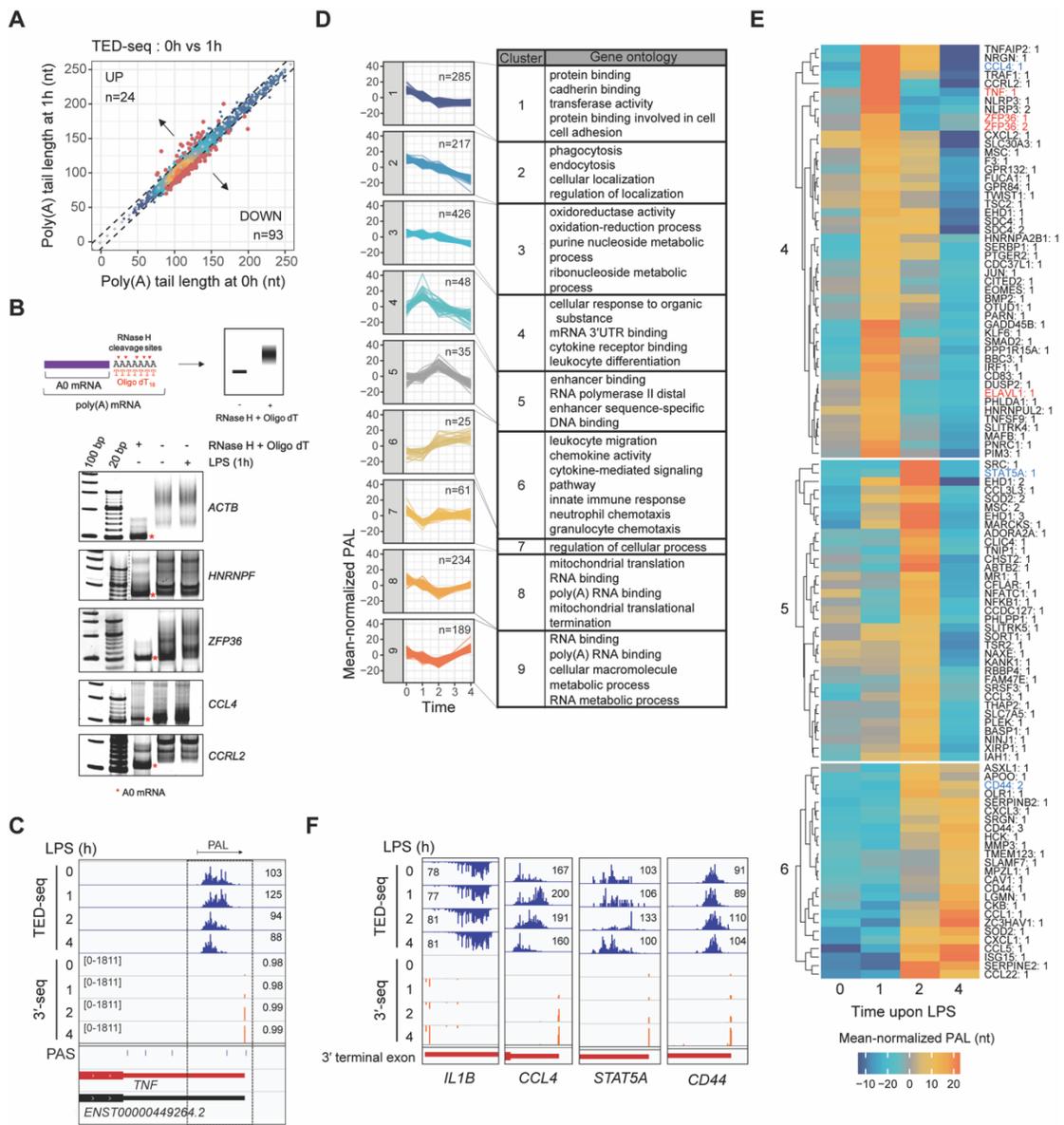


Figure 2.2. PAL dynamics during macrophage activation.

(A) Scatterplot of PALs between 0 (x-axis) and 1 hour post-activation (y-axis), averaged from two biological replicates. Each point denotes an isoform identified by 3'-seq. Point density color-coded from blue to orange (low to high). Red points indicate isoforms with significant changes ($\Delta\text{PAL} \geq 10$ nt and K-S test $\text{FDR} < 0.1$ in both replicates).

(B) Validation of TED-seq results by PAT assay for indicated genes. Final PAT-PCR products were analyzed on 6% non-denaturing polyacrylamide-TBE gel, followed by SYBR Gold staining. Deadenylated form (A0 mRNA; asterisk) was generated by treating total RNA with RNase H and Oligo dT.

(C) Genome browser tracks of *TNF*: TED-seq profiles and mean PAL (top 4), 3'-seq profiles and percentage of PAS isoform (PPI; next 4), during time-course. Values in square brackets indicate read count (y-axis) range. *De novo* PAS isoforms track (PAS) is displayed on the 9th lane.

For D and E, mean poly(A) tail lengths from two biological replicates were averaged for a given transcript isoform. The averaged PAL at each time point was mean-normalized by subtracting mean of the averaged PALs across all time points. The resulting mean-normalized PAL values were plotted.

(D) Distinct poly(A) tail length changing patterns during macrophage activation. GO terms enriched (Fisher's exact test, $\text{FDR} < 0.2$) in genes of each PAL-changing pattern listed.

(E) Heatmap of mean-normalized PALs in clusters 4, 5 and 6; labels identify genes and isoform index.

(F) Genome browser tracks of genes from clusters 4, 5 and 6 (colored in blue in panel D, x-axis) and *IL1B* as a negative control with no PAL change.

One of two biological replicates was shown for TED-seq and 3'-seq data as a representative on the genome browser for the panels C and F. Mean PAL displayed per track.

See also Figure S2.2 and S2.3.

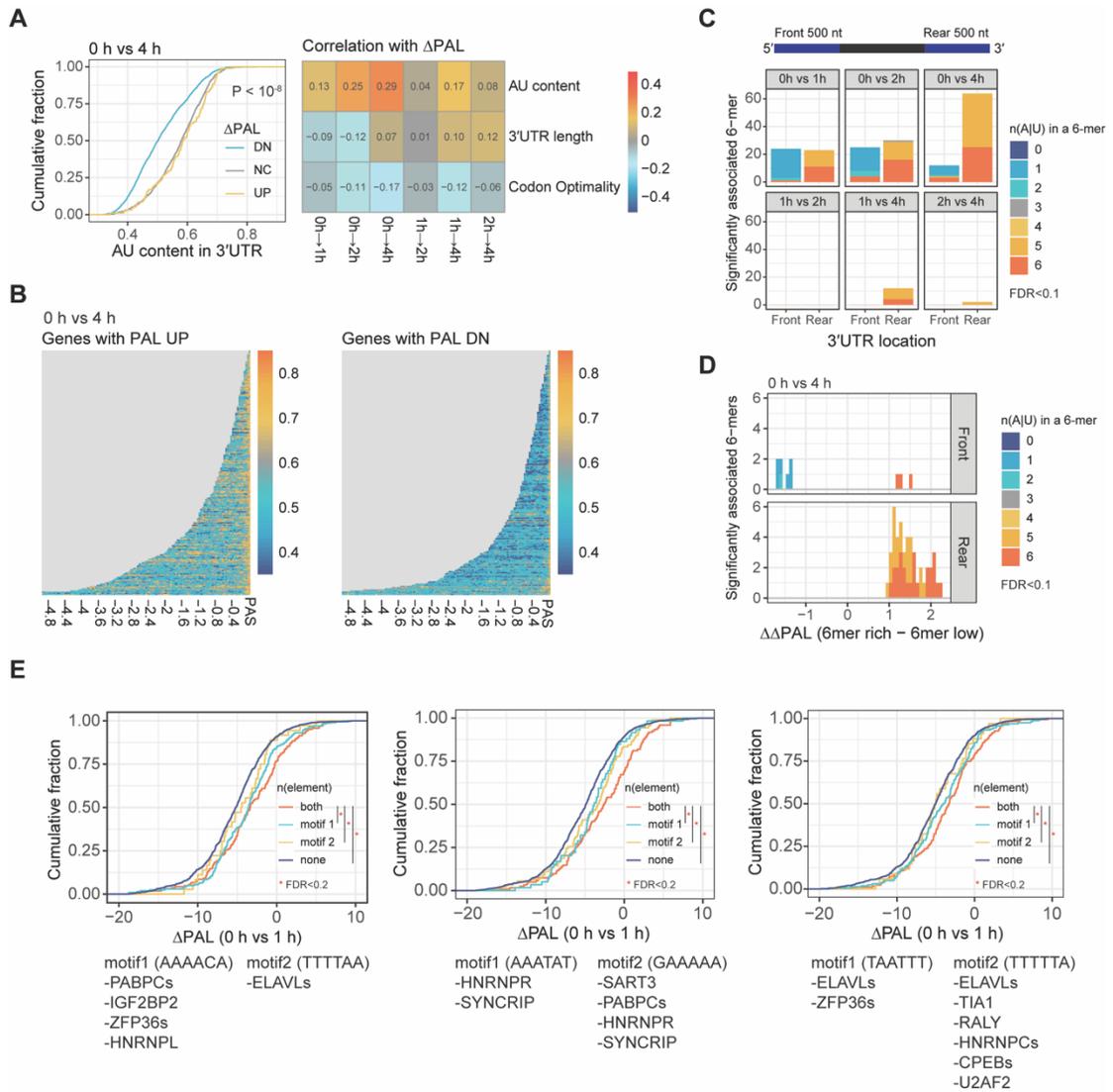


Figure 2.3. RNA features contributing to PAL control during the macrophage immune response.

(A) Association of different transcript features with Δ PAL. (left) CDF of 3'UTR AU content (x-axis) in transcripts grouped by Δ PAL between 0h and 4 h (down, DN; no change, NC; up, UP). Indicated P value denotes two-tailed K-S tests between Δ PAL UP and DN groups. (right) Heatmap of Pearson correlation coefficients between transcript features (y-axis), and Δ PAL at indicated time intervals (x-axis). When calculating Δ PAL, average value of mean PALs from two biological replicates was compared between two different time points (e.g., $PAL_{4h} - PAL_{0h}$).

(B) Heatmap of AU content across 3'UTRs in transcripts with increased and decreased PAL (left and right panels, respectively). Rows are transcripts ordered by 3'UTR lengths, and columns (x-axis) are consecutive non-overlapping 50 nt windows of 3'UTR for a given transcript isoform, aligned to the PAS. Each cell contains AU content calculated in a given window, where the value is colored from blue to orange (low to high).

(C) Association between 3'UTR motif location and Δ PAL during activation. Number of unique 6-mers associated (Student's t-test, $FDR < 0.1$) with Δ PAL (y-axis), partitioned by location (first and last 500 nt of 3'UTR; x-axis), in six time intervals (subpanels). 6-mers grouped and color-coded by number of A or U bases [n(A|U)].

(D) Distribution of $\Delta\Delta$ PAL of the unique 6-mers identified in panel C, with respect to 3'UTR locations. The comparison of 0 h and 4 h was shown as a representative. Top and bottom subpanels indicate Front and Rear 3'UTR positions, respectively. $\Delta\Delta$ PAL (x-axis) is the difference of Δ PAL (0 hr to 4 hr) values between the highest quartile isoforms with a given 6-mer and the lowest quartile isoforms with the 6-mer. 6-mers color-coded as in panel C.

(E) Pairs of 6-mers showing stronger association with Δ PAL (0 h vs 1 h) in combination than individually (K-S test, FDR<0.1). The 6-mers identified as associated with Δ PAL (0 h vs 1 h) in the last 500 nt 3'UTR windows were tested for their combinatorial effects. The top three combinatorial 6-mers were shown as a representative and the other pairs of 6-mers for other time point comparisons are provided in table S6.

See also Figure S2.3.

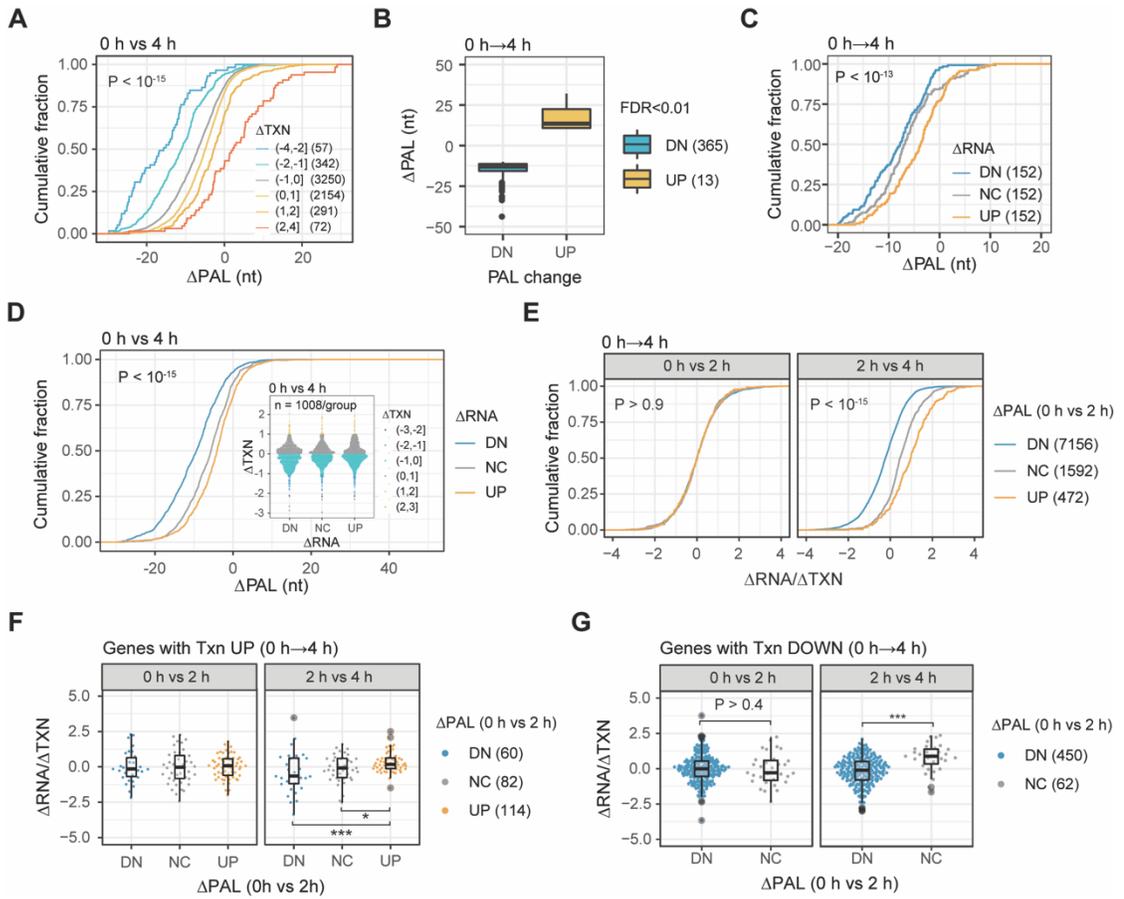


Figure 2.4. Association between changes in PAL and mRNA abundance regardless of transcriptional changes.

(A) CDF of Δ PAL (x-axis) with respect to transcriptional changes (Δ TXN; PRO-seq, \log_2 Fold Change) between 0 and 4 hours post-activation. Transcripts are binned into 6 groups based on Δ TXN (e.g., $(-4,-2]$ denoting $-4 < \Delta$ TXN ≤ -2), and their Δ PAL values are plotted as a group. K-S test P value compares Δ PAL values from lowest and highest Δ TXN bins. TXN values (RPKM) were averaged from two biological replicates of PRO-seq at each time point, and the fold change value between two different time point was \log_2 -transformed for Δ TXN calculation.

(B) Box plots of Δ PAL for transcript isoforms with PAL changes out of the genes with minimal transcriptional changes ($|\Delta$ TXN| < 0.5 in all pairwise time intervals) for 0 h and 4 h comparison. The number of transcripts are denoted. PAL changes were defined as $|\Delta$ PAL| ≥ 10 nt and K-S test FDR < 0.01 .

(C) Analysis of genes with minimal Δ TXN between 0 h through 4 h ($|\Delta$ TXN| < 0.5 in all pairwise time intervals). CDF of Δ PAL (x-axis) among the sets of transcripts grouped based on RNA abundance changes (Δ RNA; DN, NC, and UP) defined in Figure S2.4F. K-S test P value between DN and UP groups indicated.

3'-seq read count values were averaged from two biological replicates at each time point, and then the fold change between two different time points were \log_2 -transformed for Δ RNA calculation.

(D) CDF of Δ PAL (x-axis) differentiated among the three Δ RNA groups (Δ RNA; DN, NC, and UP). Transcripts grouped by Δ RNA proceeds with RSS by Δ TXN levels (between 0 and 4 hours post-activation). Inset violin plot shows Δ TXN per Δ RNA group. K-S test P value for comparison between DN and UP groups indicated.

(E) CDF of Δ RNA/ Δ TXN (x-axis; \log_2 (3'-seq FC/ PRO-seq FC)) comparing 0 to 2 hour (left) and 2 to 4 hour (right), for genes grouped by Δ PAL between 0 h and 2 h

post-LPS (DN; Δ PAL<10, NC; $|\Delta$ PAL| <5, UP; Δ PAL>10). K-S test P values between Δ PAL UP and DN groups indicated.

(F) Δ RNA/ Δ TXN with respect to Δ PAL in transcriptionally up-regulated genes. Δ RNA/ Δ TXN between 0 to 2 hour (left) and 2 to 4 hour (right), for genes grouped by Δ PAL (between 0 to 2 hour as defined in panel E). Student's t-test P values are denoted by asterisk.

(G) Δ RNA/ Δ TXN with respect to Δ PAL in transcriptionally down-regulated genes; otherwise as in panel F.

(Significant differences indicated as follows: *: P<0.05; **: P<0.01; ***: P<0.001)
Number of transcript isoforms or genes for the corresponding bins or groups are in parentheses.

FC, Fold Change; RSS, Random stratified sampling; down, DN; no change, NC; up, UP.

See also Figure S2.4 and S2.5.

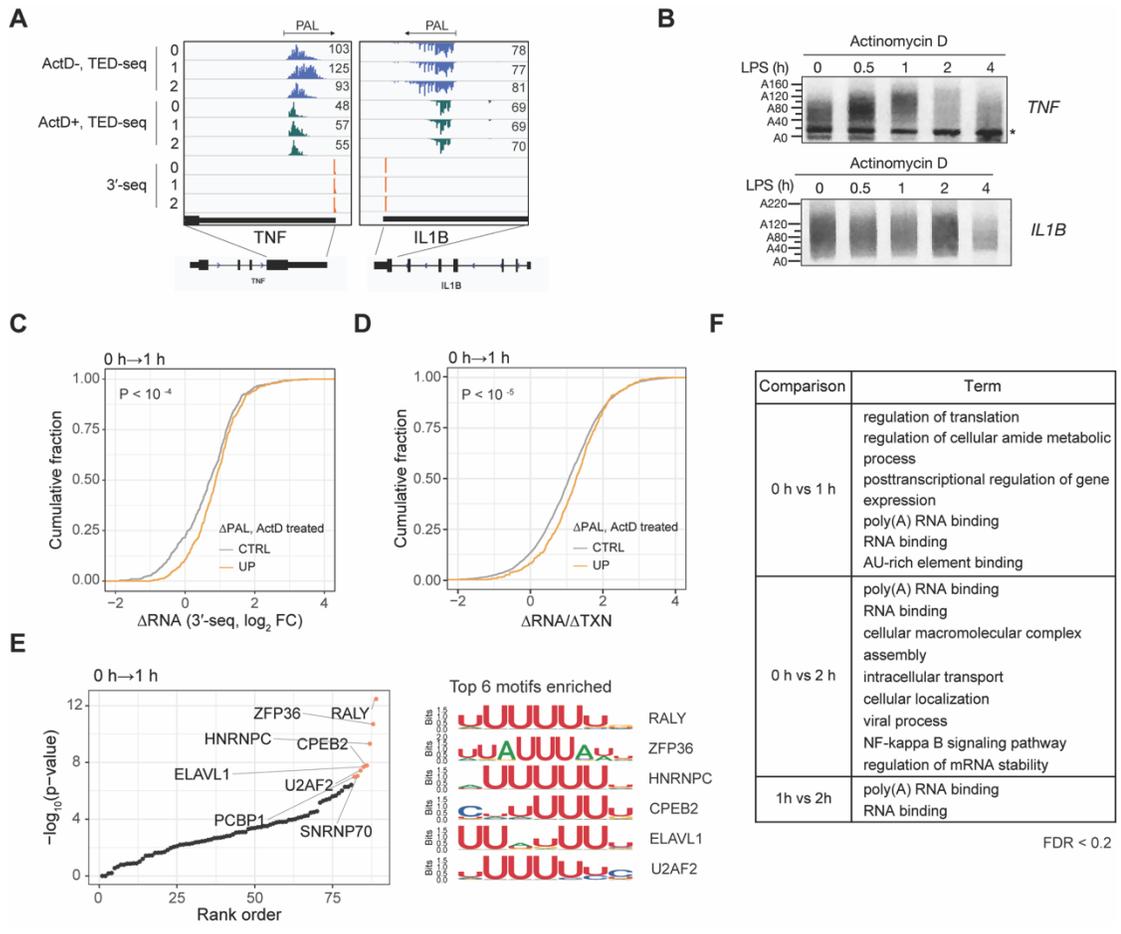


Figure 2.5. Widespread readenylation during macrophage activation.

(A) Genome browser tracks of TED-seq reads for *TNF* in the presence or absence of Actinomycin D (ActD) at different time-points post-activation (y-axis), together with 3'-seq (ActD-untreated) at the same time-points. Mean PAL was displayed at the corners of TED-seq tracks. 3'-seq peak indicates the position of PAS for the given gene. One of two biological replicates was shown on a genome browser for TED-seq and 3'-seq.

(B) PAT assay (tail length, y-axis) for *IL1B* and *TNF*, during LPS activation time-course in the presence of ActD. * denotes non-specific band.

(C) CDF of Δ RNA (x-axis) compared between readenylated (UP) and control (CTRL) transcripts. K-S test P-value for the comparison between UP and CTRL is denoted.

(D) CDF of Δ RNA/ Δ TXN (x-axis) for readenylated and control transcripts defined as in panel C (orange and grey lines, respectively). K-S test P-value for the comparison between UP and CTRL is denoted.

(E) Association of RBP motifs with readenylated transcripts. (top) Statistical significance (y-axis) of RBP motifs tested for enrichment within the 3'-terminal 500 nt of 3'UTRs of readenylated transcripts compared to control transcripts. (bottom) Sequence logos of top ranked motifs.

(F) Gene ontology terms enriched in transcripts undergoing readenylation (Fisher's exact test, FDR<0.2).

See also Figure S2.6 and S2.7.

Figure 2.6. ZFP36 mRNAs are readenylated through poly(U)-containing 3'UTR elements upon macrophage activation.

(A) Genome browser tracks of TED-seq reads for *ZFP36* in the presence or absence of Actinomycin D (ActD) at different time-points post-activation (y-axis), together with 3'-seq at the same time-points. One of two biological replicates was shown on a genome browser for TED-seq and 3'-seq.

(B) PAT assay (tail length, y-axis) for *ZFP36*, during LPS activation time-course in the presence or absence of ActD (one of three replicates as a representative). Red asterisk indicates the PCR product for completely deadenylated mRNAs, derived from RNase H treatment in the presence of oligo dT.

(C) Schematic of GFP reporters with either wild-type (WT) or mutant versions (MUT-DEL, MUT-GC) of human *ZFP36* 3'UTR. Three distinct THP-1 stable cell lines were generated with each expressing one of the GFP-ZFP36 reporters by lentiviral transduction. Known RBP motifs were searched in the 3'UTR region. For the motifs containing consecutive Us (≥ 3 Us) and with at least half of the motif length composed of Us (annotated in red), the consecutive Us were modified to have a deletion or GC substitution (annotated in green). The 3'UTR length was indicated under each construct name in brackets. The PAT assay forward primer was presented as a red line for each construct with their distances from the cleavage site in brackets.

(D) PAT assay on the human *ZFP36* 3'UTR reporter mRNAs of WT, MUT-DEL, and MUT-GC versions expressed in the differentiated THP-1 cells before and after LPS treatment (1h). The gel image of PAT assay (left panel) was digitally quantified by image J (right panel). Red asterisk denotes the PCR product of completely deadenylated mRNA (A0).

See also Figure S2.7.

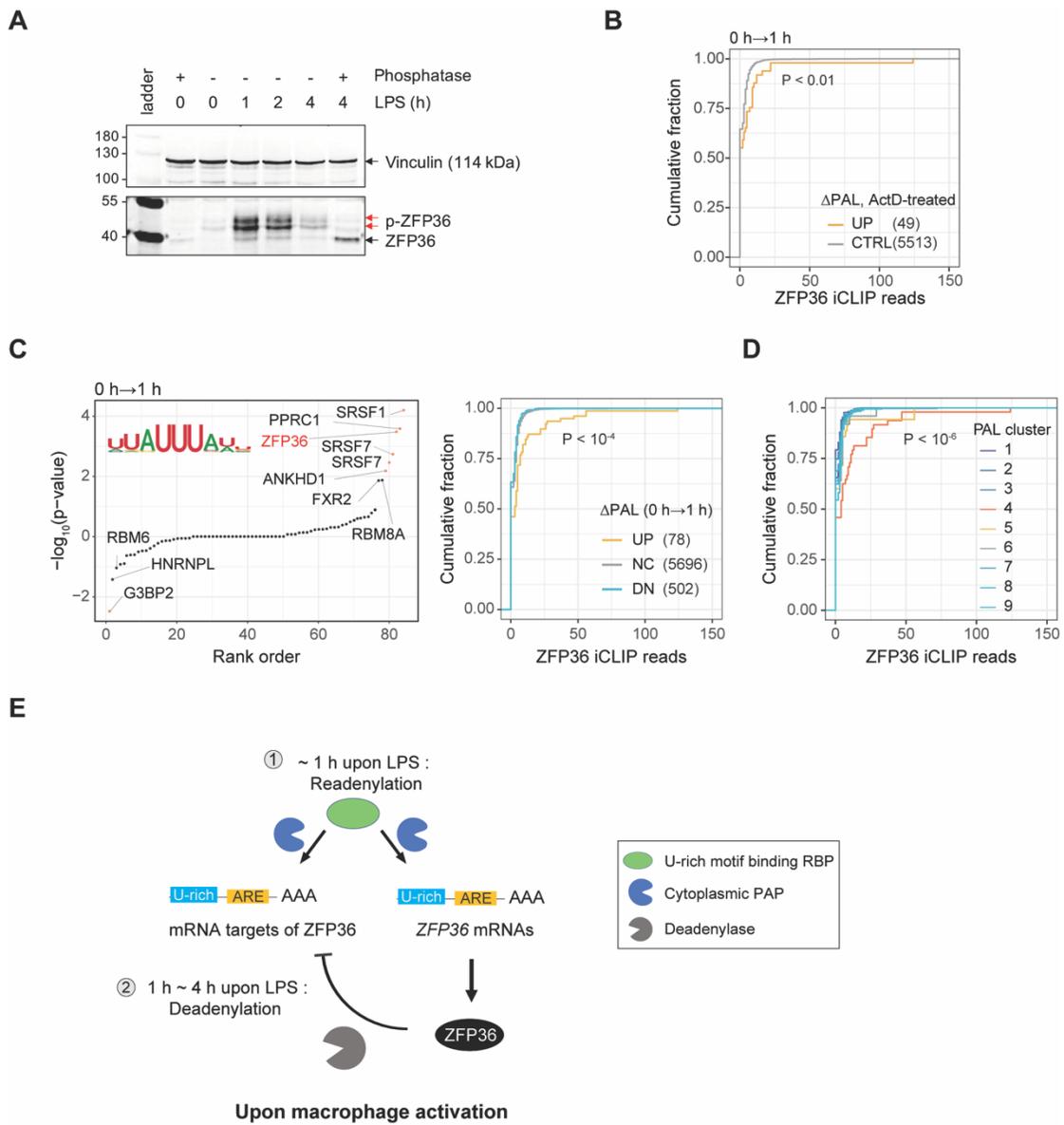


Figure 2.7. Concomitant readenylation of ZFP36 and its target mRNAs upon macrophage activation.

(A) LPS-induced changes in overall protein expression and phosphorylation level for ZFP36 protein was measured by western blot, together with staining for Vinculin as a loading control. One of three replicates was displayed as a representative.

(B) CDF of ZFP36 iCLIP binding (x-axis) in the 3'UTRs of readenylated and control transcripts. The number of transcripts in each group is displayed in the legend. K-S test P value between UP and CTRL groups is denoted.

(C) Association of RBP motifs with Δ PAL (0 to 1 hours). (left) Motif enrichment P values (Fisher's exact test) in the last 500 nt of 3'UTRs of transcripts with increased PAL compared to decreased PAL. (right) CDF of ZFP36 iCLIP binding (x-axis) in the 3'UTRs of transcripts grouped by PAL changes (DN, NC, and UP). Indicated P value denotes K-S test P value between Δ PAL UP and DN groups. The number of transcripts in each group is labeled in the legend.

(D) ZFP36 binding across PAL clusters defined in Figure 2.2D. CDF of ZFP36 binding (x-axis) in 3'UTRs of transcripts in different PAL-clusters. Wilcoxon test P-values between cluster 4 and the rest of clusters indicated. In panel D-F, ZFP36 iCLIP data from bone marrow derived macrophage (BMDM) post-activation was used.

(E) Model of post-transcriptional feedback loop via ZFP36 mRNA readenylation during macrophage activation.

See also Figure S2.7.

Supplementary figures

FIGURE S1

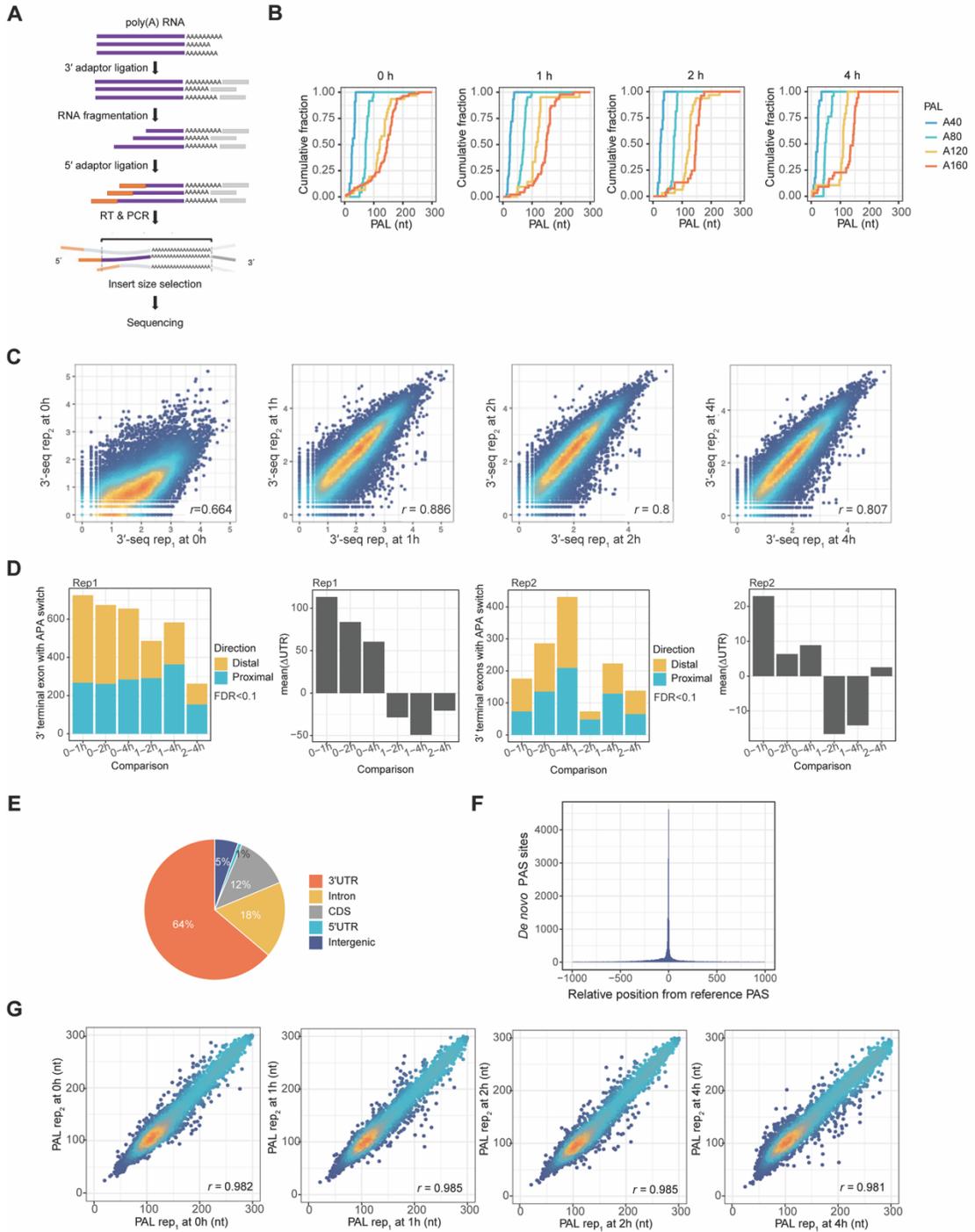


Figure S2.1. Reproducibility of TED-seq and 3'-seq, and association of poly(A) tail length with 3'UTR isoform usage; related to Figure 2.1.

A. Schematic of TED-seq library preparation (modified from Woo et al., 2018).

B. The tail length distributions of four distinct synthetic poly(A) standards, as measured by TED-seq. RNA standards with different tail lengths (Table S9) were added to a pool of poly(A)-selected RNAs. Plotted is the cumulative distribution of poly(A) tail length for each standard analyzed by TED-seq.

C. Scatterplots of 3'-seq between two biological replicates. Each point indicates 3'-seq read counts at a PAS isoform after CPM normalization ($n = 23,467$). Pearson correlation coefficient between biological replicates is shown at the bottom of the scatterplot. Local point density is color-coded from dark-blue to orange (low to high). Axes are \log_2 transformed after adding a pseudo-count of 1 to the CPM values.

D. Change in APA usage between the indicated time points, tested in individual sets of 3'-seq biological replicates. The number of 3' terminal exons with shifts in APA isoform usages, color-coded by the switching direction; left panels for each replicate set. Changes in the weighted 3'UTR length (Δ UTR) between two time points for the 3' terminal exons with two or more PAS; right panels of each replicate set. Switching to distal APA isoforms (3' UTR) depicted as an increase in Δ UTR, and switching to proximal APA isoforms as a decrease in Δ UTR (3'UTR lengthening and shortening, respectively). Barplots on the right panels are the mean of Δ UTR values for 3' terminal exons with significant shifts in APA usage ($FDR < 0.1$, and $|USI| > 0.1$). USI, tandem 3'UTR switch index

E. Pie chart of the genomic locations of *de novo* PAS determined by 3'-seq ($n=47,267$).

F. Distance of *de novo* PAS from nearest annotated site (reference PAS).

G. Comparison of poly(A) tail length measured by TED-seq between two biological replicates. Each point indicates mean poly(A) tail length of an individual PAS isoform (mapped reads ≥ 50 counts), and colored based on the density on the scatter plot (orange: high density, blue: low density). Pearson correlation coefficient between biological replicates is shown.

Figure S2

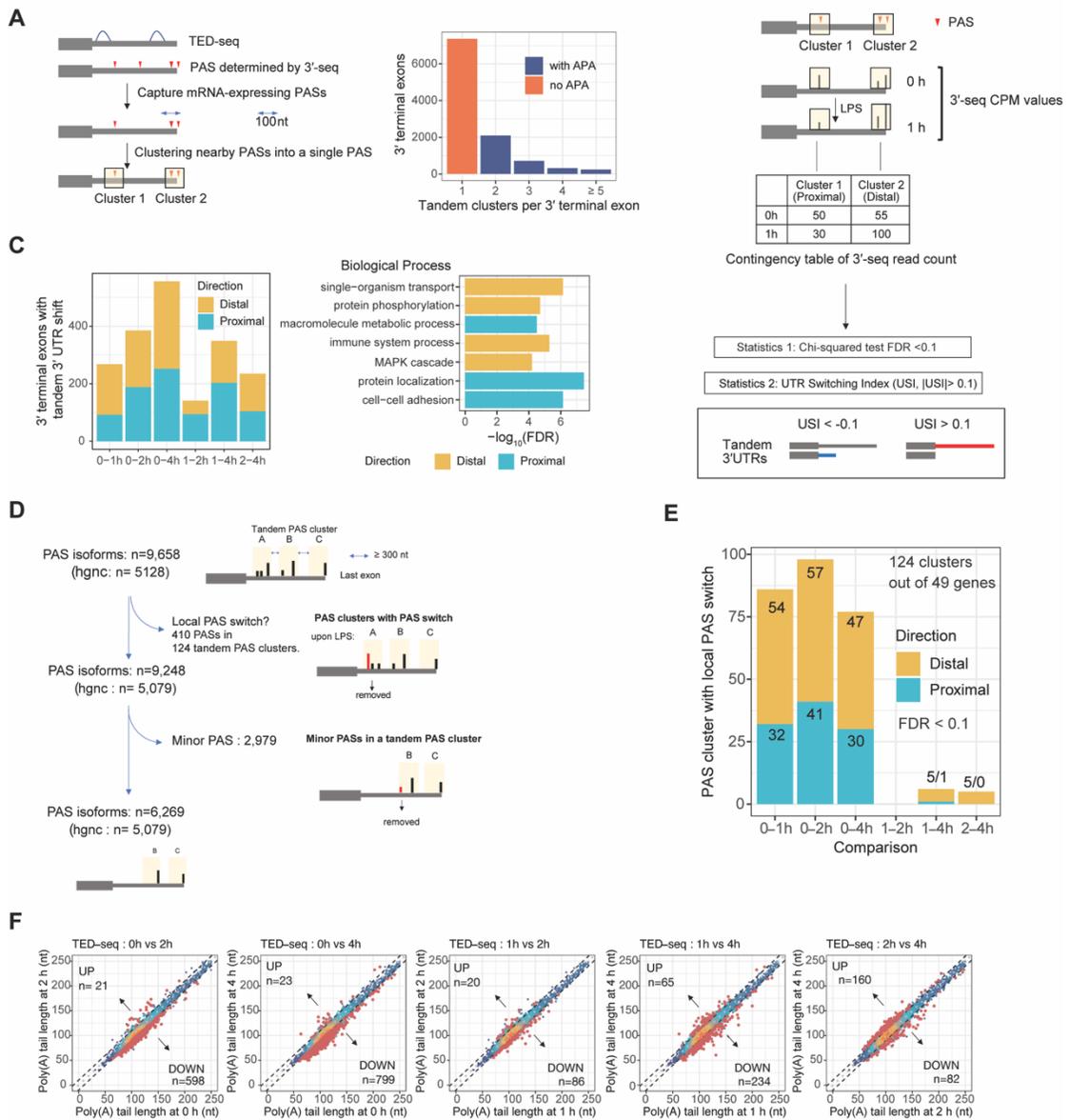


Figure S2.2. Analysis of 3'UTR isoform switching and measurements of poly(A) tail length change; related to Figure 2.2.

- A. (left) Schematic of APA isoform detection strategy and the corresponding RNA abundance. TED-seq and 3'-seq were integrated to identify APA isoforms (also known as tandem 3'UTRs). A PAS cluster represents a single APA isoform. (right) Number of tandem PAS clusters per 3' terminal exon.
- B. Schematic of strategy to identify altered APA isoform usage. See also *Analysis of 3'UTR isoforms* in the methods.
- C. (left) Number of genes with tandem 3'UTR shift. Switching direction color-coded. (right) Gene ontologies enriched in genes with tandem 3'UTR shift with their statistical significance on x-axis.
- D. Schematic of processing PAS isoforms to measure tail lengths. For a PAS isoform with one or more PASs nearby (window size: 300 nt), changes in PAS usage within the window may obscure detection of tail length changes upon stimulation. Therefore, such clustered PAS isoforms were filtered out (refer to the method section *poly(A) tail length estimation*). Next, for clustered PASs not exhibiting altered PAS usage, minor PASs within the window, defined by 3'-seq read counts, were removed to allow TED-seq reads to be applied to a single PAS isoform.
- E. Number of local PAS clusters with altered PAS usage (window size = 300 nt) for the indicated two time points. Statistical significance of altered usage of PAS isoforms in a local tandem PAS cluster between the indicated two time points were tested by Chi-squared tests (FDR <0.1). Switching directions were determined by calculating USI in a local tandem PAS cluster (yellow for distal switch, USI > 0.1; blue for proximal switch, USI < -0.1). USI, tandem 3'UTR switch index.

F. Scatterplots of mean poly(A) tail lengths (average of mean poly(A) tail lengths from two biological replicates) between the indicated time points; otherwise as described in Figure 2.2A.

Figure S3

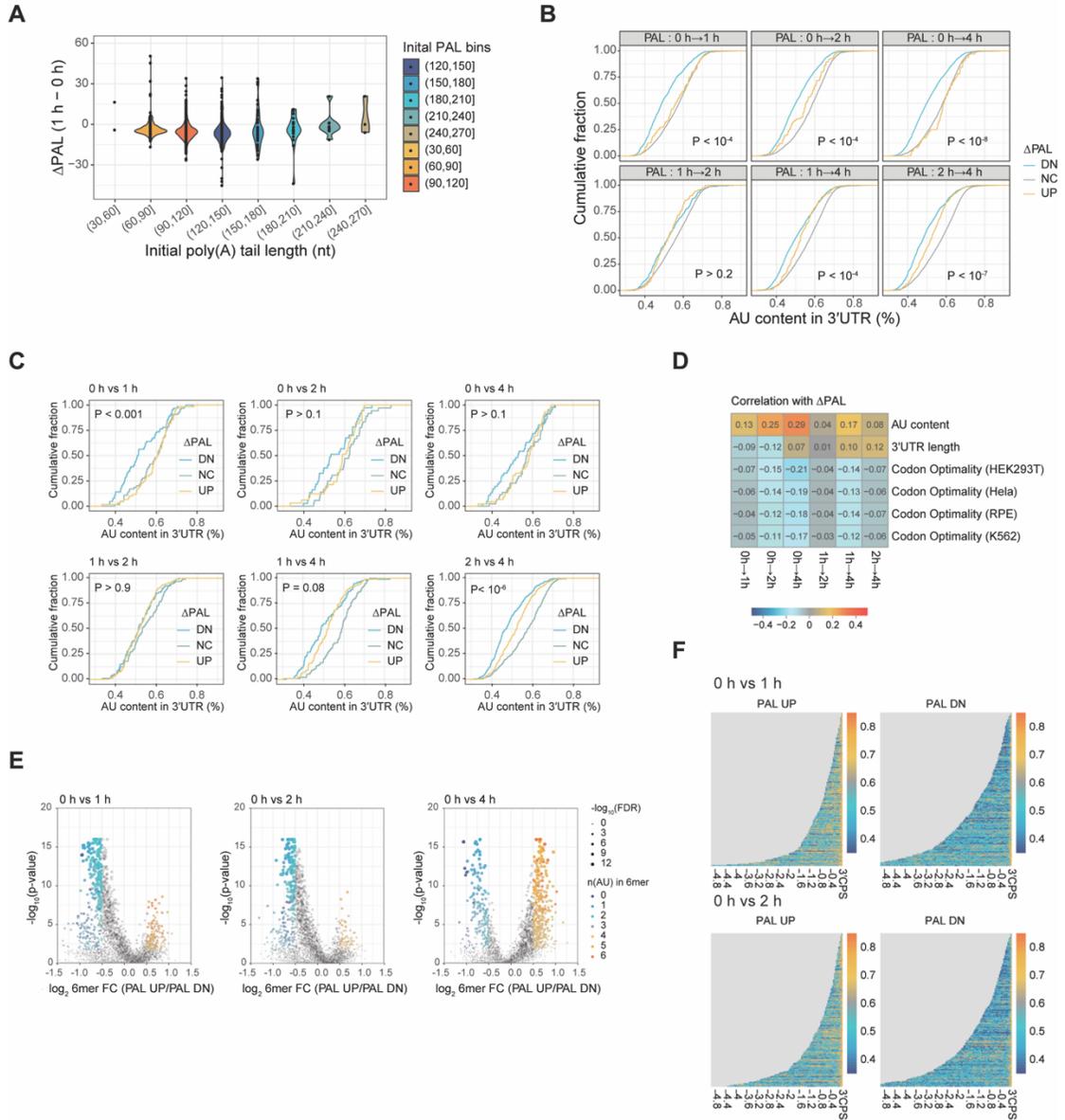


Figure S2.3. Analysis of sequence elements associated with poly(A) tail length changes; related to Figure 2.2 and Figure 2.3.

A. The relationship between initial poly(A) tail lengths and the tail length changes (Δ PAL) in the comparison of 0 h and 1 h time points. All 3' UTR isoforms were binned into 8 groups based on their initial tail lengths. For each isoform, an average of mean poly(A) tail lengths from two biological replicates at 0 h was used as the initial tail length, which was compared to that observed at 1 h post LPS.

B. Cumulative distributions of 3'UTR AU contents (x-axis) compared between transcripts grouped by Δ PAL during indicated time intervals. AU content per 3'UTR were calculated as the count of A and U in the 3'UTR divided by the length of the 3'UTR. Transcript isoforms grouped based on Δ PAL (Δ PAL \geq 10 nt for UP, $|\Delta$ PAL| $<$ 5 for NC, Δ PAL \leq -10 for DN); Δ PAL was calculated for each transcript isoform by comparing mean PALs averaged from two biological replicates between two different time points. Statistical significance assessed by K-S test, comparing UP and DN groups. down, DN; no change, NC; up, UP.

C. Cumulative distribution of 3'UTR AU contents in genes with minimal transcriptional changes compared between transcripts grouped by Δ PAL. Minimal transcriptional change (Δ TXN) was defined as $|\log_2 \Delta$ TXN| $<$ -0.5. TXN values (RPKM) were averaged from two biological replicates of PRO-seq at each time point, and the fold change value between two different time point was log₂-transformed for Δ TXN calculation.; otherwise as described in panel B.

D. Heatmap of Pearson correlation coefficients of Δ PAL and transcript features including codon optimality calculated separately for four different cell lines (y-axis) at indicated time intervals (x-axis). CSC scores (Wu et al., 2019) were used to calculate codon optimality.

E. Identification of 3'UTR 6-mers associated with Δ PAL at indicated time intervals. Subpanels are Volcano plots of 6-mers depleted or enriched in transcripts with Δ PAL ≥ 10 (UP) versus Δ PAL ≤ -10 (DN). x-axis is \log_2 6mer fold-change (UP/DN); y-axis is K-S test p-value. Each point represents a distinct 6-mer with the size and the color indicative of $-\log_{10}(\text{FDR})$ and the number of A or U [n(A|U)] in the 6-mer, respectively.

F. Heatmaps of the distributions of AU content along the 3'UTR regions (columns) in each set of the transcript isoforms (rows) categorized by Δ PAL during indicated time intervals; increased PAL (Δ PAL ≥ 10 nt for UP) and decreased PAL (Δ PAL ≤ -10 for DN). Each column (x-axis) corresponds to consecutive 50 nt windows ordered from the PAS, and each cell contains a AU content value calculated in the given 50 nt window, and the value is color-coded between blue (low) and orange (high).

Figure S4

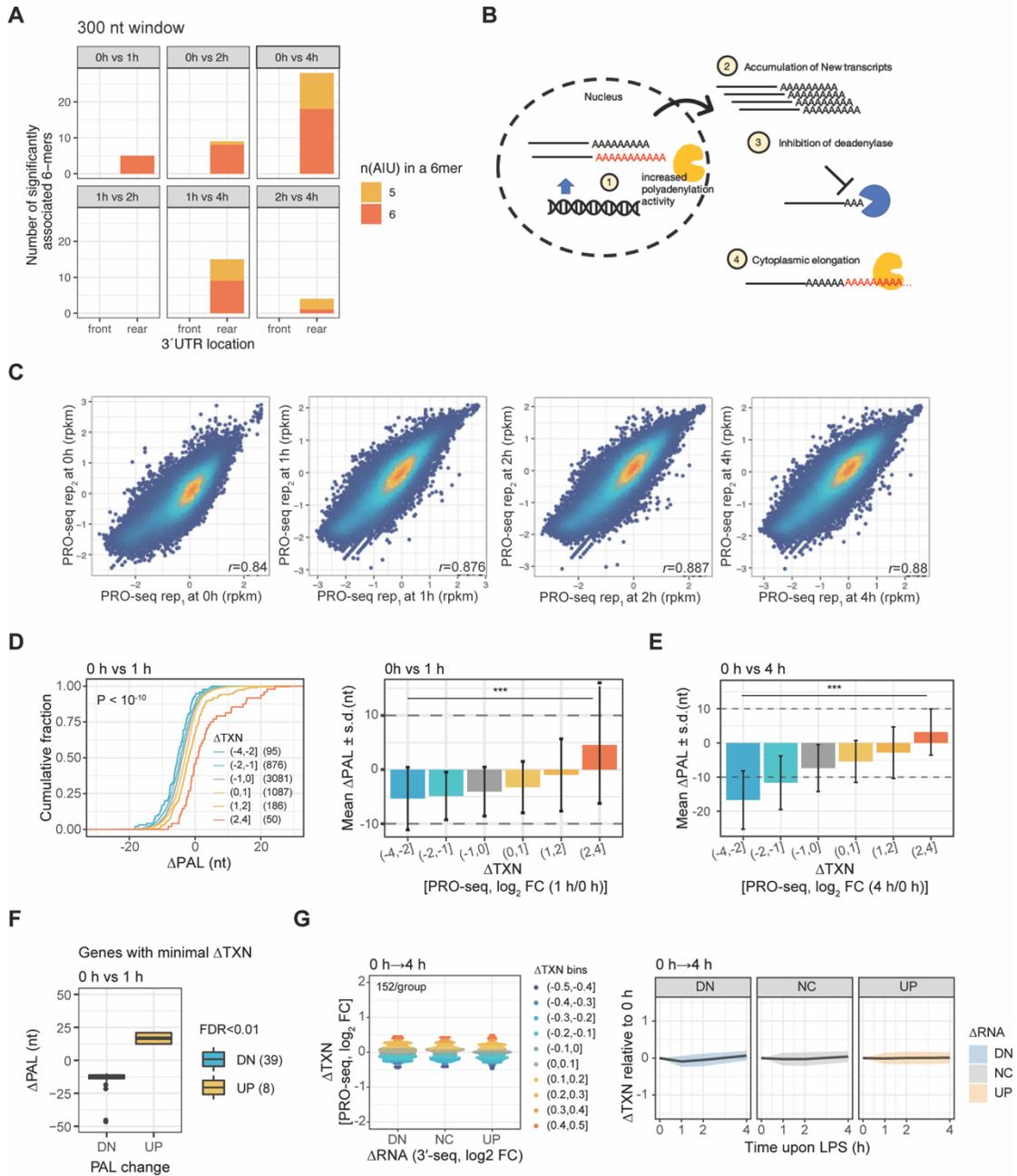


Figure S2.4. Association between transcriptional changes and changes in poly(A) tail lengths; related to Figure 2.3 and Figure 2.4.

A. Association between 3'UTR motif location and Δ PAL during activation.

Number of unique 6mers associated (Student's t-test, FDR<0.1) with Δ PAL (y-axis) in a given time interval, examined in either first or last 300 nt of 3'UTR(x-axis), and across all six time intervals (subpanels). 6mers were grouped and color-coded by number of A or U bases [n(A|U)].

B. Diagram of mechanisms that could explain increases in poly(A) tail length upon macrophage activation.

C. Scatterplots of read density per transcript isoform (RPKM) of PRO-seq between two biological replicates (n = 139,117). Each dot represents a transcript isoform, colored based on the dot density (high density in orange, and low density in blue). Pearson correlation coefficients between the biological replicates are shown on each scatterplot.

D. (left) The same analysis presented in Figure 2.4A applied for the comparison of 0 h and 1 h post-LPS. Cumulative distribution of Δ PAL (x-axis) with respect to transcriptional changes (Δ TXN; PRO-seq log₂ FC). Δ TXN was calculated as described in Figure S2.3C. (right) Bar-plots of mean Δ PAL (y-axis; error bars denote standard deviation) with standard deviation per Δ TXN bin (x-axis). ***, K-S test $P < 10^{-10}$

E. Bar-plots of mean Δ PAL (y-axis) with standard deviation per Δ TXN bin (x-axis) between 0 h and 4 h post-LPS; otherwise as described in Figure 2.4A (***, K-S test $P < 10^{-15}$).

F. Box plots of Δ PAL for transcript isoforms with PAL changes (Δ PAL \geq 10, UP; PAL \leq -10, DN) from genes with minimal transcriptional changes ($|\Delta$ TXN| < 0.5) for 0 h and 1 h comparison; otherwise as described in Figure 2.4B.

G. Analysis of genes with minimal ΔTXN between 0 h through 4 h ($|\Delta\text{TXN}| < 0.5$ in all pairwise time intervals). (left) Stratified random sampling equalized ΔTXN (y-axis; 0 h and 4h compared, PRO-seq \log_2 FC) distributions across three ΔRNA groups (\log_2 FC < -1 for DN, $|\log_2$ FC| < 0.5 for NC, and \log_2 FC > 1 for UP) based on ΔRNA (3'-seq \log_2 FC, y-axis). ΔRNA values were calculated as described in Figure 2.4C. Each dot indicates a transcript isoform, color-coded by ΔTXN . (right) The time-course changes in transcription (line, mean ΔTXN relative to 0 h; shade, standard deviation) for the sampled transcript isoforms (n=152/ ΔRNA group). down, DN; no change, NC; up, UP.

Figure S5

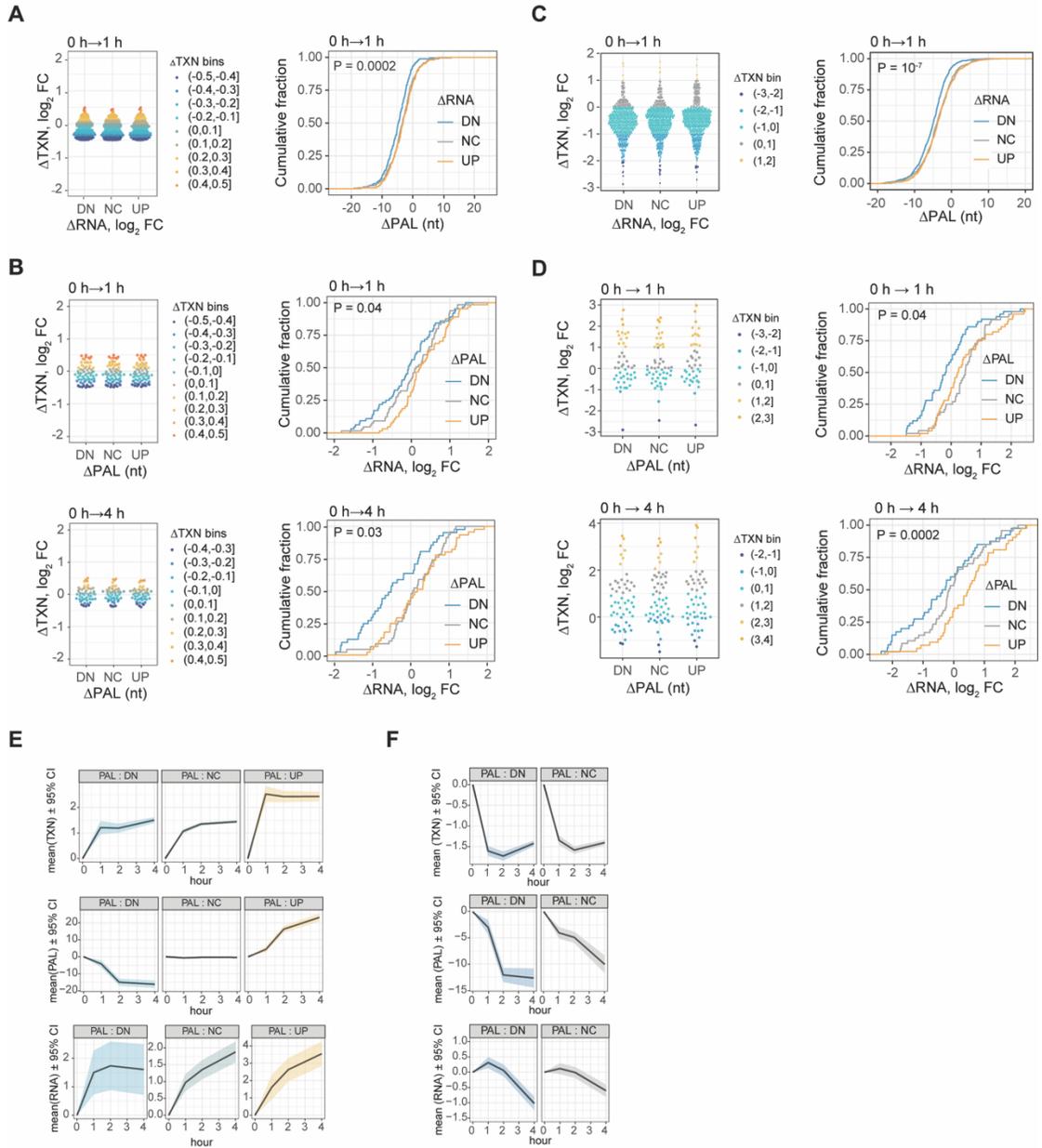


Figure S2.5. Association between changes in poly(A) tail lengths and changes in mRNA abundance; related to Figure 2.4.

A. (right) Cumulative distribution of Δ PAL (x-axis) during the indicated time interval, compared between transcript isoforms with minimal transcriptional changes, and grouped by change in RNA abundance, Δ RNA (3'-seq \log_2 FC); otherwise as described in Figure 2.4C and S2.4G. (left) Δ TXN (PRO-seq, \log_2 FC) distributions across the Δ RNA groups, to validate Δ TXN-matched sampling. P value (K-S test) denotes comparison between the UP and DN groups.

B. Cumulative distributions of Δ RNA (x-axis) during the indicated time intervals, compared between transcript isoforms with minimal transcriptional changes, and grouped by change in poly(A) tail length, Δ PAL (Δ PAL \geq 10, UP; $|\Delta$ PAL| $<$ 5, NC; Δ PAL \leq -10, DN); otherwise as described in panel A. P value (K-S test) denotes comparison between the UP and DN groups.

C. Association between Δ PAL and Δ RNA in THP1- cells upon LPS, corrected for transcriptional changes in the indicated time intervals. (left) Distribution of Δ TXN in Δ RNA groups after matched random sampling of the transcript isoforms with respect to Δ TXN. Transcript isoforms grouped by Δ RNA levels (\log_2 FC $<$ -1 for DN, $|\log_2$ FC| $<$ 0.5 for NC, and \log_2 FC $>$ 1 for UP) are sampled to have equal numbers across the 3 groups. (right) Cumulative distributions comparing Δ PAL between the three Δ RNA groups for the indicated time points. P values (K-S tests) denote comparison between the UP and DN groups.

D. Association between Δ PAL and Δ RNA in THP1- cells upon LPS treatment after Δ TXN-matched sampling in transcript isoforms groups with respect to Δ PAL. Analysis is performed as described in panel C, except that transcript grouping was based on Δ PAL.

E. Time-course dynamics of transcription, poly(A) tail length, and RNA abundance for the transcript isoforms defined in Figure 2.4F. Lines represent average values and shades represent confidence intervals. Changes in transcription level (PRO-seq, \log_2 FC; Δ TXN), poly(A) tail length (TED-seq, Δ PAL), and RNA abundance (3'-seq, \log_2 FC; Δ RNA) are plotted relative to 0 h.

F. Time-course dynamics of transcription, poly(A) tail length, and RNA abundance for the transcript isoforms defined in Figure 2.4G, plotted as described in panel E.

Figure S6

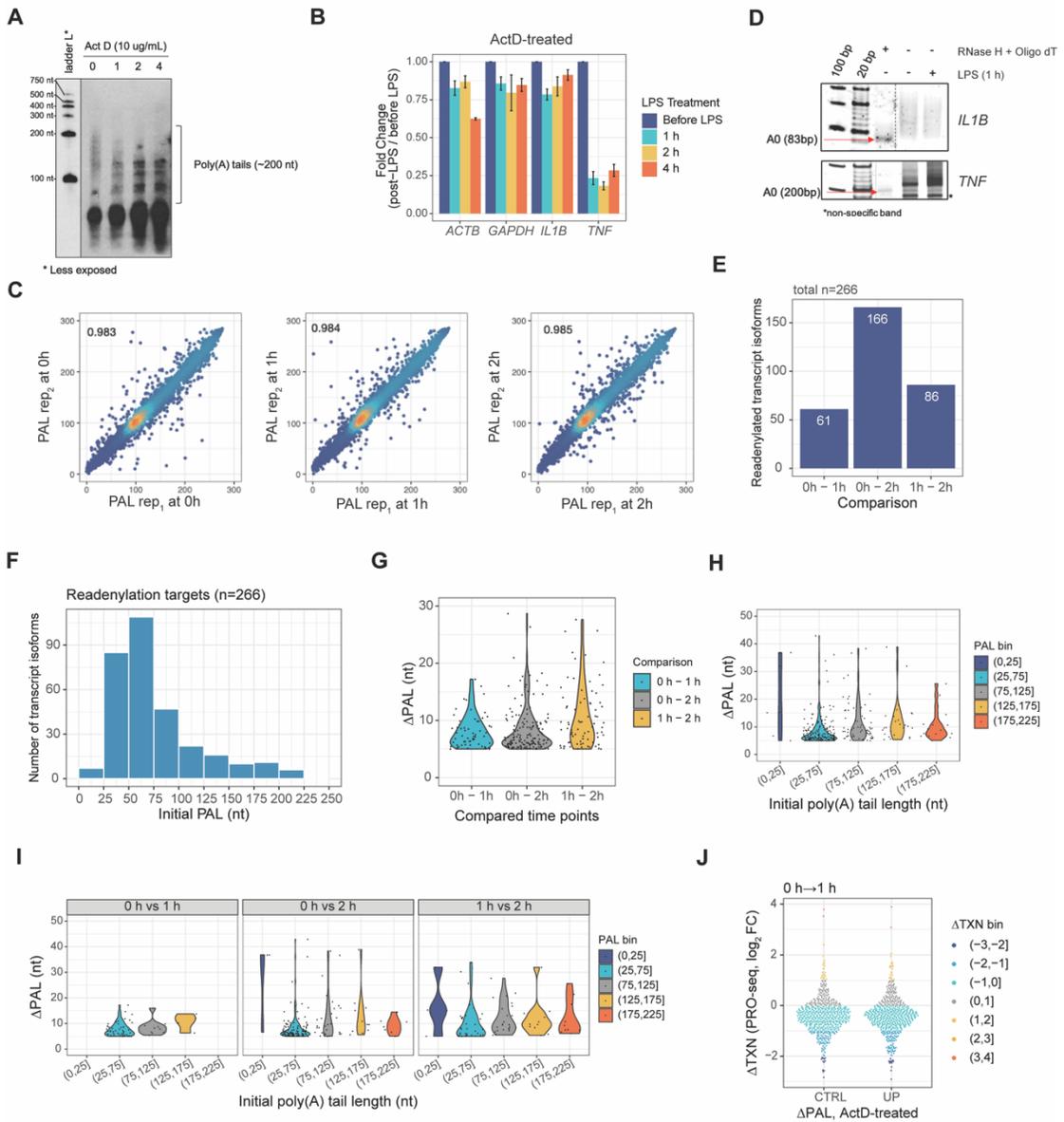


Figure S2.6. Analysis of readenylation upon LPS stimulation; related to Figure 2.5.

- A. Length changes of bulk poly(A) tails in the presence of Actinomycin D (ActD). Total RNA, 3' end labelled with biotin, was digested with RNase A/T1 mix, which preferentially degrades non-poly(A) sequences. The tails were visualized by northern blotting.
- B. RT-qPCR validation of transcription inhibition by Actinomycin D, performed with four technical replicates, repeated twice. Error bars indicate standard deviation.
- C. Scatterplots of poly(A) tail lengths measured by TED-seq between two biological replicates following LPS stimulation and ActD treatment (n=11,128 transcript isoforms from 6086 genes). Each point is the mean poly(A) tail length of a PAS isoform experimentally determined by 3'-seq.
- D. Validation of PCR product sizes for completely deadenylated (A0) transcripts of interest for PAT assays.
- E. Number of readenylation targets between two time points. Readenylation targets are defined as the PAS isoforms with poly(A) tail length increase greater than 5 nucleotides, determined by TED-seq from samples following LPS stimulation and ActD treatment.
- F. Relationship between initial tail lengths and readenylation. The 266 transcripts isoforms identified as readenylation targets by TED-seq were binned based on the initial tail (x-axis).
- G. Distribution of Δ PAL for transcripts identified as readenylation targets at indicated time point comparisons (0 h vs 1h, 0 h vs 2 h, and 1h vs 2 h; x-axis). Each dot indicates a transcript isoform with tail length increase that passed Δ PAL>5 nt, and K-S test FDR<0.2.
- H. Distribution of Δ PAL for the 266 readenylation targets binned by initial poly(A) tail length (nt, x-axis)

- I. Distribution of Δ PAL across the initial tail length bins (x-axis) for the readenylated targets with respect to time intervals.
- J. Distribution of Δ TXN (y-axis; 0 to 1 hours post-activation) for transcripts undergoing and not undergoing readenylation (x axis, UP; Δ PAL>5, CTRL; Δ PAL<1, respectively) after stratified random sampling by Δ TXN. Transcripts color-coded by Δ TXN (PRO-seq, \log_2 FC).

Figure S7

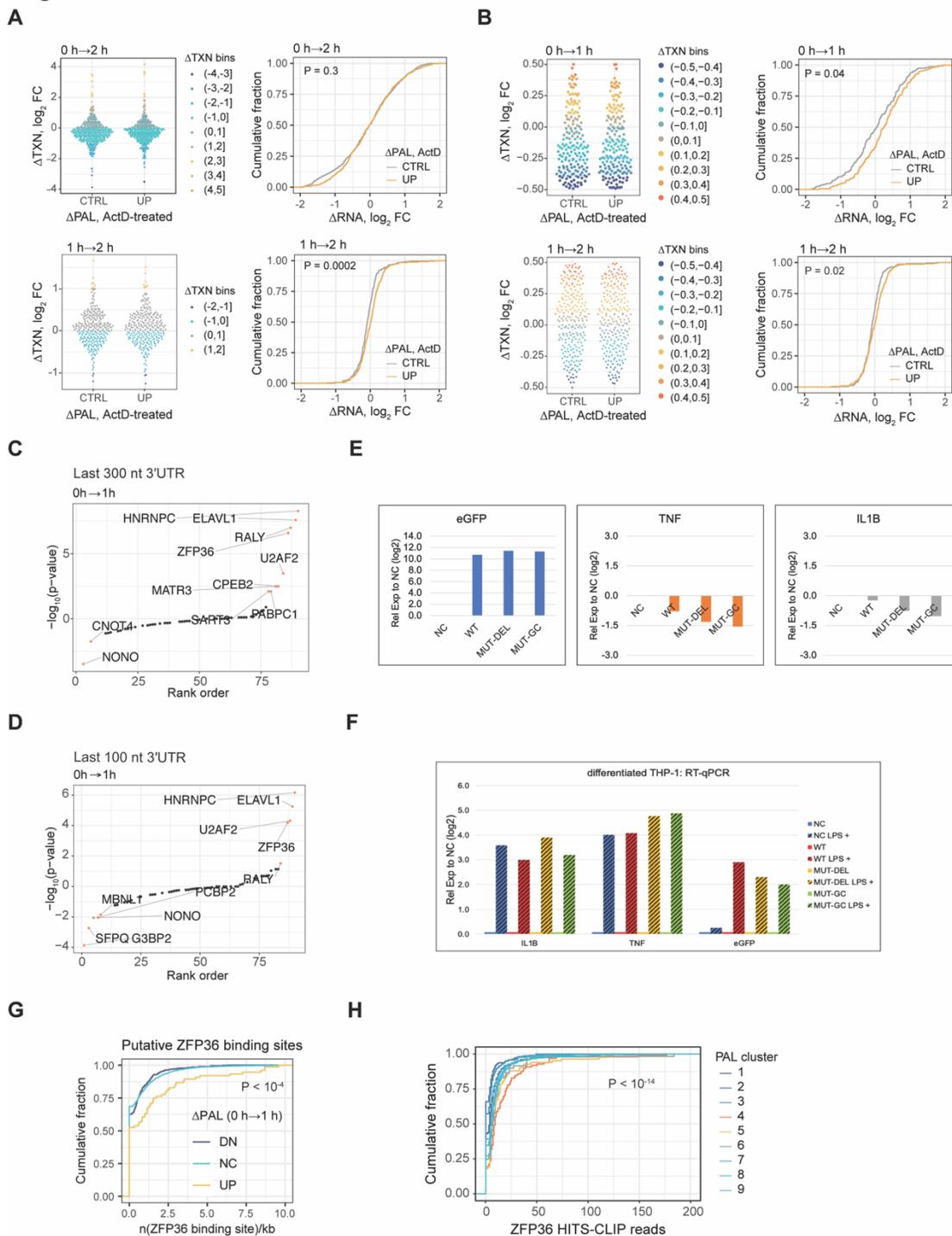


Figure S2.7. Analysis of readenylation upon LPS stimulation; related to Figure 2.5 and Figure 2.6.

A. Association between Δ RNA and readenylation. (Left panels) Distribution of Δ TXN during ActD non-treated LPS stimulation in control (CTRL) vs readenylated (UP) groups. CTRL and UP groups are defined as transcripts with Δ PAL < 1 and Δ PAL > 5 in actD pre-treated system upon LPS stimulation, respectively. (Right panels) Cumulative distribution plots of Δ RNA with respect to readenylation groups (CTRL and UP), as described in the left panels. Statistical significances are calculated by two-tailed K-S tests.

B. Association between Δ RNA and readenylation in genes with minimal transcriptional changes. Minimal transcriptional changes were defined as Δ TXN < 0.5 between two time points; otherwise as described in panel A.

C. Statistical significance (y-axis) of RBP motifs tested for enrichment within the 3'-terminal 300 nt of 3'UTRs of readenylated transcripts compared to control transcripts, respectively. Top 5 motifs were colored in red.

D. RBP motifs enriched within the 3'-terminal 100 nt of 3'UTRs of readenylated transcripts; otherwise as described in panel C.

E. Measurement of eGFP, TNF, and IL1B mRNA expression in the differentiated THP-1 cells stably expressing eGFP-3'UTR reporter genes by lentiviral transduction. After normalization by GAPDH levels, log₂-transformed fold change relative to negative control (no transduction) values were plotted.

F. Measurement of changes in eGFP, TNF, and IL1B mRNA expression upon LPS stimulation (1h) using the systems described in panel D. Log₂-transformed fold change relative to negative control (no LPS treatment) values were plotted in each of three reporter cell lines (WT, MUT-DEL, and MUT-GC) as well as original cell line (NC).

G. Cumulative distribution plots of putative ZFP36 binding site density (number of the motifs in the entire 3'UTR region divided by the 3'UTR length) in transcript isoforms with respect to poly(A) tail length changes - increase (UP), decrease (DN), or no change (NC) between 0 h and 1 h (ActD-untreated). Student t-test P-value between the UP and NC indicated.

H. ZFP36 binding across PAL clusters defined in Figure 2.2C. CDF of ZFP36 binding (x-axis) in 3'UTRs of transcripts in different PAL-clusters using HITS-CLIP data from CD4⁺ T cells after activation. Wilcoxon test P-value between cluster 4 and the rest of clusters indicated.

SUPPLEMENTARY TABLES

Tables below are available via Kwak et al (2022).

Table S1. Summary of PAS Located in 3'UTR and 1 kb Downstream Region (n=30141), Related to Figure 2.1.

For each PAS isoform determined by 3'-seq, it was assigned with transcript isoform id whose annotated PAS is the nearest to its PAS.

Table S2. BED12-formatted Customized Transcript Annotation, Related to Figure 2.1.

BED12-formatted transcript annotation was customized to terminate at 3'-seq determined PAS. The 13th column including gene name was added.

Table S3. RNA Abundance of All 3'UTR isoforms indexed based on their PAS positions. During Macrophage Activation, Related to Figure 2.1.

Each PAS isoform was assigned with tandem 3'UTR index and ALE index based on their relative location, and its relative abundance to the other PAS isoforms derived from the same gene at each time point is included in the dataset.

Table S4. Time-course Poly(A) Tail Length Profile, Related to Figure 2.2 and 2.3.

For each PAS isoform, the mean poly(A) tail lengths were averaged from the biological replicates.

Table S5. RBP motifs Aligning to the 6-mers Identified in the Last 500 nt 3'UTR Regions as Being Associated with Δ PAL upon Macrophage Activation, Related to Figure 2.3.

For each 6-mer identified as associated with Δ PAL, RBP with the motif score >5 were listed with the motif score.

Table S6. Combinatorial Effects of Two Different 6-mers in the Association with Δ PAL, Related to Figure 2.3.

In a set of the 6-mers whose frequency was individually identified as associated with Δ PAL in the last 500 nt 3'UTR region, All possible pairs of 6-mers were tested for the identification of two 6-mers (e.g, [A, B]), where presence of two different types ([A,B]= [1,1]; number of the tested two different 6-mers in brackets) have stronger association with Δ PAL than that of one type ([A,B]= [0,2] or [2,0]). FDR-corrected K-S test P-values for all statistical tests and the predicted RBPs that well align to given 6-mers were listed.

Table S7. The Integrated Datasets of Mean Poly(A) Tail Length, Transcription, and mRNA abundance, Related to Figure 2.4.

For each PAS isoform, mean poly(A) tail lengths, transcription, and mRNA abundance, are in the unit of nt, \log_2 (RPKM in gene body), and CPM, respectively.

Table S8. Readenylation Targets Identified by TED-seq in the Actinomycin D-treated THP-1 Cells, Related to Figure 2.5 and 2.6.

PAS isoforms that passed the criteria of K-S test FDR <0.2, and Δ PAL \geq 5 nt upon LPS stimulation were listed with the corresponding Δ PAL values in the table.

Table S9. Oligonucleotides or Sequences Used in This Study.

REFERENCES

- Batra R, Stark TJ, Clark E, Belzile J-P, Wheeler EC, Yee BA, Huang H, Gelboin-Burkhart C, Huelga SC, Aigner S, et al. 2016. RNA-binding protein CPEB1 remodels host and viral RNA landscapes. *Nat Struct Mol Biol* **23**: 1101–1110.
- Blumberg A, Zhao Y, Huang Y-F, Dukler N, Rice EJ, Chivu AG, Krumholz K, Danko CG, Siepel A. 2021. Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. *BMC Biology* **19**: 30.
- Braun JE, Huntzinger E, Fauser M, Izaurralde E. 2011. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell* **44**: 120–133.
- Brooks SA, Blakeshear PJ. 2013. Tristetraprolin (TTP): Interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim Biophys Acta* **1829**: 666–679.
- Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A. 1986. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc Natl Acad Sci U S A* **83**: 1670–1674.
- Carballo E, Lai WS, Blakeshear PJ. 1998. Feedback Inhibition of Macrophage Tumor Necrosis Factor- α Production by Tristetraprolin. *Science* **281**: 1001–1005.
- Carpenter S, Ricci EP, Mercier BC, Moore MJ, Fitzgerald KA. 2014. Post-transcriptional regulation of gene expression in innate immunity. *Nat Rev Immunol* **14**: 361–376.
- Chang H, Lim J, Ha M, Kim VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* **53**: 1044–1052.
- Charlesworth A, Meijer HA, de Moor CH. 2013. Specificity factors in cytoplasmic polyadenylation. *Wiley Interdiscip Rev RNA* **4**: 437–461.
- Chrestensen CA, Schroeder MJ, Shabanowitz J, Hunt DF, Pelo JW, Worthington MT, Sturgill TW. 2004. MAPKAP kinase 2 phosphorylates tristetraprolin on in vivo sites including Ser178, a site required for 14-3-3 binding. *J Biol Chem* **279**: 10176–10184.

- Corbett AH. 2018. Post-transcriptional regulation of gene expression and human disease. *Curr Opin Cell Biol* **52**: 96–104.
- Crawford EK, Ensor JE, Kalvakolanu I, Hasday JD. 1997. The Role of 3' Poly(A) Tail Metabolism in Tumor Necrosis Factor- α Regulation. *J Biol Chem* **272**: 21120–21127.
- Dai X-X, Jiang J-C, Sha Q-Q, Jiang Y, Ou X-H, Fan H-Y. 2019. A combinatorial code for mRNA 3'-UTR-mediated translational control in the mouse oocyte. *Nucleic Acids Res* **47**: 328–340.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Eichhorn SW, Subtelny AO, Kronja I, Kwasnieski JC, Orr-Weaver TL, Bartel DP. 2016. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos ed. E. Izaurralde. *eLife* **5**: e16955.
- Eisen TJ, Eichhorn SW, Subtelny AO, Bartel DP. 2020a. MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. *Molecular Cell* **77**: 775-785.e8.
- Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, Bartel DP. 2020b. The Dynamics of Cytoplasmic mRNA Metabolism. *Molecular Cell* **77**: 786-799.e10.
- Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766–D773.
- Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, Xu A. 2011. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* **21**: 741–747.
- Gallie DR. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* **5**: 2108–2116.
- Geissler R, Grimson A. 2016. A position-specific 3'UTR sequence that accelerates mRNA decay. *RNA Biol* **13**: 1075–1077.

- Geissler R, Simkin A, Floss D, Patel R, Fogarty EA, Scheller J, Grimson A. 2016. A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes Dev* **30**: 1070–1085.
- Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**: 75–79.
- Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91–105.
- Hake LE, Richter JD. 1994. CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* **79**: 617–627.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**: 1760–1774.
- Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
- Jalkanen AL, Coleman SJ, Wilusz J. 2014. Determinants and Implications of mRNA Poly(A) Tail Size - Does this Protein Make My Tail Look Big? *Semin Cell Dev Biol* **0**: 24–32.
- Jia X, Yuan S, Wang Y, Fu Y, Ge Y, Ge Y, Lan X, Feng Y, Qiu F, Li P, et al. 2017. The role of alternative polyadenylation in the antiviral innate immune response. *Nat Commun* **8**. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333124/> (Accessed July 31, 2020).
- Katsanou V, Papadaki O, Milatos S, Blackshear PJ, Anderson P, Kollias G, Kontoyiannis DL. 2005. HuR as a negative posttranscriptional modulator in inflammation. *Mol Cell* **19**: 777–789.
- Kim D, Lee Y-S, Jung S-J, Yeo J, Seo JJ, Lee Y-Y, Lim J, Chang H, Song J, Yang J, et al. 2020. Viral hijacking of the TENT4-ZCCHC14 complex protects viral RNAs via mixed tailing. *Nat Struct Mol Biol* **27**: 581–588.

- Kojima S, Green CB. 2015. Analysis of Circadian Regulation of Poly(A) Tail Length. *Methods Enzymol* **551**: 387–403.
- Kondrashov A, Meijer HA, Barthet-Barateig A, Parker HN, Khurshid A, Tessier S, Sicard M, Knox AJ, Pang L, De Moor CH. 2012. Inhibition of polyadenylation reduces inflammatory gene induction. *RNA* **18**: 2236–2250.
- Kontoyiannis D, Pasparakis M, Pizarro TT, Cominelli F, Kollias G. 1999. Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* **10**: 387–398.
- Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K, Rychlewski L, Ginalski K. 2016. FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res* **44**: 3534–3548.
- Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**: 950–953.
- Lai WS, Kennington EA, Blackshear PJ. 2003. Tristetraprolin and its family members can promote the cell-free deadenylation of AU-rich element-containing mRNAs by poly(A) ribonuclease. *Mol Cell Biol* **23**: 3798–3812.
- Legnini I, Alles J, Karaikos N, Ayoub S, Rajewsky N. 2019. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nature Methods* **16**: 879–886.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li Y, Sun Y, Fu Y, Li M, Huang G, Zhang C, Liang J, Huang S, Shen G, Yuan S, et al. 2012. Dynamic landscape of tandem 3' UTRs during zebrafish development. *Genome Res* **22**: 1899–1906.
- Lim J, Ha M, Chang H, Kwon SC, Simanshu DK, Patel DJ, Kim VN. 2014. Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. *Cell* **159**: 1365–1376.
- Lim J, Lee M, Son A, Chang H, Kim VN. 2016. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev*.
<http://genesdev.cshlp.org/content/early/2016/07/14/gad.284802.116> (Accessed December 30, 2019).

- Lima SA, Chipman LB, Nicholson AL, Chen Y-H, Yee BA, Yeo GW, Collier J, Pasquinelli AE. 2017. Short Poly(A) Tails are a Conserved Feature of Highly Expressed Genes. *Nat Struct Mol Biol* **24**: 1057–1063.
- Liudkovska V, Dziembowski A. Functions and mechanisms of RNA tailing by metazoan terminal nucleotidyltransferases. *WIREs RNA* **n/a**: e1622.
- MacDonald CC, McMahon KW. 2010. Tissue-specific mechanisms of alternative polyadenylation: testis, brain, and beyond. *Wiley Interdiscip Rev RNA* **1**: 494–501.
- Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-Pair Resolution Genome-Wide Mapping Of Active RNA polymerases using Precision Nuclear Run-On (PRO-seq). *Nat Protoc* **11**: 1455–1476.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.
- Mayr C. 2016. Evolution and Biological Roles of Alternative 3'UTRs. *Trends in Cell Biology* **26**: 227–237.
- Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTR s by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**: 673–684.
- Millevoi S, Vagner S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**: 2757–2774.
- Moore MJ, Blachere NE, Fak JJ, Park CY, Sawicka K, Parveen S, Zucker-Scharff I, Moltedo B, Rudensky AY, Darnell RB. 2018. ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *Elife* **7**.
- Muhlrad D, Parker R. 2005. The yeast EDC1 mRNA undergoes deadenylation-independent decapping stimulated by Not2p, Not4p, and Not5p. *EMBO J* **24**: 1033–1045.
- Mukherjee N, Jacobs NC, Hafner M, Kennington EA, Nusbaum JD, Tuschl T, Blackshear PJ, Ohler U. 2014. Global target mRNA specification and regulation by the RNA-binding protein ZFP36. *Genome Biol* **15**: R12.

Parameswaran N, Patial S. 2010. Tumor Necrosis Factor- α Signaling in Macrophages. *Crit Rev Eukaryot Gene Expr* **20**: 87–103.

Park J-E, Yi H, Kim Y, Chang H, Kim VN. 2016. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Molecular Cell* **62**: 462–471.

Patel RK, West JD, Jiang Y, Fogarty EA, Grimson A. 2020. Robust partitioning of microRNA targets from downstream regulatory changes. *Nucleic Acids Res* **48**: 9724–9746.

Piqué M, López JM, Foissac S, Guigó R, Méndez R. 2008. A combinatorial code for CPE-mediated translational control. *Cell* **132**: 434–448.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172–177.

Rissland OS, Subtelny AO, Wang M, Lugowski A, Nicholson B, Laver JD, Sidhu SS, Smibert CA, Lipshitz HD, Bartel DP. 2017. The influence of microRNAs and poly(A) tail length on endogenous mRNA–protein complexes. *Genome Biology* **18**: 211.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. *Nat Biotechnol* **29**: 24–26.

Sandler H, Kreth J, Timmers HTM, Stoecklin G. 2011. Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. *Nucleic Acids Res* **39**: 4373–4386.

Sanduja S, Blanco FF, Dixon DA. 2011. The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdiscip Rev RNA* **2**: 42–57.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.

Smibert P, Miura P, Westholm JO, Shenker S, May G, Duff MO, Zhang D, Eads BD, Carlson J, Brown JB, et al. 2012. Global Patterns of Tissue-Specific Alternative Polyadenylation in *Drosophila*. *Cell Rep* **1**: 277–289.

- Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**: D535-539.
- Stoecklin G, Stubbs T, Kedersha N, Wax S, Rigby WF, Blackwell TK, Anderson P. 2004. MK2-induced tristetraprolin:14-3-3 complexes prevent stress granule association and ARE-mRNA decay. *EMBO J* **23**: 1313–1324.
- Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**: 66–71.
- Tian B, Pan Z, Lee JY. 2007. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res* **17**: 156–165.
- Tiedje C, Diaz-Muñoz MD, Trulley P, Ahlfors H, Laaß K, Blackshear PJ, Turner M, Gaestel M. 2016. The RNA-binding protein TTP is a global post-transcriptional regulator of feedback control in inflammation. *Nucleic Acids Res* **44**: 7418–7440.
- Tiedje C, Ronkina N, Tehrani M, Dhamija S, Laass K, Holtmann H, Kotlyarov A, Gaestel M. 2012. The p38/MK2-Driven Exchange between Tristetraprolin and HuR Regulates AU-Rich Element-Dependent Translation. *PLOS Genetics* **8**: e1002977.
- Warkocki Z, Liudkovska V, Gewartowska O, Mroczek S, Dziembowski A. 2018. Terminal nucleotidyl transferases (TENTs) in mammalian RNA metabolism. *Philos Trans R Soc Lond B Biol Sci* **373**.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6232586/> (Accessed July 31, 2020).
- Weill L, Belloc E, Bava F-A, Méndez R. 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol* **19**: 577–585.
- Wells DG, Dong X, Quinlan EM, Huang YS, Bear MF, Richter JD, Fallon JR. 2001. A role for the cytoplasmic polyadenylation element in NMDA receptor-regulated mRNA translation in neurons. *J Neurosci* **21**: 9541–9548.
- Woo YM, Kwak Y, Namkoong S, Kristjánisdóttir K, Lee SH, Lee JH, Kwak H. 2018. TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Reports* **24**: 3630-3641.e7.

Wu L, Fan J, Belasco JG. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A* **103**: 4034–4039.

Wu L, Wells D, Tay J, Mendis D, Abbott MA, Barnitt A, Quinlan E, Heynen A, Fallon JR, Richter JD. 1998. CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. *Neuron* **21**: 1129–1139.

Wu Q, Medina SG, Kushawah G, DeVore ML, Castellano LA, Hand JM, Wright M, Bazzini AA. 2019. Translation affects mRNA stability in a codon-dependent manner in human cells eds. N. Sonenberg, K. Struhl, and J.S. Weissman. *eLife* **8**: e45396.

Xu N, Chen CY, Shyu AB. 1997. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol* **17**: 4611–4621.

Zhang D, Guelfi S, Garcia-Ruiz S, Costa B, Reynolds RH, D'Sa K, Liu W, Courtin T, Peterson A, Jaffe AE, et al. 2020. Incomplete annotation has a disproportionate impact on our understanding of Mendelian and complex neurogenetic disorders. *Science Advances* **6**: eaay8299.

Zheng D, Ezzeddine N, Chen C-YA, Zhu W, He X, Shyu A-B. 2008. Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. *The Journal of Cell Biology* **182**: 89.

CHAPTER 3²

Genome-Wide Identification of Polyadenylation Dynamics with TED-Seq

3.1 Abstract

Polyadenylation and deadenylation of mRNA are major RNA modifications associated with nucleus-to-cytoplasm translocation, mRNA stability, translation efficiency, and mRNA decay pathways. Our current knowledge of polyadenylation and deadenylation has been expanded due to recent advances in transcriptome-wide poly(A) tail length assays. Whereas these methods measure poly(A) length by quantifying the adenine (A) base stretch at the 3' end of mRNA, we developed a more cost-efficient technique that does not rely on A-base counting, called tail-end-displacement sequencing (TED-seq). Through sequencing highly size-selected 3' RNA fragments including the poly(A) tail pieces, TED-seq provides an accurate measure of transcriptome-wide poly(A)-tail lengths in high resolution, economically suitable for larger-scale analysis under various biologically transitional contexts.

3.2 Introduction

Poly(A) tail is one of the longest known classical mRNA modifications with multiple molecular functions. It is the binding site for poly(A) binding proteins (PABP) that protects the 3' ends of mRNA from exonuclease mediated decay, and serves as a recruiter of translation initiation factors [1, 2]. Deadenylation—removal of poly(A) tail—is a critical process in mRNA decay mediated by CCR4-NOT deadenylases with

² This work is adaptation of a manuscript of the same name published on October 26th, 2021 in *Methods in Molecular Biology*. The authors of this manuscript were Yeonui Kwak and Hojoong Kwak.

3'–5' exonuclease activity, resulting in the shortening of poly(A) tails [3].

Polyadenylation, on the other hand, occurs co-transcriptionally during nascent RNA synthesis, coupled to 3' cleavage/polyadenylation site (PAS) formation [4].

Cytoplasmic polyadenylation can also occur later in the mRNA life span under specific biologically transitional contexts, most well known in early embryogenesis [5]. Therefore, monitoring the poly(A) tail lengths is increasingly becoming more important as one of the critical measures of post-transcriptional processes.

Traditionally, poly(A) tails have been measured through nuclease protection assays [(Swanson et al. 2002)]. mRNA poly(A) tail, hybridized to oligo- dT DNA probes, is susceptible to RNase H degradation, and the comparison between the poly(A) intact and poly(A) degraded RNA will yield the poly(A) tail length. More recent procedures include 3' ligation-mediated reverse transcription, and use of polymerase chain reaction (lmRT-PCR) to amplify the 3' fragments of the mRNA including the full poly(A) tail of the specific transcript of interest [6]. However, these methods have limited throughput and require prior knowledge of the gene of interest.

Through the advances of the next-generation sequencing technologies, new transcriptome-wide poly(A) tail length methods have been developed. Measuring the poly(A) tail length through RNA sequencing is conceptually simple: count the number of A-bases. However, there are complications in accurately counting A-base homopolymer stretches using the currently prevailing next-generation sequencing technology that depends on fluorescent-base incorporation cycles. Incomplete quenching of the fluorescence signal from the previous base read cycle can result in a contaminated signal in the following read cycle, and this can become more problematic in reading through the repeat of the same bases. This sequencing

ambiguity has been circumvented by methods such as PAL-seq or TAIL-seq [7, 8], where the poly(A) tails are either indirectly quantified using an additional fluorescent signal reporting A-base homopolymer abundance, or incorporating a custom fluorescence signal analysis of the ambiguous read cycles. However, these methods are device dependent, requiring modifications to the next-generation sequencing devices or analysis software. Direct long-read sequencing methods using nanopores, such as FLAM-seq [9], can also be used to count poly(A) lengths, but currently has lower throughput than the previous methods, which makes larger-scale applications cost-limited.

TED-seq is based on simple arithmetic that the length of an RNA fragment, including poly(A) tail and its flanking region, is equal to the addition of the poly(A) tail length and the distance from fragment start site to 3' PAS (Figure 3.1) [10]. Once the cDNA fragments are precisely size-selected, only sequencing and mapping the start of the cDNA fragment is necessary to calculate poly(A) tail length without the need to sequence through the poly(A) tail. The precise size-selection of the cDNA is performed by high-resolution polyacrylamide gel electrophoresis (PAGE). Knowing the selected size, and distance from the mapped 5' end of the fragment to the 3' CPS, poly(A) tail length is easily calculated for each mapped sequence read which represents each mRNA molecule. To map the 5' end of the fragment to the genome, 30-40 base reads are sufficient rather than sequencing up to 250 bases of A-bases, which greatly enhance the cost efficiency of TED-seq.

Outline of the major procedure: poly(A) RNAs are purified using Oligo d(T)-linked beads; purified poly(A) RNAs are ligated to 3' adaptor; RNA fragmentation by base hydrolysis; the resultant is subjected to 5' end repair and 5' adaptor ligation; reverse

transcription and PCR amplification selectively amplify cDNA fragments including poly(A) tail; high-resolution native PAGE purification of the poly(A)-cDNA library at a specific length (300 base pair); next-generation sequencing of the library from the 5' end (Figure 3.2). The sequence reads can be mapped to the reference genome or reference transcriptome sequences, and the distance from the mapped sites to the downstream 3' CPS is subtracted from 300 base pairs to yield the poly(A) tail length.

TED-seq can perform as an RNA quantification method as well as the marker of posttranscriptional RNA regulation. Conceptually, it is similar to 3' sequencing methods (3'-seq) [11] in that the reads are derived from regions near 3' PAS, which is known to perform robustly for RNA quantification. Quantification pipelines established in 3'-seq can be adopted for TED-seq with minimal modification. Also, TED-seq can be versatile for conjugating with other modified RNA detection methods, such as after RNA immunoprecipitation [12] or metabolic RNA labeling [13, 14]. TED-seq is compatible with standard, unmodified next-generation sequencing platforms, which makes it easier to merge into existing RNA methods as a library preparation module. TED-seq can be even more powerful when used in combination with other RNA-seq methods, such as nascent RNA sequencing [15] and RNA stability measures [16], which will provide a complete set of mRNA regulation from its synthesis to decay. It will also be compatible with any upgrades in the next-generation sequencing devices, since TED-seq is device-independent.

TED-seq may have drawbacks in terms of resolution, isoform distinction, and the requirement of input material amount. The resolution of poly(A) tail length measurement is dependent on the precision of the library size selection by PAGE, typically about 20 bases. While it does not reach the single-base resolution of poly(A)

length, the 20 base resolution may be sufficient since the binding footprint of PABP encompasses about 20 bases, and mechanistically, the deterministic factor of the role of poly(A) length is dependent on the number of bound PABPs [17]. Also, TED-seq depends on preidentified 3' PAS, and if the 3' PAS is either ambiguous or multiple 3' end isoforms are clustered, it will be difficult to assign the distance between the mapped reads and 3' PASs. However, the majority of the annotated transcripts have sharply defined 3' PASs within 10–20 bases, and the 3' alternative polyadenylation or alternatively spliced isoforms are usually hundreds of bases apart. The amount of required input RNA may be in a rather higher range (5–10µg of total RNA or >200 ng of poly(A) RNA). This is due to the size selection step, where only a fraction of the cDNA from the fragmented RNA is recovered. However, adding an additional amplification and cDNA fragmentation step prior to the size selection can resolve the input requirement, which will make it compatible with less amount of input material. Overall, while limitations of TED-seq do exist, they are outweighed by its cost-effectiveness, versatility, and the potential for further improvements. Herein, we describe the experimental details of TED-seq and present a preliminary data processing pipeline.

3.3 Materials

3.3.1 Poly(A) RNA Isolation

1. Poly(A) RNA isolation kit (Ambion, Dynabeads mRNA purification kit): oligo-dT Dynabead, Binding buffer, Wash buffer (see Note 1).
2. Magnetic tube rack.
3. Tube rotator.
4. Heat block set to 65°C.

5. Fluorometric nucleic acid quantification device and high sensitivity RNA detection reagent.

6. (Optional) poly(A) spike-in RNA (see Note 2).

3.3.2 3' RNA Adaptor Ligation

1. 3' RNA adaptor (RA3): 10 μ M of 5'-/phosphate/-
rUrGrGrArArUrUrCrUrCrGrGrGrUrGrCrCrArArGrG-/inverted-dT/-3' (see Note 3).

2. 10X T4 RNA ligase buffer.

3. 50% PEG-8000 (New England Biolabs).

4. 10 mM ATP.

5. RNase inhibitor (10 units/ μ l).

6. T4 RNA ligase I.

7. Heat block set to 65°C.

8. Thermocycler.

3.3.3 RNA Cleanup

1. TRIzol reagent (Invitrogen).

2. Chloroform.

3. GlycoBlue (Ambion).

4. 100% isopropanol.

5. 75% ethanol.

3.3.4 RNA Fragmentation

1. 1 N NaOH.

2. 1 M Tris-HCl, pH 6.8.

3. Micro Bio-Spin P-30 Gel Column (Bio-Rad).

4. Poly(A) RNA isolation kit (Ambion, Dynabeads mRNA purification kit).

5. Heat block set to 65°C.

3.3.5 5' RNA Phosphorylation

1. Polynucleotide kinase (PNK).
2. 10X PNK buffer.
3. 10 mM ATP.
4. RNase inhibitor (10 units/ μ l).

3.3.6 5' RNA Adaptor Ligation

1. 5' RNA adaptor (RA5): 10 μ M of 5'-
rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCrNrNrNrNrNrNrNr
N-3' (see Note 4).
2. RT primer (RTP): 50 μ M of 5'-GCCTTGGCACCCGAGAATTCCA-3'.
3. 10X T4 RNA ligase buffer.
4. 50% PEG-8000 (New England Biolabs).
5. 10 mM ATP.
6. RNase inhibitor (10 units/ μ l).
7. T4 RNA ligase I.
8. Heat block set to 65°C.
9. Thermocycler.

3.3.7 Reverse Transcription

1. Superscript II Reverse Transcriptase (Invitrogen).
2. 0.1 M DTT.
3. 5X First Strand buffer (Invitrogen).
4. RNase inhibitor (10 units/ μ l).
5. 12.5 mM dNTP mix: 12.5 mM dATP, 12.5 mM dCTP, 12.5 mM dGTP, 12.5 mM dTTP.
6. Thermocycler.

3.3.8 First Round Amplification of the Library

1. Short RP1 primer (shRP1 primer): 10 μ M of 5'-GTTCAGAGTTCTACAGTCCGA-3'.
2. RTP primer for PCR: 10 μ M of 5'-GCCTTGGCACCCGAGAATTCCA-3'.
3. High-fidelity hot-start PCR premix (2X).
4. Thermocycler.

3.3.9 PCR Cleanup Using SPRI Beads

1. Ampure XP beads (Beckman Coulter).
2. 75% ethanol.
3. Magnetic tube rack.

3.3.10 Size Selection of the Library

1. 5X TBE: 0.45 M tris–borate pH 8.3, 10 mM EDTA.
2. 6% PAGE gel, 16–20 cm of vertical height: 6% acrylamide, 0.5X TBE, 1% APS, 0.1% TEMED.
3. Vertical gel electrophoresis module, 16–20 cm height.
4. Power supply.
5. 6X gel loading dye, orange G.
6. 100 bp DNA ladder.
7. 25 bp DNA ladder.
8. (Optional) 10 bp DNA ladder.
9. Fluorescent DNA gel staining reagent.
10. Blue light gel illuminator.
11. TE-TW buffer: 10 mM tris, pH 8.0, 1 mM EDTA, 0.01% Tween 20.
12. 37°C incubation chamber with rotator.
13. Spin X column (Sigma-Aldrich).
14. Ampure XP beads (Beckman Coulter).

15. 75% ethanol.

16. Magnetic tube rack.

3.3.11 Second Round Full Amplification of the Library

1. RP1 primer: 10 μ M of 5'-

AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA-
3'.

2. RPI-index primer: 10 μ M of 5'-

CAAGCAGAAGACGGCATAACGAGATJJJJJGTGACTGGAGTTCCTTGGCACCC
GAGAATTCCA-3' (see Note 5).

3. High-fidelity hot-start PCR premix (2X).

4. Thermocycler.

3.3.12 Second Size Selection and PCR Cleanup

1. 5X TBE: 0.45 M tris–borate pH 8.3, 10 mM EDTA.

2. 6% PAGE gel, 16–20 cm of vertical height: 6% acrylamide,
0.5X TBE, 1% APS, 0.1% TEMED.

3. Vertical gel electrophoresis module, 16–20 cm height.

4. Power Supply.

5. 6X gel loading dye, orange G.

6. 100 bp DNA ladder.

7. 25 bp DNA ladder.

8. (Optional) 10 bp DNA ladder.

9. Fluorescent DNA gel staining reagent.

10. Blue light gel illuminator.

11. TE-TW buffer: 10 mM tris, pH 8.0, 1 mM EDTA, 0.01%
Tween 20.

12. 37 °C incubation chamber with rotator.

13. Spin X column (Sigma-Aldrich).
14. Ampure XP beads (Beckman Coulter).
15. 75% ethanol.
16. Magnetic tube rack.
17. Fluorometric nucleic acid quantification device and highsensitivity DNA detection reagent.

3.3.13 TED-seq Data Analysis

1. UNIX compatible computing system with the following GNU software installed and accessible from the \$PATH variable:

bash shell, awk, sort, samtools [18], bedtools [19], STAR aligner [20].

2. Reference genome sequence file in a fasta format (`./genome/genome.fasta`).
3. Gene annotation file of the 3' cleavage polyadenylation site (CPS) in a bed format (`./gene/gene.bed`) (see Note 6).

3.4 Methods

Prepare all solutions in ultrapure DNase and RNase free water. Prepare and store all reagents on ice unless indicated otherwise. Use 1.5 ml microcentrifuge tubes unless indicated to use 0.2 ml PCR tubes compatible with thermocyclers. Use heat blocks to incubate at temperatures higher than room temperature, except when indicated to use a thermocycler. Use DNase and RNase free plasticwares.

3.4.1 Poly-A RNA Isolation

1. Adjust RNA volume to 50 μ l in water (5–10 μ g of total RNA). Heat RNA at 65°C for 2 min and place the tube on ice.
2. Prepare Oligo d(T) Dynabeads. Transfer 100 μ l of the beads from the kit to a microcentrifuge tube and place on a magnetic rack for 30 s. Discard the supernatant and wash the beads once with 50 μ l Binding Buffer. Place the tube on the magnetic

rack, and remove the supernatant after the supernatant gets clear. Add 50 μ l Binding Buffer to the beads, and mix beads thoroughly. Scale up accordingly if multiple samples are processed at once.

3. Mix the 50 μ l beads with 50 μ l RNA, followed by rotation on a mini rotator for 3 min at room temperature.

4. Wash the beads. Place the tube on the magnet for 30 s, and remove the supernatant. Resuspend the beads in 100 μ l Washing Buffer B, and remove the buffer. Repeat once more for a total of two washes.

5. Add 11.5 μ l water to the bead for the elution. Mix thoroughly by gently pipetting. Heat the beads at 65°C for 2 min in a heat block, and immediately place the tube on the magnetic rack. When the beads are clearly separated from the water, collect 10 μ l of supernatant containing eluted RNA.

6. Quantify poly-A isolated RNA using flourometer.

7. (Optional) Add poly(A) spike-in RNAs to the eluted RNAs (~1 ng spike-in per ~100 ng of poly-A isolated RNA; see Note 2).

3.4.2 3' RNA Adaptor Ligation

1. Add 4 μ l of 10 μ M RA3 to 9 μ l of the poly-A RNA from Subheading

3.4.1. Heat the RNA mix at 65°C for 40 s on a heat bock, then cool down on ice immediately.

2. Add the following reagents to the 13 μ l RNA mix in a PCR tube for the total reaction volume of 30 μ l: 3 μ l of 10X T4 RNA ligase buffer, 6 μ l of 50% PEG-8000, 3 μ l of 10 mM ATP, 2.5 μ l of RNase inhibitor, 2.5 μ l of T4 RNA ligase I (see Note 7).

3. Incubate at 20 °C for 6 h, followed by an infinite hold at 4°C in a thermocycler. The reaction can remain at 4°C up to overnight.

3.4.3 RNA Cleanup

1. Add 500 μ l TRIzol to the RNA ligation reaction and mix well (see Note 8).

2. Add 100µl chloroform and vortex for 30 s, followed by centrifugation at 15,000 x g for 5 min at 4°C. Collect aqueous layer in a new microtube (~300 µl).
3. Mix with 2 µl GlycoBlue, then add 1X volume of 100% isopropanol (~300 µl) and vortex. Centrifuge at 15,000 x g for 20 min at 4°C, discard the supernatant, and wash pellet by gently pipetting 100µl 75% ethanol. The bluish gray RNA pellet in 75% ethanol can be stored in -80 °C for up to at least 2–3 weeks.
4. Completely remove supernatant and air dry for 5 min. Dissolve the RNA pellet in 20 µl water.

3.4.4 RNA Fragmentation

1. Heat the RNA at 65°C for 40 s for denaturation, then cool down on ice immediately.
2. Add ice cold 5 µl 1 N NaOH to the RNA, and incubate on ice for 10 min. Add 25 µl 1 M Tris–HCl, pH 6.8 to stop the base hydrolysis reaction.
3. Prepare a P-30 minicolumn by allowing the column storage buffer to flow by gravity. Centrifuge the column at 1000 x g for 2 min at room temperature to remove all buffers. Place the column on a new tube.
4. Transfer 50 µl of base hydrolyzed RNAs to the column. Pass the RNA through the P-30 column by centrifuging at 1000 x g for 2 min at room temperature.
5. Using the poly(A) RNA isolation kit, repeat Subheading 3.4.1, and elute in 20 µl water (see Note 9).

3.4.5 5' RNA Phosphorylation

1. Add the following reagents to the 20 µl of poly(A) RNA from the previous step: 3 µl of 10X PNK buffer, 3 µl of 10 mM ATP, 2 µl of RNase inhibitor, and 2 µl PNK enzyme (see Note 10).
2. Incubate the reaction at 37°C for 1 h.

3. RNA cleanup by repeating Subheading 3.4.3 (see Note 11). Do not dissolve the RNA pellet in water. Proceed with the precipitated RNA pellet.

3.4.6 5' RNA Adaptor Ligation

1. Add 3 μ l of 10 μ M RA5 adaptor and 1 μ l of 50 μ M RT primer to the RNA pellet. Incubate at RT for 1 min, then vortex and spin down to dissolve the RNA with the adaptor and the primer. Heat the sample at 65°C in a heat block for 20 s, and cool down on ice immediately. Transfer all (4 μ l) of the mix to a PCR tube.
2. Add the following reagents to the sample (total reaction volume: 10 μ l): 1 μ l of 10X RNA ligase buffer, 2 μ l of 50% PEG-8000, 1 μ l of 10 mM ATP, 1 μ l of RNase inhibitor, and 1 μ l of T4 RNA ligase (see Note 12).
3. Incubate at 20°C for 6 h, and hold out at 4°C on a thermocycler overnight.

3.4.7 Reverse Transcription (RT)

1. Directly add the following reagents to the 10 μ l RNA ligation: 2.8 μ l of 5X FS buffer, 1.4 μ l of 0.1 M DTT, 0.8 μ l of 12.5 mM dNTP, 1 μ l of RNase inhibitor, and 1 μ l of Superscript II RTase (see Note 13).
2. Proceed with the RT reaction by incubating at 50°C for 1 h.

3.4.8 First Round Amplification of the Library

1. Make a PCR mix by mixing the following reagents: 25 μ l of 2X PCR premix, 2.5 μ l of 10 μ M short RP1 primer, 2.5 μ l of 10 μ M RTP primer, and 3 μ l of water. Add 33 μ l of the PCR mix to the 17 μ l RT reaction (see Note 14).
2. Incubate in the thermocycler at 98°C 2 min, 8 cycles of 98°C for 30 s, 64°C for 30 s, and 72°C for 30 s, followed by 72°C extension for 5 min (see Note 15).

3.4.9 PCR Cleanup Using SPRI Beads

1. Add 1X volume (50 μ l) of Ampure XP beads to the PCR reaction thoroughly by pipette mixing. Place the reaction at room temperature for 5 min.
2. Incubate on a magnetic rack for 5 min and discard the clear resolution.

3. Add 200 μ l of 75% ethanol and incubate for 30 s while the tube is placed on the magnetic rack. Discard the ethanol and wash once more with 75% ethanol.
4. Briefly spin down, place on the magnetic rack, and completely remove any trace of ethanol. Allow the bead to dry for 5 min on the magnetic rack with the cap open.
5. Add 12 μ l of water to each tube with dried magnetic SPRI beads and incubate at room temperature for 2 min. Place the mix on the magnet for 1 min and collect 10 μ l of the eluted supernatant.

3.4.10 Size Selection of the Library

1. Prepare a 6% native TBE polyacrylamide gel for a 16–20 cm length vertical electrophoresis unit (45 ml). Prerun the gel (Protean II xi gel) for 20 min at 40 mA in 0.5X TBE prepared from 5X TBE.
2. Mix the 10 μ l sample from the previous step with 2 μ l of 6X gel loading buffer. Load the sample with 25 bp and 100 bp DNA ladders at its both sides, which will be used later for measuring cDNA library size (Figure 3.3).
3. Run PAGE for 1 h 30 min (40 mA), or until 10–15 min after the orange dye completely passes through the gel.
4. Disassemble the unit and take out the gel. Stain the gel with SYBR gold reagent for 3–5 min (see Note 16).
5. Precisely excise a rectangular gel piece between 350 and 360 base pairs according to the DNA ladder (Figure 3.3; see Note 17). Place the excised gel piece into a 0.2 ml DNase-free PCR tube and grind it with a sterile pipette tip (Figure 3.4 a). Add 100 μ l TE-TW buffer to the gel and incubate it overnight at 37 °C with rotation.
6. Place the PCR tube with the cap removed, upside down on a Spin X column (Figure 3.4 b–g). Pass the eluate through a Spin X column using microcentrifuge for 5 min at 15,000x g. Approximately 100 μ l of eluant will be collected in the microcentrifuge tube.

7. Add 1x volume (100 μ l) of Ampure XP beads to the passthrough, and repeat steps in Subheading 3.4.9 for the DNA cleanup. Elute in 17 μ l of water.

3.4.11 Second Round Full Amplification of the Library

1. Make PCR mix by mixing the following reagents: 25 μ l of 2X PCR premix, 2.5 μ l of 10 μ M RPI primer, and 3 μ l of water (see Note 14).
2. Add 30.5 μ l PCR mix to the size selected cDNA library (356 bp) from Subheading 3.4.10.
3. Add 2.5 μ l of RPI-1 or other RPI-index primer (e.g., RPI-2, RPI-3, RPI-4) to each sample.
4. Perform PCR as follows: at 98°C 2 min, 5 cycles of 98°C for 30 s, 64°C for 30 s, and 72°C for 30 s, followed by 72°C extension for 5 min (see Note 15).

3.4.12 Second Size Selection and Cleanup

1. Clean up the PCR products using 1.8X volume (90 μ l) of Ampure XP beads, otherwise proceed as Subheading 3.4.9; elute in 10 μ l of water.
2. Pre-run the large 6% PAGE gel for 15–20 min.
3. Mix the 10 μ l sample with 2 μ l 6X loading dye. Load each sample into each well with the DNA size markers as described in Subheading 3.4.10, step 2. Electrophorese for 90 min at 40 mA.
4. Stain the gel with 1X SYBR Gold reagent diluted in 1X TBE for 3–5 min.
5. Using sterile forceps and a cutting blade, excise out between 420 and 440 bp (see Note 18). Place the excised gel piece in a PCR tube and grind it with a pipette tip. Add 50 μ l of TE-TW buffer to the excised gel followed by incubation overnight at 37°C on a rotator.
6. Pass the gel mixture through a Spin X column. DNA cleanup using 1X volume (50 μ l) of Ampure XP beads; repeat steps in Subheading 3.4.9 and elute in 10 μ l of water.

7. Quantify the library using a DNA fluorometer, and normalize the library to 2 ng/ μ l. Multiplex the libraries as needed. Proceed with the single-end Illumina sequencing compatible with TRU-seq small RNA adaptors.

3.4.13 TED-seq Data Visualization

1. Download the result Illumina sequencing fastq file (TEDseq.fastq) to the working directory (./) of the UNIX compatible system.
2. Extract first 8 bases of Unique Molecular Identifier (UMI) from the sequence reads and append it to the sequence identifiers by executing the following awk command:

```
awk '{id=$1;getline;tag=substr($1,1,8); \
seq=substr($1,9);getline; \
phred=substr($1,9);if(length(seq)>=16) \
printf id":tag"\n"seq"\n+\n"phred"\n"}' \
./TEDseq.fastq > ./TEDseq.UMItag.fastq
```

3. Generate genome index file for STAR aligner by executing the following command (see Note 19):

```
STAR --runMode genomeGenerate \
--genomeDir ./genome \
--genomeFastaFiles ./genome/genome.fasta
```

4. Align TED-seq reads to the genome:

```
STAR -- genomeDir ./genome \
--readFilesIn ./TEDseq.UMItag.fastq
--outFilterMultimapMax 1 \
--outFileNamePrefix TEDseq
```

5. Collapse identical UMIs of the alignment file by executing the following lines (note that temporary files are created and deleted):

```
samtools view -S TEDseq.Aligned.out.sam | \
```

```

awk '{n=length($1);print substr($1,n-7,8)"\t"$0;}' \
> _sam.tmp
samtools view -SH TEDseq.Aligned.out.sam \
> _umi_unique_sam.tmp
sort -k4,4 -k5,5n -k1,1 -u _sam.tmp | \
cut -f2- >> _umi_unique_sam.tmp
samtools view -Sb _umi_unique_sam.tmp >
TEDseq.uniqueUMI.bam
rm _sam.tmp _umi_unique_sam.tmp

```

6. Generate strand specific bedgraph files that can be loaded on genome browser softwares, such as Integrative Genomics Viewer (IGV) [21].

```

bedtools genomecov -ibam TEDseq.uniqueUMI.bam -bg -
strand + -5 \
> TEDseq.pl.bedgraph
bedtools genomecov -ibam TEDseq.uniqueUMI.bam -bg -
strand - -5 | \
awk '{print $1"\t"$2"\t"$3"\t"$4*-1}' > \
> TEDseq.mn.bedgraph

```

7. Browse the plus and minus strand bedgraph files on a genome browser. (Figure 3.5; see Note 20).

3.5 Notes

1. Oligo(dT) magnetic beads can be made custom by conjugating 3'-amino modified oligo-dT and carboxylic acid linked magnetic beads using N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide (EDC).

2. Poly(A) spike-in RNA can be prepared by in vitro transcription of PCR amplified unique sequence template with T7 promoter sequence at the 5' end of the forward primer and poly-dT sequence of the desired length at the 5' end of the reverse primer. The unique sequence template can be an arbitrary sequence of 700–800 base pairs from a plasmid backbone, that is not found in the target organism's genome, and that does not contain more than 4–5 consecutive T's on the sense strand which might serve as an internal termination signal for T7 RNA polymerase.
3. The RA3 adaptor is an RNA oligonucleotide with 5' phosphorylation and 3' inverted dT modifications. Alternatively, preadenylated DNA oligonucleotide can be used, but will need to adjust the RNA ligation reaction by replacing the RNA ligase I with the truncated version of RNA ligase II enzyme without the presence of ATP.
4. The RA5 adaptor is an RNA oligonucleotide that contains 8 random nucleotide sequences (N's) that serve as Unique Molecular Identifier (UMI). N's are equal compositions (25%) of A, C, G, and U bases.
5. RPI-index primers contain sample barcode index sequences (JJJJJ) that comply with single ended Illumina TRU-seq small RNA sequencing primers (RPI-1: CGTGAT, RPI-2: ACATCG, RPI-3: GCCTAA, RPI-4: TGGTCA, RPI-5: CACTGT, RPI-6: ATTGGC; etc.)
6. Ideally, 3' CPS annotation should be from a 3' end sequencing data in the same biological sample. Alternatively, the last 300 bases of transcripts from reference gene annotation, such as RefSeq, can be used. The TED-seq reads will be positioned in the 300 bases region.
7. Mix thoroughly by gently pipetting as PEG-8000 is highly viscous. The reagents can be premixed when processing multiple samples at once. Scale up accordingly, and add 17 μ l of the premix to each reaction.

8. The remaining steps in Subheading 3.4.3 can be replaced by other column-based or SPRI bead-based RNA cleanup procedures. Adjust the final volume to 20 μ l in water. TRIzol procedure is preferred if the procedure cannot resume immediately, and extended storage of the material is needed at Subheading 3.4, step 3.
9. This step removes hydrolyzed RNA fragments that do not contain poly(A) tail regions, and remaining RA3 adaptors. It is possible to skip this step (but not currently recommended), which results in products with higher internal RNA reads and adaptor dimers.
10. The reagents can be premixed when processing multiple samples at once. Scale up accordingly, and add 10 μ l of the premix to each reaction.
11. The next step requires small volume reactions, and precipitated RNA suits better for this purpose. It is also possible to use other column-based or SPRI bead-based RNA cleanup procedures. In these cases, use appropriate mixtures of RA5 adaptor and RTP primer in water for the final elution.
12. As in Note 7, make sure to mix thoroughly by gently pipetting. The reagents can be premixed when processing multiple samples at once. Scale up accordingly, and add 6 μ l to each reaction.
13. The reagents can be premixed when processing multiple samples at once. Scale up accordingly, and add 7 μ l to each reaction.
14. 2X PCR premix from commonly obtainable sources or custom mixes containing hot-start high-fidelity thermostable DNA polymerase can be used. For multiple samples, scale up accordingly.
15. Optimal number of PCR cycles may be determined empirically. Typically, 1 μ l of the input material is serially diluted by 4 folds, and subject to 20–25 cycles of PCR. Alternatively, 5 μ l of test PCR products in \sim 50 μ l PCR reaction can be taken out every 2 cycles. Over-amplified products will appear as an upshift of the smear due to self-

priming. Determine the number of cycles by subtracting 4–6 cycles before over-amplification happens.

16. Make sure that the plate for gel staining is clean to minimize the possible contamination to the cDNA library.

17. Target product size is precisely at 356 bp (insert 300 bp + PCR primer and UMI 56 bp). 10 bp DNA ladder may not show up to this size (typically up to 330 bp). Using a 25 bp DNA ladder and 350 bp mark (Figure 3.3, dashed line) as a guide, cut out a thin (~1 mm) slice of the gel. Loading the DNA ladder on both sides of the sample and cut along a straight line between the two 350 bp marks. The insert size can be other than 300 bp, such as 250 bp, as long as the products are precisely size-selected. The advantage of using a shorter size is that the on-gel manipulation is easier. The drawback of using shorter size (e.g., 250 bp) is that it will not be possible to detect poly(A) tails longer than the insert size, since $L = S - D$ (Figure 3.1). However, most eukaryotic poly(A) tails are known to be shorter than 250 bp, and may not affect many transcripts.

18. There will be a single band at 426 bp (cDNA insert: 300 bp, PCR primers: 126 bp), that may form a thicker smear due to self-priming. Use the 25 bp DNA ladder as a guide to cut out around 425 bp band. Try to include as much as possible, but cut out the tails of the smear to preserve the correct insert size.

19. This step needs to be done once, and the indexed genome files can be used for subsequent alignments. If the indexed genome file already exists, this step can be omitted.

20. Poly(A) tail length distribution appears as a displaced distribution of TED-seq reads by 300 bases upstream from the 3'CPS. For most transcripts, calculating the distance from the TED-seq reads to its downstream 3'CPS (D) is sufficient to yield poly(A) tail length (L) and using the formula $L = S - D$ (Figure 3.1). If there are

multiple 3'CPS within the 300 bp window, either use the most dominant 3'CPS or exclude the transcript from the analysis. On rare occasions, there may be a spliced intron within the 300 bp window. On those genes, there may appear to be accumulations of TED-seq reads at the splice junction due to the partial alignment of TED-seq reads. In such cases, spliced reads need to be treated separately to be mapped to the correct 5' end positions.

FIGURES

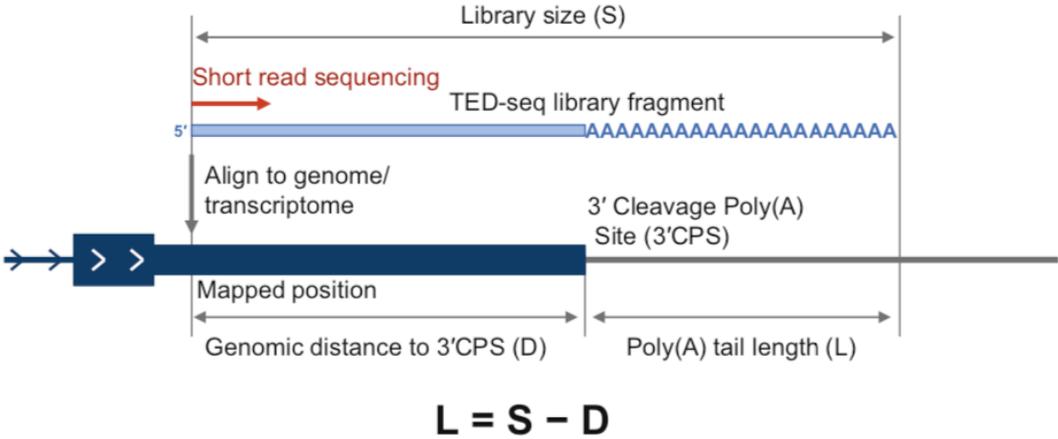


Figure 3.1. Schematics of poly(A) tail length calculation in TED-seq. TED-seq library fragment encompassing poly(A) tail is aligned at the 3' end of a gene. Typically, sequencing the first 40 bases is sufficient for the alignment (red arrow). Dark blue bar indicates the gene annotation: arrowheads pointing to the sense direction of transcription, thicker body reflect the coding sequence, thinner body reflect untranslated region (UTR), and line overlaid by arrowhead indicate spliced intron. From this diagram of poly(A) length (L), library size (S), and the distance from 3' cleavage polyadenylation site (CPS) to TED-seq read (D), $L = S - D$

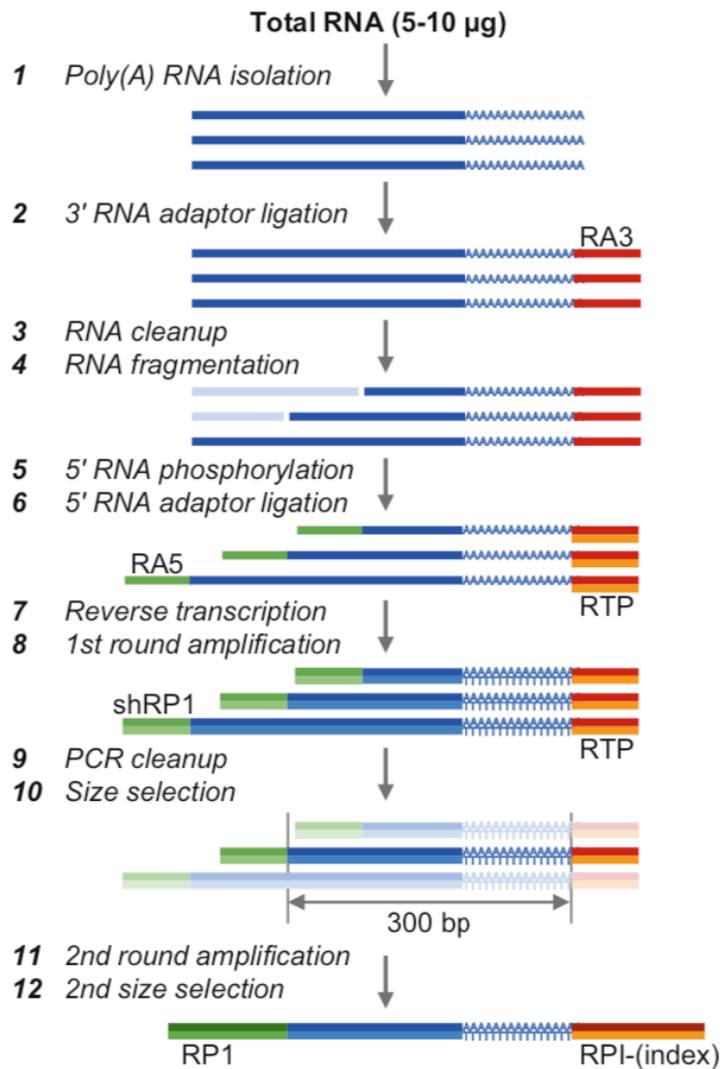


Figure 3.2. Overview of TED-seq experimental procedures shows all the experimental steps in Subheading 3.4. Blue bars with poly(A) tail represent mRNA. 3' and 5' RNA adaptors are shown as red (RA3) and green (RA5) bars, respectively. Reverse transcription primer (RTP) is in orange after step 6, annealed to the ligated RA3. RTP is also used as a PCR primer with the short RP1 primer (shRP1; step 8). Full length RP1 and RPI-index primers are used in the second amplification step (step 11)

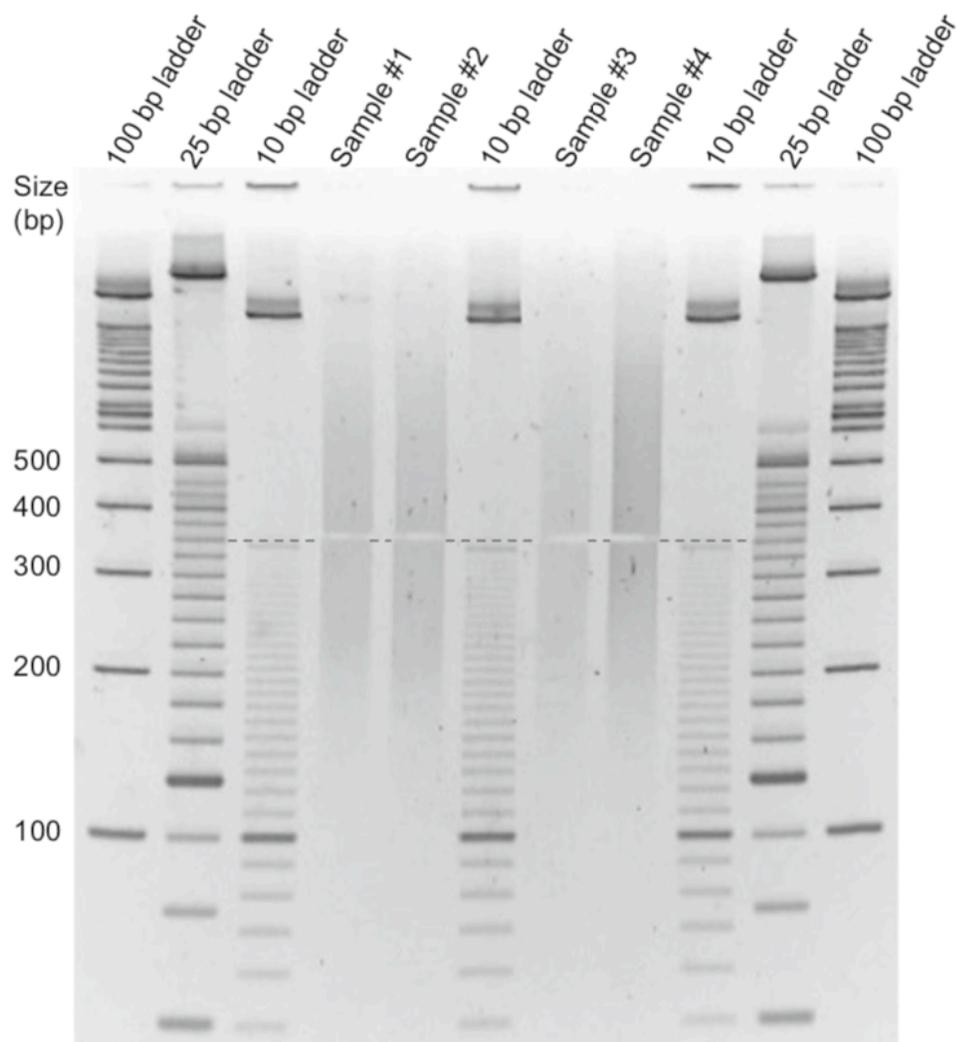


Figure 3.3. Electrophoresis for the size selection. Shown is a post-excision polyacrylamide gel. DNA size markers and samples are labeled on the top, and DNA size labeled on the left size. Dashed lines indicate 350 bp. Note the excised region within the smear of the sample (modified from Woo et al. (2018) Cell Rep)

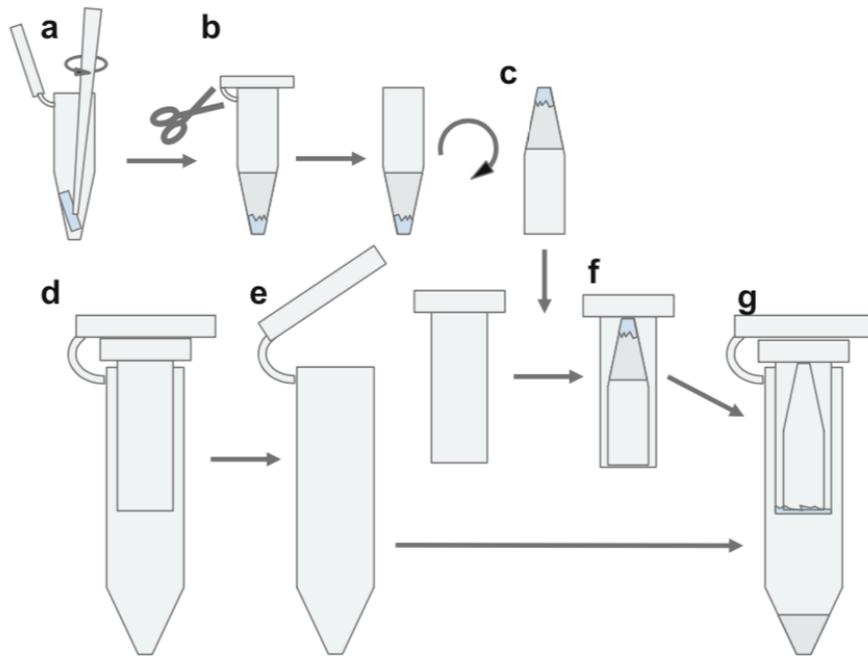


Figure 3.4. DNA elution from a polyacrylamide gel. (a) Grinding a small gel piece in a 0.2 ml PCR tube (Subheading 3.4.10, step 5). (b) Cutting out the cap of the PCR tube in Subheading 3.4.10, step 6. (c) Inserting the decapped PCR tube in the microspin filter unit. (d) Microspin (Spin-X) filter unit. (e) Inner filter unit detached. (f) Decapped PCR tube in the filter unit. (g) Reassembly before centrifugation.

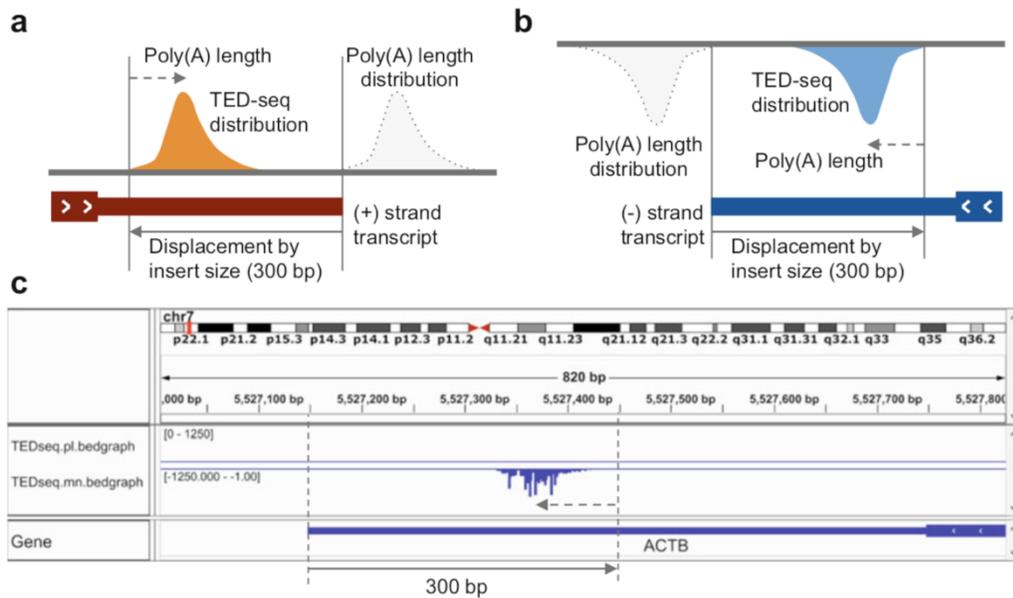


Figure 3.5. Example of a TED-seq result in a genome browser. (a) Schematics of visualizing poly(A) tail length distribution by TED-seq on the genome browser. Red bar on the bottom indicates the gene annotation of a (+) strand transcript. Dotted distribution reflects the poly(A) tail length distribution of the transcripts. Orange filled distribution reflects the TED-seq distribution which is the poly(A) tail length distribution shifted upstream by the library insert size (300 bp). (b) Schematics of TED-seq browser view of a (-) strand transcript. Blue bar indicates the gene annotation, and the light blue filled distribution is the TED-seq distribution, otherwise as described in (a). Note that the (-) strand reads are inverted horizontally on the negative side of the y-axis for visualization. (c) Example of TED-seq at human ACTB gene (data from Woo et al. (2018) Cell Rep) on the Integrative Genomics Viewer [21]. Note that ACTB gene is on the (-) strand. TED-seq distribution relative to the position 300 bp upstream of the 3' CPS (dashed line) is the poly(A) tail length distribution (dashed arrow) of ACTB transcripts

REFERENCES

1. Eckmann CR, Rammelt C, Wahle E (2011) Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* 2(3):348–361
2. Weill L, Belloc E, Bava FA et al (2012) Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol* 19(6):577–585
3. Mugridge JS, Collier J, Gross JD (2018) Structural and molecular mechanisms for the control of eukaryotic 5'-3' mRNA decay. *Nat Struct Mol Biol* 25(12):1077–1085
4. Colgan DF, Manley JL (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11(21):2755–2766
5. Richter JD (1999) Cytoplasmic polyadenylation in development and beyond. *Microbiol Mol Biol Rev* 63(2):446–456
6. Murray EL, Schoenberg DR (2008) Assays for determining poly(A) tail length and the polarity of mRNA decay in mammalian cells. *Methods Enzymol* 448:483–504
7. Subtelny AO, Eichhorn SW, Chen GR et al (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508 (7494):66–71
8. Chang H, Lim J, Ha M et al (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* 53 (6):1044–1052
9. Legnini I, Alles J, Karaïskos N et al (2019) FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat Methods* 16(9):879–886
10. Woo YM, Kwak Y, Namkoong S et al (2018) TED-Seq identifies the dynamics of poly(A) length during ER stress. *Cell Rep* 24 (13):3630–3641. e3637

11. Lianoglou S, Garg V, Yang JL et al (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* 27(21):2380–2396
12. Köster T, Haas M, Staiger D (2014) The RIP-per case: identification of RNA-binding protein targets by RNA immunoprecipitation. *Methods Mol Biol* 1158:107–121
13. Schwahnhauser B, Busse D, Li N et al (2011) Global quantification of mammalian gene expression control. *Nature* 473 (7347):337–342
14. Garibaldi A, Carranza F, Hertel KJ (2017) Isolation of newly transcribed RNA using the metabolic label 4-Thiouridine. *Methods Mol Biol* 1648:169–176
15. Mahat DB, Kwak H, Booth GT et al (2016) Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* 11 (8):1455–1476
16. Ross J (1995) mRNA stability in mammalian cells. *Microbiol Rev* 59(3):423–450
17. Mangus DA, Evans MC, Jacobson A (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* 4(7):223
18. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079
19. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842
20. Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21
21. Robinson J, Thorvaldsdóttir H, Winckler W et al (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–2

CHAPTER 4

Conclusions and Future Directions

4.1 Conclusions

Great progress has been made in the past few years in understanding the distribution and dynamics of poly(A) tail length in early developmental processes such as oocyte maturation and early embryogenesis, which was found conserved across different species. Now, we have a foundation to explore this regulation in various physiological contexts beyond embryonic systems. However, in post-embryonic systems, due to the presence of active and dynamic transcriptional regulation, quantifying post-transcriptional tail length change is challenging. Multiple scenarios could explain poly(A) tail length changes in somatic cells upon stimulation. For example, increased poly(A) tail length can be explained by (1) rapid accumulation of new transcripts derived from a signal-induced transcriptional burst, (2) increased co-transcriptional polyadenylation, (3) decreases in deadenylation, and/or (4) cytoplasmic readenylation. In this thesis, I analyzed the dynamics of poly(A) tail length in human macrophage activation on a genome-wide scale by employing various transcriptomic approaches, including measurements of mRNA levels, poly(A) tail length dynamics, and transcriptional regulation. A combined analysis of these data uncovered widespread and dynamic gene regulation coupled with changes in poly(A) tail length across the macrophage activation, including evidence of readenylation for many mRNAs.

First, I described the development of a novel technique to investigate poly(A) tail dynamics. Although recent advances in transcriptome-wide poly(A) tail length assays, including PAL-seq (Subtelny et al. 2014) and TAIL-seq (Chang et al. 2014), expanded our knowledge of poly(A) tail length control, these previous methods are not broadly

available to other researchers due to the requirement of modifications of sequencing hardware or software. Also, these methods cannot discern the mRNA isoforms to which the poly(A) tails are attached. In an effort to resolve the disadvantages of previous methods, we developed a new poly(A) tail length profiling method called tail-end-displacement sequencing (TED-seq). TED-seq is an accessible, affordable, and quantitative version of global poly(A) tail length profiling method that is compatible with commonly used short-read sequencing platform (i.e., Illumina) and thus can be broadly accessible to other researchers (Woo et al. 2018; Kwak and Kwak 2022). TED-seq still employs short-read sequencing, but by fixing the library insert size and integrating the 3' cleavage site information, it estimates the tail length at the 3'UTR isoform level. Therefore, TED-seq allows the broader community to have easy access to a global-scale and accurate poly(A) tail length measurement.

In the second part, I applied TED-seq to macrophage activation, revealing that a large fraction of the transcriptome undergoes changes in poly(A) tail length upon macrophage activation, including transient increases for pro-inflammatory genes. By quantifying nascent transcription, we found that even the genes with minimal transcriptional regulation exhibit significant changes in poly(A) tail length upon stimulation, indicative of the change of poly(A) tail length being post-transcriptional. Additionally, increases in tail length correlated with mRNA levels regardless of transcriptional activity, indicating a post-transcriptional regulation of mRNA abundance via poly(A) tail length control. Poly(A) tail length control is often mediated by interactions between 3'UTR *cis* elements and *trans*-acting factors. AU-rich elements (AREs) are well characterized 3'UTR *cis*-elements that mediate rapid decay of many short-lived mRNAs, in particular, of cytokine and chemokine transcripts during immune responses (Caput et al. 1986; Xu

et al. 1997). ARE-binding RBPs such as ZFP36 recognize AREs and destabilize mRNAs by recruiting deadenylation and decay factors (Lai et al. 2003; Sanduja et al. 2011). My analysis showed that 3'UTR AU content is globally associated with LPS-induced PAL changes and this association manifests near the 3' end of 3'UTRs. We also identified that many mRNAs undergo tail extension, and they tend to encode proteins necessary for immune function and post-transcriptional regulation. Readenylation was associated with increased mRNA abundance, implicating readenylation as a process responsible for stabilizing transcripts during macrophage activation. In humans, CPEB1 mediates cytoplasmic polyadenylation by binding to poly(U)-containing *cis* elements called cytoplasmic polyadenylation elements (CPE) during oocyte maturation (Hake and Richter 1994), and host-viral mRNA readenylation in cytomegalovirus infection (Batra et al. 2016). Our motif enrichment analysis identified poly(U) motifs in the 3'UTR as a mediator of readenylation upon activation in macrophage cells, validated by performing 3'UTR reporter assays. Strikingly, our data showed that *ZFP36* transcripts undergo tail extension upon activation. Our analyses indicate that many mRNAs undergoing tail lengthening are, in turn, degraded by elevated levels of ZFP36, constituting a post-transcriptional feedback loop that ensures transient regulation of transcripts integral to macrophage activation. Taken together, this study establishes the complexity, relevance and widespread nature of poly(A) tail dynamics, and the resulting post-transcriptional regulation during macrophage activation.

Altogether, my dissertational work introduced (1) a broadly accessible technique allowing systematic poly(A) tail length studies across the labs, and (2) an analytic framework to study post-transcriptional control of poly(A) tail length in transcriptionally active, cellular processes. By employing these new technique and analytic approaches, this work provides evidence that readenylation can be widely used, exerting a profound effect on gene expression in a non-developmental context.

4.2 Future Directions

There are many interesting avenues that future studies can explore for better understanding of poly(A) tail length regulation. Here I described several important questions and challenges that remain to be addressed in relevance to my work.

Developing TED-seq 2.0

One path of future exploration is to improve the current TED-seq method to decrease technical limitations or bias intrinsic to the TED-seq protocol. Like every other sequencing method does, TED-seq consists of numerous molecular steps in library preparation with each step having an aspect to bias the tail length measurement. Among them, I would like to discuss several steps that could generate potential artifacts in the sequencing library, and how we could improve them.

First, TED-seq enriches mRNAs out of ribosomal RNAs and small RNA molecules by capturing poly(A) tailed mRNAs with oligo(dT)-conjugated beads. However, this method can enrich longer poly(A) tailed mRNAs more preferentially than shorter poly(A) tails. Also, transcripts with short poly(A) tails (<25 nt) could be discarded during the mRNA purification, and thus their abundance may be underestimated than the actual level as well as their poly(A) tail length can be overestimated in the post-purification mRNA pool. We could resolve this technical bias by performing rRNA depletion instead of the use of oligo(dT) in TED-seq 2.0.

After mRNA isolation, the RNA samples undergo several molecular reactions: 3' adaptor ligation, random RNA fragmentation, 5' phosphorylation, 5' adaptor ligation. These molecular reactions generate a 3' RNA fragment between a 5' adaptor and a 3' adaptor on each end. The 3' RNA fragments have a full poly(A) tail, as well as the gene-derived sequence as an insert. Reverse transcription generates cDNAs, which are PCR-amplified using 5' adaptor and 3' adaptor sequences as PCR primers. One issue in the protocol is that the homopolymeric poly(A) region is not efficiently amplified, and therefore short-tailed fragments could be enriched during PCR, which could bias the tail length and mRNA abundance in the library pool. After the PCR reaction, the library proceeds with size selection to include inserts of a specific size with a narrow range of error (± 10 nt nucleotides), followed by sequencing. In TED-seq 2.0, we could modify the protocol in a way that poly(A) tail region doesn't go through the PCR reaction, eliminating the size bias. What if we size-select the library before cDNA synthesis so that the poly(A) tail regions do not need to be in the library? Once we know the size of RNA library, we can digest away the poly(A) regions before PCR amplification. One possible workflow would be as follows. After 5' adaptor ligation, the resulting RNA fragments undergo size selection through Urea PAGE gel. Reverse transcription using 3' adaptor sequence as a primer would generate first-strand cDNA. At this point, adding oligo-d(A) primers to the cDNA library would make the formation of double-strand DNAs only at the poly(A) tail, subject to double-strand specific DNases, resulting in poly(A) tail degradation. Subsequently, the reaction product would proceed with ligation at the 3' end with a single-strand DNA adaptor sequence, and finally with PCR amplification using the adaptor sequences as PCR handlers. Although requiring more input RNA, this approach will enable more accurate, less biased tail length estimation. This is only one alternative protocol, and the technique could benefit from additional modifications.

Lastly, TED-seq requires precise poly(A) site to estimate poly(A) tail length. Since current TED-seq does short-read sequencing, many of the reads do not read through the cleavage and polyadenylation sites, and therefore poly(A) site information should be obtained independently by 3'-seq. In TED-seq 2.0, we could adopt long-read sequencing (~300 nt) or paired-end sequencing (75 nt per end x 2), which enable us to process tail length information and cleavage and polyadenylation site information at one time.

New approach to measure poly(A) tail length dynamics in transcriptionally active systems

Owing to the extreme lability of poly(A) tail *in vivo* and the lack of proper methodologies to rigorously monitor poly(A) tail length along the time course of the mRNA lifecycle, current genome-scale studies profile poly(A) tails for every mRNA molecule in the cell, calculating the average poly(A) tail length for mRNAs derived from the same gene without telling poly(A) tail length dynamics across the mRNA lifespan. It cannot tell us how many new transcripts are synthesized entering the mRNA pool, how long poly(A) tails are initially added during nuclear 3' end formation, how poly(A) tail lengths change upon export to the cytoplasm. Therefore, most genome-wide studies inevitably assumed that averaged poly(A) tail length quantitatively summarizes post-transcriptional poly(A) tail length control, which seems reasonable in transcriptionally silent systems such as early developmental contexts, but inappropriate to establish the actual relationship between poly(A) tail and mRNA fate probably in most somatic cellular processes. In somatic cells, new transcripts continuously enter the mRNA pool in a somatic system, and the extent of transcription varies across the transcriptome, which variably affects the pool of mRNAs, and thus poly(A) tail length distribution across the transcriptome. Indeed, the

differences in mean poly(A) tail length well explain the variation in mRNA regulation such as mRNA stability or translation in early embryonic systems but not in somatic systems. Very recently, Eisen et al. first studied the global relationship between deadenylation rate and mRNA stability in steady-state cultured mouse fibroblasts by employing pulse metabolic labeling and computational modeling (Eisen et al. 2020a, 2020b). This study revealed that differences in deadenylation rates explain a substantial fraction of variation in mRNA stability with faster deadenylation corresponding to faster mRNA turnover. Therefore, poly(A) tail length control plays significant roles in gene expression regulation in a somatic cell line, and this information can be correctly quantified only when looking at the kinetics of poly(A) tail length across a time-course of mRNA lifecycle without being confounded by new transcripts.

This issue is more problematic in a study examining poly(A) tail length dynamics in response to environmental cues. Since environmental cues cause dramatic changes in transcriptional activity across the transcriptome, the distribution of poly(A) tail lengths can easily fluctuate to transcriptional changes, and thus it is challenging to discern post-transcriptional tail length changes from new transcripts' effect on poly(A) tail length distribution. In this thesis, I measured post-transcriptional tail length changes by quantifying nascent transcription and poly(A) tail length in a regular transcription condition, or by monitoring poly(A) tail length in a transcription shut-off condition using Actinomycin D to stop transcription. However, global inhibition of transcription itself is a major perturbation to the cell, and this has been shown to induce a general stress response, which could confound our results by affecting mRNA decay rates (Sun et al. 2012). Additionally, as reported in several papers, Actinomycin D has other nonspecific effects in the cells (Sun et al. 2012; Harigaya and Parker 2016), which could confound the results. Therefore, a more rigorous and sophisticated approach is

needed that can examine post-transcriptional poly(A) tail length control in a less disruptive manner.

A promising approach for capturing poly(A) tail length control in a less perturbing way is to adopt a classical “pulse-chase” transcription approach to a transcriptome level (Chen and Shyu 2011; Wilson and Treisman 1988; Greenberg and Ziff 1984). In the past, the dynamics of mRNA deadenylation or cytoplasmic polyadenylation have been studied on a gene-by-gene basis across species from yeast to mammals, where pulsed transcription was made using a Tet promoter-driven reporter gene or in vitro transcripts injection, and poly(A) tail lengths were subsequently examined (“chased”) for the pulsed reporter transcripts (Xu et al. 1998; Loflin et al. 1999). We could adopt this pulse-chase approach for endogenous genes and scale up across the transcriptome by combining a metabolic pulse-chase RNA labelling with a global poly(A) tail profiling method (Lugowski et al. 2018). This approach could directly quantify tail length dynamics without much assumption for computational modeling, and especially useful in the cellular response to an environmental cue, whereas the “pulse-only” approach used by Eisen et al. requires many extreme assumptions (e.g. constant transcription rate) for building a computational model to estimate poly(A) tail length dynamics (Eisen et al. 2020a, 2020b), some of which are not applicable in cells upon stimulation.

Characterizing poly(A) tail length dynamics in various biological contexts

Several groups have shown that poly(A) tail length control can be dynamically modulated in response to cellular cues using early developmental processes as a model. They found that transcripts are globally stored as translationally repressed and short-tailed mRNAs, waiting for the signals to activate readenylation for translational entry (Sheets et al. 1995; Lim et al. 2016). However, this phenomenon is thought to be

rare with one or two genes reported to be readenylated in some specific post-embryonic cases. Such bias in the biological contexts can partially be explained by the systematic advantage intrinsic to early developmental systems, which are transcriptionally dormant and thus allow for monitoring post-transcriptional poly(A) tail length changes without the need to consider confounding transcriptional changes. Somatic systems are more challenging to monitor for changes in poly(A) tail length, since new transcripts with longer tails continuously enter the mRNA pool. Now we can start to understand how widespread post-transcriptional tail length control is in other post-embryonic contexts through 1) the combination of poly(A) tail length profiling and nascent transcription measurements, 2) measuring the tail lengths in transcription shut-off condition, or 3) possibly employing a pulse-chase tail length measurement. First, it is of utmost importance to unveil poly(A) tail length dynamics in other non-developmental contexts. For example, how dynamically can the deadenylation process be modulated in response to cellular cues? How common is readenylation in other physiological process? how is the poly(A) tail length dynamics associated with cytoplasmic mRNA fate (e.g. translation, mRNA stability, and mRNA subcellular localization) in various somatic cellular contexts? Lastly, it is highly important to characterize differential poly(A) tail length control across tissues (or cell types). For example, is poly(A) tail length controlled in a tissue-specific manner? If it is, what are the molecular mechanisms and functional consequences of tissue-specific poly(A) tail length control? What are the trans-acting factors responsible for tissue-specific poly(A) tail length control? All these works would broaden our knowledge of the physiological relevance of appropriate poly(A) tail length control.

Deciphering cis and trans factors involved in poly(A) tail length control

In our study, we generated a model that poly(U) sequence mediates readenylation in macrophage activation, and partially validated the model using a reporter mRNA with or without poly(U) elements in the 3'UTR of ZFP36 transcripts, one of the readenylation candidates. However, there are several parts that remain elusive. First, we do not know yet the identity of RBP that binds to poly(U) sequence and mediate readenylation process, and the identity of poly(A) polymerase that catalyze readenylation in this system. To answer the first part of these questions, future studies could knock out or knock down RBPs known to bind poly(U) motifs using siRNA or CRISPR systems and repeat the tail length measurements during macrophage activation. It should be noted that my approach to finding cis-acting element and trans-acting factors involved in readenylation relies on the currently available database of RBP-binding sequence motifs. However, there might be many more RBPs with unknown binding motifs, thus not included in my motif search analysis. Additionally, the structural motifs were completely excluded in the analysis. Therefore, we should be aware that there might be other sequence or structural motifs and novel cognate RBPs that could play more significant roles than poly(U) sequence motifs in readenylation during macrophage activation.

While there are numerous proteins involved in nuclear polyadenylation, deadenylation and the regulatory steps in between that have already been described, more trans-acting factors that affect the poly(A) tail are yet to be covered. New RBPs involved in poly(A) tail length control could be discovered by combined analysis of poly(A) tail length and proteome across cell types/tissue and various physiological contexts involving spatial and/or temporal dimensions. Except for a few rare contexts such as oocytes and neuronal dendrites, the dependence of poly(A) tail length control on mRNA subcellular localization is poorly characterized, and thus it would be important

to develop a fine-resolution method to sharply probe RNAs on a transcriptome scale along spatial axis (e.g. neuron into soma and dendrites) or into subcellular compartments (e.g. cytoplasm into membrane-bound organelles, free cytosol, and membrane-less granules). Recently, an RNA proximity labeling technique, called APEX-seq, was developed to precisely resolve the localization of RNAs within the cell (Fazal et al. 2019). APEX-seq uses the ascorbate peroxidase APEX2 fused to a spatial marker protein to catalytically label the transcripts nearby the enzyme, which can precisely probe the spatial organization of the transcriptome within the cells, even in various subcellular compartments impossible to purify via biochemical fractionation-based approaches. Combining APEX-seq with poly(A) tail length measurements would allow us to reveal the range of the localized transcripts, their functional categories, and the relevance of subcellular localization with poly(A) tail length control while proteomic version of APEX-seq (APEX-MS) would find the proteins differentially expressed (DE) or differently post-translationally modified between the subcellular fractions (Kalocsay 2019). To this end, the knockdown screening of these DE proteins would unveil the RBPs directing the spatial control of poly(A) tail length.

In summary, given the integral role of poly(A) tail length control in various aspects of mRNA lifecycle, it would be great to systematically unveil 1) poly(A) tail length dynamics in more biological contexts from its initial biogenesis to dynamic control in the cytoplasm and ultimately decay, 2) understand the role of poly(A) tail length control on mRNA lifecycles, and ultimately, 3) determine the *cis* and *trans* factors underlying poly(A) tail length dynamics. These questions could be better addressed with the advent of new methodologies and technologies that accurately capture poly(A) tail kinetics and dynamics with high spatial and temporal resolution.

Particularly, identifying the relevance of abnormal tail length regulation in human diseases would be of utmost importance, possibly leading to the novel disease biomarkers that can be used as diagnostic tools in the future. Also, advanced knowledge in the molecular mechanisms of poly(A) tail length control would contribute to our understanding of how certain diseases develop.

REFERENCE

- Batra R, Stark TJ, Clark E, Belzile J-P, Wheeler EC, Yee BA, Huang H, Gelboin-Burkhart C, Huelga SC, Aigner S, et al. 2016. RNA-binding protein CPEB1 remodels host and viral RNA landscapes. *Nat Struct Mol Biol* **23**: 1101–1110.
- Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A. 1986. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc Natl Acad Sci U S A* **83**: 1670–1674.
- Chang H, Lim J, Ha M, Kim VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* **53**: 1044–1052.
- Chen C-YA, Shyu A-B. 2011. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* **2**: 167–183.
- Eisen TJ, Eichhorn SW, Subtelny AO, Bartel DP. 2020a. MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. *Molecular Cell* **77**: 775-785.e8.
- Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, Bartel DP. 2020b. The Dynamics of Cytoplasmic mRNA Metabolism. *Molecular Cell* **77**: 786-799.e10.
- Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, Chang HY, Ting AY. 2019. Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**: 473-490.e26.
- Greenberg ME, Ziff EB. 1984. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* **311**: 433–438.
- Hake LE, Richter JD. 1994. CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* **79**: 617–627.
- Harigaya Y, Parker R. 2016. Analysis of the association between codon optimality and mRNA stability in *Schizosaccharomyces pombe*. *BMC Genomics* **17**: 895.
- Kalocsay M. 2019. APEX Peroxidase-Catalyzed Proximity Labeling and Multiplexed Quantitative Proteomics. *Methods Mol Biol* **2008**: 41–55.
- Kwak Y, Kwak H. 2022. Genome-Wide Identification of Polyadenylation Dynamics with TED-Seq. In *Post-Transcriptional Gene Regulation* (ed. E. Dassi), *Methods in Molecular Biology*, pp. 281–298, Springer US, New York, NY https://doi.org/10.1007/978-1-0716-1851-6_15 (Accessed November 6, 2021).

- Lai WS, Kennington EA, Blackshear PJ. 2003. Tristetraprolin and its family members can promote the cell-free deadenylation of AU-rich element-containing mRNAs by poly(A) ribonuclease. *Mol Cell Biol* **23**: 3798–3812.
- Lim J, Lee M, Son A, Chang H, Kim VN. 2016. mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev*. <http://genesdev.cshlp.org/content/early/2016/07/14/gad.284802.116> (Accessed December 30, 2019).
- Loflin PT, Chen CY, Xu N, Shyu AB. 1999. Transcriptional pulsing approaches for analysis of mRNA turnover in mammalian cells. *Methods* **17**: 11–20.
- Lugowski A, Nicholson B, Rissland OS. 2018. Determining mRNA half-lives on a transcriptome-wide scale. *Methods* **137**: 90–98.
- Sanduja S, Blanco FF, Dixon DA. 2011. The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdiscip Rev RNA* **2**: 42–57.
- Sheets MD, Wu M, Wickens M. 1995. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature* **374**: 511–516.
- Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**: 66–71.
- Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Larivière L, Maier KC, Seizl M, Tresch A, Cramer P. 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* **22**: 1350–1359.
- Wilson T, Treisman R. 1988. Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. *Nature* **336**: 396–399.
- Woo YM, Kwak Y, Namkoong S, Kristjánssdóttir K, Lee SH, Lee JH, Kwak H. 2018. TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Reports* **24**: 3630–3641.e7.
- Xu N, Chen CY, Shyu AB. 1997. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol* **17**: 4611–4621.
- Xu N, Loflin P, Chen CY, Shyu AB. 1998. A broader role for AU-rich element-mediated mRNA turnover revealed by a new transcriptional pulse strategy. *Nucleic Acids Res* **26**: 558–565.

APPENDIX 1

Dynamic Regulation of Alternative 3'UTR Isoform Expression During Macrophage Activation and Relationship between Poly(A) tail Length and 3'UTR Isoform Identity

Abstract

More than a half of human genes contain multiple poly(A) sites in the 3' untranslated regions (3'UTRs), alternative use of which generates mRNA isoforms with different 3'UTR lengths. Alternative 3'UTR isoforms can have difference in RNA stability, translation, and mRNA trafficking through differences in the 3'UTR sequence. Although 3'UTRs often regulate gene expression via the deadenylation pathway, the relationship between alternative 3'UTR processing and poly(A) tail length control is not well understood. In the first part of this study, we measure the differences in poly(A) tail length regulation between alternative 3'UTR isoforms in unstimulated macrophage cells. Global regulation of alternative 3'UTR isoform expression is largely uncharacterized in many cellular contexts, including macrophage activation. In the second part, we examine dynamics of alternative 3'UTR isoform expression during macrophage activation.

Introduction

More than half of all human genes have multiple functional poly(A) site (PAS) in the 3' untranslated regions (3'UTRs), which can be alternatively cleaved and polyadenylated to generate different 3'UTR isoforms (Tian and Manley 2017). Particularly, alternative polyadenylation (APA) occurring in 3'UTRs leads to the production of mRNA isoforms with different 3'UTR lengths, which can have regulatory consequences. As 3' UTRs often have binding sites for microRNAs (miRNAs) and RNA-binding proteins (RBPs),

longer 3' UTRs typically contain more of these regulatory sequences than shorter 3'UTRs (Di Giammartino et al. 2011). Differential usage of poly(A) signals in the 3'UTR results in different cis regulatory elements in the mature mRNA, which may substantially affect mRNA stability, translation efficiency, and mRNA localization. For example, in neuronal cells, long isoform of BDNF, RAN, and several cytoskeleton transcripts tend to be transported into dendrites and axons rather than the cell body via a “zipcode” element present exclusively in the long isoform (Arora et al. 2022).

Recent global studies have shown that APA is widespread across the transcriptome; APA is known to be tissue-specific with the most notable examples being brain and testis preferentially expressing longer and shorter 3'UTR isoforms, respectively. APA is also globally regulated in various cellular conditions, such as cell proliferation and differentiation, and in response to extracellular cues (Sandberg et al. 2008; Mayr and Bartel 2009; Mayr 2016; Lianoglou et al. 2013). For example, before activation, genes important for T cell activation express long 3'UTR isoforms, many of which are subject to rapid mRNA degradation due to the presence of miRNA binding sites, leading to a low level of protein production (Domingues et al. 2016, 5). However, upon stimulation of T cells, these genes undergo 3'UTR shortening through APA (usage of proximal PAS), resulting in the removal of miRNA binding sites, and subsequent increase in mRNA abundance and translation leading to elevated protein levels. In addition, abnormal APA regulation is found in many oncological, immunological, neurological, and hematological diseases. There are likely numerous other cellular contexts involving widespread change in the APA landscape that are yet to be uncovered.

PAS choice is often attributed to the relative strength of cis-acting elements controlling site selection and the availability of trans-acting factors in concert with transcription dynamics, shaping the 3'UTR isoform expression preferences across different tissues and cell conditions. One important mechanism of APA regulation

involves modulation of the expression of core polyadenylation factors, with the most prominent example being Cleavage Factor I (CFI). Two of its constituent proteins, CFI-25 and CFI-68, display widespread regulation of APA (Zhu et al. 2018). CFI expression leads to enhanced usage of distal PASs over proximal PASs, since it has a binding affinity for UGUA motifs, and these motifs are more enriched near distal PASs over proximal PASs, thus leading to preferential recruitment of CPSF to distal PASs (Zhu et al. 2018; Li et al. 2015). Importantly, CFI-25 expression is downregulated in glioblastoma cells, leading to the usage of upstream PASs which enhances tumorigenicity and increases tumor size (Masamha et al. 2014); conversely, CFI-25 overexpression inhibits tumor growth. In addition, copy number variations of NUDT21 (the gene encoding CFI-25) were found in individuals with certain neuropsychiatric syndromes (Gennarino et al.). In lymphoblastoid cells of these individuals, increased CFI-25 levels led to higher expression of a long isoform of the mRNA encoding methyl CpG-binding protein 2 (MECP2), resulting in reduced production of MECP2, probably owing to the presence of numerous miRNA target sites in its alternative UTR segment between proximal and distal poly(A) sites. Because MECP2 levels need to be tightly regulated in the brain and small fluctuations in abundance can lead to neurological malfunctions, NUDT21 was suggested to be a candidate gene for causing intellectual disability and neuropsychiatric diseases. As such, APA can be significantly regulated by modulating expression of core cleavage factors, and their appropriate expression is essential for normal cell activity. A growing number of other APA regulatory trans factors were identified using a siRNA knockdown system. Some of the factors function globally, whereas others act in a gene-specific manner. For example, changes in the expression of core polyadenylation factors (e.g. CSTF64, a subunit of the CSTF complex) or other proteins (PABPN1, and PABP) can make a global shift in PAS usage by either

promoting proximal or distal PAS usage for many genes. For example, increased CSTF64 expression plays a role in 3'UTR shortening as well as intronic polyadenylation during B cell maturation, and interestingly is often found in cancers showing global 3'UTR shortening. Of note, certain transcripts are more sensitive to regulation by CSTF64 expression level, w more U- and GU-rich elements, the recognition motifs of CSTF64, near PASs. In addition, various aspects of mRNA regulation, including transcription, splicing, polyadenylation, and cytoplasmic mRNA regulation, were reported to be interconnected with APA regulation, including promoter activities (Nagaike et al. 2011), transcription elongation rate (Liu et al. 2017), splicing factors (U1 snRNP), poly(A) tail binding proteins (PABPN1, and PABP1), and other RBPs (Ren et al. 2020; Tian and Manley 2013). Thus, multiple layers of transcriptional and post-transcriptional regulation diversify and fine-tune APA regulation depending on the cellular and molecular contexts.

In addition to APA, which occurs in the same exon, another important mode of generating alternative 3'UTR isoforms is through alternative 3' terminal exons (also called alternative last exons, ALEs). In contrast with tandem polyadenylation sites, which are located within the 3'UTR, ALE affects coding regions through alternative splicing and subsequent termination in intronic PAS, giving rise to protein isoforms differing in protein coding regions as well as sharing no common sequences in the subsequent 3'UTR regions. In general, the distal last exon of a gene is viewed as producing the full-length isoform, while internal ALEs (also called proximal ALEs) lead to less abundant and truncated isoforms, sometimes with dominant-negative properties. Intronic PAS sites are much less conserved between human and murine compared to downstream PAS sites. Although more than 3000 human genes have alternative 3' terminal exons and there are broad differences in the ALE usage across the tissue (Tian et al. 2007; Kalsotra and Cooper 2011), our understanding of how

dynamically ALE can be regulated, what their functional significance is, and how these processes are regulated remain largely unknown.

Growing evidence suggests that both types of alternative 3'UTR isoforms, ALE and tandem 3'UTRs, exhibit distinct regulatory impacts on mRNA localization, stability and/or translation between distal and proximal isoforms (Mayr, 2016; Tian and Manley, 2017; Zheng et al., 2018). 3'UTR-mediated gene regulation often involves poly(A) tail length control with tail lengthening leading to increase in translation and mRNA stability and vice versa. Interestingly, genome-wide studies on poly(A) tail length and APA revealed that distal tandem 3'UTR isoforms preferentially exhibit longer mean poly(A) tail length (Legnini et al., 2019), which suggests a global role of APA in poly(A) tail length regulation. However, it is unclear whether this relationship applies equivalently to ALE choice and poly(A) tail length.

While APA can be analyzed with data from microarrays (Sandberg et al. 2008) and serial analysis of gene expression (SAGE) (Ji et al. 2009) or RNA-seq (Katz et al. 2010; Xia et al. 2014), these techniques were not specifically designed to identify PAS sites, leading to incomplete analysis. These methods are particularly ineffective when PAS sites of different isoforms are located close to one another. However, isoforms using different PAS sites within a short window (~100 nt) have been shown to have quite different mRNA decay rates (Geisberg et al. 2014), making it necessary to examine APA isoforms with precise tools. A number of deep sequencing methods have been developed to specifically sequence the 3' end of transcripts (Fu et al. 2011). These methods can identify PAS sites and also examine gene expression. Most methods use primers containing the oligo(dT) sequence for reverse transcription (RT). While efficient, oligo(dT) can prime at internal A-rich sequences (Nam et al. 2002), leading to false PAS identification. This issue is usually addressed computationally by eliminating putative PAS in A-rich regions (Lee et al. 2007). However, this approach

cannot guarantee full elimination of false positives caused by internal priming, and can also discard bona fide PAS sites in a stretch of As. To overcome this issue, some sequencing methods, including 3P-seq [poly(A)-position profiling by sequencing] (Jan et al. 2011) and 3'READS variants (3' region extraction and deep sequencing; Hoque et al. 2013), were developed more recently. Both approaches involve removal of most of the poly(A) tail sequence by RNase H followed by ligation of an adapter to the 3' end of digested RNA. A short poly(A) sequence unalignable to the genome is used as evidence for the poly(A) tail, which is important for identification of genuine PAS sites.

In our study, we investigated the global landscape of alternative 3'UTR isoforms in human macrophage cells, and its dynamic change during the early stage of macrophage activation using 3'-seq, a variant of 3p-seq. In conjunction, we characterized the relationship between alternative 3'UTR isoforms and poly(A) tail lengths by combining TED-seq and 3'-seq data generated along the same time-course during LPS stimulation of human macrophage THP-1 cells.

Results

Global identification of alternative 3'UTR isoforms used in human macrophage cells

To generate a global and high-resolution view of poly(A) tail length and 3'UTR isoform identity during macrophage activation, we stimulated the human macrophage cell line THP-1, with LPS, and collected RNA samples 0,1,2, and 4 hours after LPS stimulation. These temporal RNA samples were used for TED-seq and 3'-seq library preparation to measure poly(A) tail lengths and identify 3' mRNA ends, respectively (Figure 1). In our initial analysis, we focused on unstimulated cells, categorizing genes into those that generate tandem 3'UTR pairs and those that generate ALE pairs (Figure

2A). Transcripts from 5,438 genes were detected in both TED-seq and 3'-seq under resting conditions, with the majority of genes containing a single terminal exon (n=4,809) and smaller number of genes exhibiting multiple ALE isoforms (n=629). Moreover, out of all 3' terminal exons (n=6,871 transcript isoforms, from 5,438 genes), most express a single 3'UTR isoform (n=5,986), while a minority were found to express multiple 3'UTR isoforms by APA (n=895) (Figure 2B and 2C). Thus, alternative 3'UTRs generated by both alternative splicing and APA are prevalent in unstimulated macrophages, affecting 11.6% and 13% of the transcripts, respectively.

Measuring the poly(A) tail lengths of alternative 3'UTR isoforms

Alternative 3'UTR isoforms, ALE and tandem 3'UTRs, exhibit distinct regulatory impacts on mRNA localization, stability and/or translation between distal and proximal isoforms. Therefore, we examined the association between 3'UTR isoform identity and poly(A) tail length. To do this, we used our TED-seq and 3'-seq data generated in human macrophage THP-1 cells responding to LPS. By focusing on unstimulated cells, we investigated the relationship separately for the genes that generate tandem 3'UTR pairs and those that generate ALE pairs. We found that distal tandem 3'UTR isoforms, in general, tend to possess longer poly(A) tails compared to the proximal isoform set (Figure 2D). However, this association was not observed when we compared poly(A) tail lengths between sets of distal and proximal ALE isoforms (Figure 2E). Importantly, when we compared pairs of isoforms generated from the same gene, we again found that poly(A) tails from distal APA isoforms were longer than their shorter proximal counterparts (Figure 2F and S1E). Moreover, differences in poly(A) tail lengths were weakly but significantly correlated ($R = 0.17$, $P = 9.9 \times 10^{-7}$) with differences in alternative 3'UTR lengths between the isoform pairs, with isoform pairs with larger differences in lengths tending to have greater

differences in poly(A) tail lengths (Figure 2H). However, this relationship was not observed for pairs of isoforms generated by ALEs (Figure 2G, Figure S1F). Given that there is no 3'UTR length bias for proximal vs. distal ALE isoforms (Figure 2G left), unlike for APA isoform pairs (Figure 2F left), the preference of distal APA isoforms for longer poly(A) tails may be linked to 3'UTR length and the differential abundance of regulatory elements influencing degradation. Alternatively, these isoforms may be adenylated at different rates. In sum, these results emphasize the importance of considering alternative 3'UTRs generated by APA and splicing as distinct groups. We also discovered trends common to both ALE and tandem 3'UTR isoform pairs. In unstimulated THP-1 macrophage cells, distal ALE and distal tandem 3'UTR isoforms were more abundant, as indicated by the color of the dots (more orange) (Figure 2D, Figure 2E), a result more evident in the pairwise analyses (Figure 2F-2G, Figure S1EF-S1FG). Moreover, more transcript isoforms with poly(A) tails greater than 200 nt were found in the sets of proximal ALE and APA isoforms (Figure 2I left, Figure 2J left). One characteristic of these long-tailed transcripts is their reduced abundance compared to transcripts with shorter poly(A) tails (Figure 2I right, Figure 2J right). These trends were also observed for genes with three tandem or ALE isoforms (Figure S1G and Figure S1H). It is worth noting that these results are in line with observations in steady-state somatic cells, in which transcripts with long median poly(A) tails and with a broad distribution of tail lengths were found to exhibit relatively rapid deadenylation and decay (Lima et al., 2017). Taken together, our data revealed that proximal tandem 3'UTRs exist typically with shorter poly(A) tails concomitant with reduced transcript abundance, and revealed a close association between 3'UTR length and poly(A) tail length control, although it remains to be determined whether these associations are directly caused by sequence-

mediated regulation, physical length effects of the alternative 3'UTR, or perhaps a more exotic mechanism.

Changes in alternative 3'UTR isoform usage during macrophage activation

The relative usage of 3'UTR isoforms is often modulated during cellular processes. For example, both cellular transformation and proliferation are correlated with preferential generation of proximal ALEs and proximal APA isoforms, whereas differentiation, senescence and quiescence are associated with shifts toward distal ALEs and distal APA isoforms (Fu et al. 2011; Ji et al. 2009; Mayr and Bartel 2009; Sandberg et al. 2008; Taliaferro et al. 2016; Chen et al. 2018). Global shifts towards shorter 3'UTR isoforms have been reported in macrophages as they respond to infection, conclusions reached using human primary monocyte-derived macrophages (MDM) treated with Macrophage Colony Stimulating Factor (M-CSF) (Jia et al., 2017; Pai et al., 2016). Our macrophage activation model, using human THP-1 cells, is accomplished in two steps: first, differentiation using PMA (phorbol 12-myristate-13-acetate; Starr et al., 2018), and second, activation using LPS. Accordingly, we investigated whether PMA-mediated differentiation and LPS activation also manifested similar alterations in 3'UTR isoform preferences.

We focused on 3'UTR isoforms that passed the criteria of PAS isoform defined from our 3'-seq pipeline, and which had ≥ 50 read counts in at least one TED-seq dataset across the induction time-course. PAS isoforms within 100 nt were considered to be from the same or similar transcripts, and thus collapsed into a single cluster (tandem PAS cluster, Figure 3A). We found that during LPS activation, 31% of all 3' terminal exons (3,333/10,717 terminal exons) possessed multiple tandem clusters and that 14 % (1,283/9,226 genes) expressed multiple ALE isoforms at one or more time-points (Figure 3B). We defined relative APA isoform abundance as the fraction of 3'-seq

read counts from a given tandem cluster over counts from all clusters within the corresponding terminal exon. Relative abundance of proximal versus distal APA isoforms revealed that most genes with APA isoforms preferentially express the most distal PASs (resulting in the longest 3'UTRs) in all time points before and after LPS activation (Figure 3C, right). Likewise, we defined relative ALE usage as the fraction of 3'-seq reads from a given terminal exon over counts from all ALEs from the corresponding gene. Distal ALE isoforms were more abundant than proximal ALE isoforms across the macrophage activation time-course (Figure 3C left). Given the systematic preference for distal 3'UTRs in many differentiated cells, this bias towards distal 3'UTR isoforms, observed for both ALE and APA isoforms, is unsurprising considering THP-1 cells are differentiated cells, although the degree of distal isoform usages in the undifferentiated cells was not determined in this study. In support of this, a previous study showed that PMA-induced THP-1 cells are less proliferative and exist in a more pronounced differentiation state than those induced with M-CSF. Next, we assessed the change in relative abundance of APA isoforms located within a 3' terminal exon across our stimulation time-course (Figure 3A). The degree of differential isoform usage between any two time points was assessed using the 3'UTR switch index (USI; Fu et al., 2011; Jia et al., 2017), which indicates a shift toward distal APA isoform for positive values (USI >0.1) and proximal APA isoform for negative values (USI < -0.1) (Figure 3D for an example of APA switch). Strikingly, and in stark contrast to the M-CSF-induced MDM cell model which reported a switch to shorter 3'UTR isoforms upon infection (Jia et al., 2017; Pai et al., 2016), our data showed that LPS stimulation induced more 3'UTR lengthening during the early stages of macrophage activation (1h), followed by a gradual increase in the number of genes exhibiting 3'UTR shortening (Figure D and Figure 3E left). However, the number of genes with 3'UTR lengthening (n=566, FDR < 0.1) surpasses those with 3'UTR

shortening (n=464, FDR < 0.1). This trend, observed using the averaged values of two biological replicates of 3'-seq data, was consistently reproduced with an individual set of 3'-seq replicates despite the relatively low correlation between the biological replicates at zero time point (Figure S2A and S2B). These results indicate phenotypic heterogeneity between M-CSF-induced MDM cells and PMA-differentiated THP-1 cells. We also examined the changes in ALE usage, finding relatively fewer genes with ALE switch (n=32, Figure S2C and S2D; Figure S2E for an example of APA switch). Thus, macrophage activation involves extensive changes in 3'UTR isoform usage, but the majority (97%) of these changes are mediated by APA, rather than alternative splicing. It should be also noted that we did not observe a consistent relationship between the mRNA abundance changes and changes in APA isoform abundance (Figure S2F), indicating this relationship cannot be simplified, which agrees with previous studies (Jia et al. 2017).

Notably, genes that exhibited changes in 3'UTR isoform usage during macrophage activation were significantly enriched for gene ontologies involved in immune responses, metabolic processes and protein transport/localization. In particular, immune-related terms were highly enriched for the set of genes exhibiting 3'UTR lengthening, whereas protein transport/localization ontologies were preferentially enriched for the set displaying 3'UTR shortening (Figure 3E right). These enrichments are consistent with changes in 3'UTR isoform usage over the time course, contributing to the phenotypic alterations of macrophages during immune activation.

DISCUSSION

In this study, we set out to systematically examine the prevalence and dynamics of alternative 3'UTR isoform expression in human macrophage cells, and its relationship with poly(A) tail lengths and RNA expression, in a complex regulatory environment.

We selected a cell culture model of macrophage activation, which we examined across a time-course of stimulation. To accomplish our goals, we simultaneously measured 3'UTR isoform abundance and poly(A) tail length in unstimulated macrophage cells and during a time-course following LPS stimulation. Importantly, our approach enabled us to profile the poly(A) tail across the transcriptome and with 3'UTR isoform resolution, which is an important aspect of the study given that alternative 3'UTR isoforms typically undergo different post-transcriptional regulation. Using these approaches, we were able to reveal extensive regulation of transcript isoform abundance associated with the control of poly(A) tails.

Association between alternative 3'UTR isoform choice and poly(A) tail lengths

The regulatory elements within the 3'UTR are the major specifiers of post-transcriptional events that impact poly(A) tail lengths (Eisen et al., 2020; Legnini et al., 2019; Pai et al., 2016). In addition, 3'UTRs themselves are not static, with multiple alternative isoforms existing for each gene; more importantly, isoform preferences change across cell types and during cellular transitions (Lianoglou et al. 2013; Mayr and Bartel 2009; Pai et al. 2016; Sandberg et al. 2008; Taliaferro et al. 2016). Thus, a focus of this study was to examine 3'UTR isoform preferences across the macrophage activation time-course, and to investigate the relationship between alternative 3'UTR choice and poly(A) tail length. Consistent with previous reports, we demonstrated that distal APA isoforms exist preferentially with longer poly(A) tails (Legnini et al., 2019). Importantly, steady-state poly(A) tail lengths are determined primarily by the balance between co-transcriptional input of nascent polyadenylated transcripts and post-transcriptional processing leading to deadenylation. Thus, it has been ambiguous whether differences in poly(A) tail lengths between proximal and distal 3'UTR isoforms derive from alternative post-transcriptional fates, or alterations associated

with new transcripts entering the pool. Our results in steady-state cells indicate that poly(A) tail length differences between proximal and distal APA isoforms increase as the differences in 3'UTR isoform lengths increase, supporting the association between poly(A) tail length control and the length of 3'UTR, a major post-transcriptional specifier. However, it remains unclear which aspect of longer 3'UTRs are functionally relevant: the physical length of the 3'UTR itself, or the presence of additional cis-acting elements within the extended 3'UTR. For example, some studies indicated that mRNA deadenylation and decay machineries assembled close to the stop codon are less efficient at promoting the decay of transcripts with long 3'UTRs, perhaps as a consequence of increased distance to the poly(A) tail (Mishima and Tomari 2016). Alternatively, it is also likely that cis-acting elements found within longer alternative 3'UTRs may facilitate deadenylation and decay of distal 3'UTR isoforms. Paradoxically, accelerated deadenylation-dependent decay can result in a transcript pool with longer poly(A) tails, a consequence of the increased relative fraction of newly synthesized long-tailed transcripts compared to the rapid loss of older transcripts (Lima et al. 2017). Taken together, in concert with other previous studies (Legnini et al. 2019; Lima et al. 2017; Mishima and Tomari 2016; Woo et al. 2018), our results provide evidence showing the importance of APA choice and/or 3'UTR length in poly(A) tail length control in steady state cells, but the underlying mechanisms remain to be further explored.

Dynamics in alternative 3'UTR isoform expression during macrophage activation

Notably, our data on 3'UTR isoform usage, based on PMA-differentiated THP-1 cells stimulated with LPS, differ from those obtained using differentiated MDM cells induced by M-CSF, an alternative model of macrophage activation. We suspect that the physiological differences between the two human macrophage models are related

to the differences in post-transcriptional control. In particular, our data indicates clearly that activation triggers a shift to longer 3'UTR isoforms, whereas the MDM-based model appears to show the opposite trend (Alasoo et al. 2015; Pai et al. 2016). In support of our observations, PMA-differentiated THP-1 macrophages activated by exposure to *Mycobacterium tuberculosis* also induce 3'UTR lengthening (Kalam et al. 2017), indicating phenotypic differences between the two alternative macrophage models. Moreover, M-CSF induced MDM cells are more resistant to bacterial infection than PMA-induced cells (Starr et al. 2018), suggesting that the global shift to shorter 3'UTR isoforms in M-CSF induced MDM cells may be functionally relevant to their cellular state. Previous studies have shown that differentiation and/or senescence accompany 3'UTR lengthening, whereas cancer transformation and increased proliferative capacity are associated with 3'UTR shortening (Chen et al. 2018; Hoque et al. 2013; Mayr 2016; Taliaferro et al. 2016). Moreover, de-differentiation is coincident with a shift towards proximal PAS, resulting in shorter 3'UTR isoforms, implicating a close coupling between 3'UTR isoform usage and differentiation (and/or proliferation activity). Although the basis for the different trend between the two macrophage models is unclear, it may be associated with variations in the potential to de-differentiate and proliferate between the two macrophage models. That is, perhaps there is a global preference for an increase in proximal APA isoforms in LPS stimulated macrophages, as observed in multiple other non-developmental cellular activation systems (Zheng et al. 2018). Therefore, M-CSF-induced MDM cells may retain potential to de-differentiate, exhibiting a global shift toward proximal APA isoforms. In contrast, differentiation elicited by PMA may specify a more pronounced differentiation state, and thus less potential to de-differentiate, favoring the use of distal APA isoforms in response to LPS stimulation.

However, dynamic changes in 3'UTR isoform abundance in the cells should be carefully understood since it is unclear whether the altered 3'UTR isoform abundance is a consequence of a shift in poly(A) site usage during nuclear 3' end processing or a change in mRNA stability of the 3'UTR isoforms. For example, in the case microRNA sites are present only in the longer isoform, increased microRNA expression could cause rapid degradation of the longer isoform upon activation, leading to an apparent shortening of 3'UTR isoforms without affecting APA. Second, given that polyadenylation is regulated by complex mechanisms that involve various trans-acting factors along with transcription dynamics, it will be important to determine the molecular mechanisms that underlie environmental- or tissue-specific 3' end formation. Lastly, it will be important to understand if and how differential 3'UTR usage itself affects the functional properties of macrophage activation and other diverse cellular contexts.

Methods

Cell lines, Cell culture and Compound Treatment

THP-1 cells (ATCC, TIB-202) were cultured in RPMI1640 (Gibco, 11875093) supplemented with 10% FBS (VWR) and 1% antibiotics (Gibco, 15240062). THP-1 cells were differentiated to macrophage-like cells by incubating them overnight in complete media containing 200 ng/ml PMA- (Sigma Aldrich, P1585-1MG), followed by 3 days incubation in fresh media without PMA. The resulting differentiated cells were stimulated with 200 ng/mL LPS (Sigma-Aldrich, A 9415) and collected at four time points: 0 hours post-stimulation (no stimulation), and 1, 2, and 4 hours post-stimulation. Total RNA was extracted with TRIzol (Invitrogen, 15596018). To inhibit transcription, THP-1 cells were incubated in media with 1ug/mL Actinomycin D (Sigma Aldrich, A9415) for 15 minutes prior to stimulation with LPS.

3'-sequencing library preparation

Total RNA was extracted using TRIzol (Invitrogen, 15596018) from differentiated THP-1 cells throughout the LPS stimulation time-course (0, 1, 2 and 4 hours). For each sample, poly(A) RNA was isolated from 10 ug of total RNA (Dynabeads™ mRNA Purification Kit; Invitrogen, 61006) followed by RNA fragmentation with 0.1 N NaOH, 5' RNA phosphorylation (NEB, M0201S), and 5' RNA ligation (NEB, M0204L) to VRA5 (5'-CCUUGGCACCCGAGAAUCCA-3'). After heat denaturation at 65oC for 2 minutes, 5' adapter-ligated poly(A)-containing RNA fragments were reverse transcribed by superscript II enzyme (Invitrogen,18064-014) using RT primer (CPS_RTP: 5'-GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNNT8VN-3') at 50oC for 1 hour. The 3'-terminal ten nucleotides of CPS_RTP were designed to anneal to the junction between the poly(A) tail and the site of cleavage and polyadenylation within the transcript, and also to contain an eight-nucleotide (nt) unique molecular index (UMI) barcode for PCR deduplication, with the remaining sequence designed for PCR amplification (NEB, M0530L). The resulting cDNA molecules were amplified by PCR for 14 cycles with RP-1 primer (See Table S7) and RPI-X primers (See Table S7) using Phusion High-Fidelity DNA Polymerase (NEB, M0530), followed by gel purification of 200 to 500 bp products on a 6% PAGE gel in TBE buffer. PCR products were eluted from the excised gel in TE-TW buffer overnight at 37oC, and then filtered through a DNase-free spin X column (Corning Costar Spin-X centrifuge tube filters; Corning, CLS8160) and purified using Ampure XP beads (Beckman Coulter, A63881). The purified, barcoded libraries were quantified and pooled prior to Illumina small RNA sequencing on a Next500 platform (75 bp single-end reads).

Unless otherwise stated, enzymatic reactions were performed as described in the manufacturer's protocols.

3'-seq data processing

5' RNA adaptor sequence was removed from the 3' end of sequencing reads using Cutadapt with option `-e 0.10, --overlap 2, --minimum-length=10, --nextseq-trim 20`. After adaptor removal, low quality reads were removed (those with quality scores at any position <20). The first 30 nt, containing the 8 nt UMI, were used to deduplicate the reads. After trimming 16 nucleotides (8 nt UMI and 8 nt corresponding to the dT8 portion of the adapter oligonucleotide) from the 5' end of each read, reads with at least 10 nt remaining were mapped to the human genome (hg38) released by UCSC, using STAR aligner (Dobin et al., 2013) with the option `--sjdbGTFfile "gencode.v26.annotation.gtf" --alignSJDBoverhangmin 3, --outFilterMultimapNmax 1`. The aligned reads were represented by their 5' end mapping coordinate on the opposite strand, and converted to BedGraph format, where the mapping position and the corresponding read counts of a 3'-seq peak were used to determine the cleavage and polyadenylation site and mRNA abundance of a transcript isoform.

CPS position analysis

To avoid potential contamination of 3'-seq reads by the annealing of oligo-dT primers to internal A-rich sequences, we filtered the 3'-seq reads to remove internally primed reads from A-rich internal regions, as previously described (Li et al., 2012, Fu et al. 2011). Briefly, we searched for consecutive A sequences (>5 consecutive A nt) downstream of 3'-seq peaks, filtering out these reads from our 3'-seq reads. Then the 3'-seq read counts are normalized by counts per million mapped reads (CPM). The 3'-seq peaks were collapsed across all samples (0, 1, 2, and 4 h) with the read count per

position totaled. Next, each 3'-seq peak position was converted to a 10 nucleotide-wide window, and the overlapping windows of 3'-seq peaks within the window merged, totaling the merged 3'-seq peaks, retaining the midpoint of the merged window as the PAS coordinate. Merged windows with five or more normalized reads (final PAS window) were retained. In each 3'-seq timepoint data, the read counts of the 3'-seq peaks mapped in a final PAS window were summed to represent mRNA abundance of the PAS isoform expressed in the given sample. All final PAS located in the reference (GENCODE V26)-annotated 3'UTR(s) + 1kb downstream region, were considered as distinct 3'UTR isoforms expressed in THP-1 cells. Finally, the custom transcript isoform annotation (bed12) file was built by modifying the reference transcript isoforms to terminate at our experimentally determined 3'-seq PAS sites. Poly(A) tail lengths were estimated for this comprehensive set of experimentally determined PAS isoforms.

Analysis of ALE isoforms and tandem 3'UTR isoforms in unstimulated cells

By analyzing our custom transcript isoform annotation, each gene was scrutinized for ALE and tandem 3'UTR expression. If a gene had multiple transcript isoforms whose 3' terminal exons start at different genomic locations, it was considered to express alternative last exon isoforms (ALEs). If a gene had multiple transcript isoforms that shared the same 3' terminal exon (5' splice site, thereof), but were cleaved and polyadenylated at different positions, these transcript isoforms were considered as tandem 3'UTR isoforms. For analytic stringency, we only considered transcript isoforms with ≥ 50 TED-seq reads within the 300 nt region upstream of the PAS for further analysis. This approach allowed us to remove any potential decay intermediates and other artifacts derived from internal priming from our collection of PAS isoforms, and focus on a more confident set of transcript isoforms.

In cases with multiple transcript isoforms located within ≤ 300 nt, we determined the isoform with the highest read counts as the major isoform, and the rest as minor isoforms. Poly(A) tail length was only calculated for the major isoforms calculation, to ensure that TED-seq reads were applied to only a single isoform. Finally, for the final set of major transcript isoforms, each transcript isoform was indexed based on their genomic location, as follows. For a gene manifesting ALEs, the proximal isoform (closest to the 5' prior exon), was indexed as 1, and those further away from the prior exon were indexed as 2, 3, etc, increasing the index with distance from the prior exon. In the case of a 3' terminal exon expressing multiple PAS isoforms, the PAS closest to the stop codon was indexed as 1, incrementing the index with increasing distance from the stop codon. Poly(A) tail length, RNA abundance and 3'UTR length were measured as follows: for a tandem 3'UTR isoform, mRNA abundance, poly(A) tail length and 3'UTR were calculated as the read counts of the given isoform, mean poly(A) tail length, and the distance from stop codon to the PAS, respectively. For a 3' terminal exon, mRNA abundance of the 3' terminal exon was defined as the sum of all tandem 3'UTR isoforms positioned in the given 3' terminal exon. Poly(A) tail length of a given 3' terminal exon was defined as the average of mean poly(A) tail lengths of tandem 3'UTRs weighted by the read counts of tandem 3'UTR isoforms (weighted poly(A) tail length). Likewise, 3'UTR length was weighted by the read counts of the tandem 3'UTR isoforms to represent 3'UTR length per 3' terminal exon with tandem 3'UTR isoforms. For a 3' terminal exon possessing tandem 3'UTRs, relative RNA abundance of a tandem 3'UTR isoform was calculated by dividing its RNA level by the RNA level of the 3' terminal exon. Likewise, relative RNA abundance of a 3' terminal exon was defined as RNA level of the exon relative to overall expression of the gene, the sum of all ALE isoforms derived from the gene.

Assessment of 3'UTR isoform switching

Switching in either tandem poly(A) sites or in ALEs was tested based on previously described approaches (Jia et al., 2017, Fu et al., 2011, Agresti et al., 2002). Briefly, for cases in which a terminal exon contained multiple poly(A) sites (or multiple tandem 3'UTRs), PAS isoforms with more than 50 TED-seq reads were considered as the expressed mRNA isoform, and PAS close to each other (≤ 100 nt) were clustered together to represent a single APA isoform (tandem PAS cluster). For genes with multiple tandem PAS clusters in the 3' terminal exon, the shift in tandem 3'UTR isoforms between two time points following LPS stimulation was assessed using the following statistical parameters given the relative abundance of individual tandem 3'UTR isoforms time-course. First, polychoric correlation coefficients, a version of Pearson correlation coefficient with discrete UTR indices, were calculated to estimate the switching direction, and defined as USI, UTR switching index (described as TSI in Jia et al., 2017), and second, a Chi-squared test was performed to determine the statistical significance of the 3'UTR switch. Under the criteria of the adjusted p-value (FDR) < 0.1 , a positive USI value (USI > 0.1) indicated a switch to the longer tandem 3'UTRs (distal), while a shift to short tandem 3'UTR (proximal) has a negative USI value (USI < -0.1).

TED-seq library preparation

Tail End Displacement sequencing (TED-seq) was applied to total RNA samples (5 - 10 ug) collected at multiple time points (0, 1, 2 and 4 h) after LPS treatment. TRIzol-Purified RNA was subjected to poly(A) RNA purification using the manufacturer's protocol (Dynabeads™ mRNA Purification Kit, Invitrogen), and ligated with the adaptor molecule (RA3; Table S7) to their 3' terminus. The products of the ligation reaction were purified using TRIzol, and fragmented with 0.1 N NaOH. Fragmented

RNAs were purified with a P-30 column (Bio-Rad, 732-6251), and poly(A)-containing fragments enriched using Dynabeads mRNA purification kit. T4 polynucleotide kinase (PNK; NEB, M0201S) was used to phosphorylate the 5' terminus of RNA fragments, enabling ligation of the 5' terminus to the adaptor oligonucleotide containing UMIs (RA5; Table S7). The resulting RNA libraries were reverse transcribed and PCR amplified using KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems, KR0370), using no more than 8 cycles of amplification. Prior to sequencing, 350-360 bp DNA molecules were purified using PAGE. Following PCR amplification, PAGE-mediated size selection was repeated on the amplified DNA. The resulting size-selected libraries were pooled, and sequenced on an Illumina NextSeq500 (75 bp single-end reads).

Synthesis of spike-in poly-(A) standards

Poly(A) spike-in RNAs of 40, 80, 120 and 160 nt were generated by in vitro transcription of PCR amplified double stranded DNA template composed of T7 promoter sequence, unique sequences for alignment from plasmid vector backbones (mRFP for A40, pEGFP-C1 at EGFP ORF for A80, pEGFP-C1 at NeoR/KanR ORF for A120, and pGL4.23 for A160), and poly(A) repeats of desired lengths (Table S7). In vitro transcribed RNA products (MAXIscript™ T7 Transcription Kit; Invitrogen, AM1314) were PAGE-purified and quantified using Nanodrop. The spike-in RNAs were added to purified mRNAs from samples used for TED-seq library generation (1 ng of each spike-in RNA species per 100 ng of poly(A)-selected RNA).

TED-seq data processing

For sequencing reads ending with >10 A residues, consecutive (A) sequences were trimmed from the 3' end. After poly-(A) tail trimming, reads with a length ≥ 15 nt and

mean quality score > 20 were retained for further analysis. PCR duplicates were filtered out using the first 15 nt of the trimmed reads, which includes an 8 nt UMI. Nucleotides correspond to the UMI were then trimmed from the 5' end of the deduplicated reads, followed by removal of trimmed reads shorter than 15 nt. The resulting reads were mapped to the human genome (hg38) using STAR (2.4.2a; ref) with the option `--sjdbGTFfile "gencode.v26.annotation.gtf"` `--alignSJDBoverhangmin 3`, `--outFilterMultimapNmax 1`. BWA (Li et al., 2009) was used to align the reads corresponding to the poly(A) spike-in standards. The reads aligned to the genome were represented by their 5' terminus mapping coordinate, and visualized with the IGV genome browser (Robinson et al., 2011). GENCODE V26 annotation (Frankish et al., 2019; Harrow et al., 2012) of the human transcriptome (bed12 format) was amended to reflect experimentally determined THP-1 cleavage and polyadenylation sites identified using 3'-seq. Finally, a frequency table of TED-seq read 5' termini located within the 3' terminal 500 nt of 3' exons within the custom transcriptome annotation was constructed.

Poly(A) tail length estimation

Transcript isoforms with ≥ 50 mapped reads in the terminal 500 nt region were used to calculate the mean value of poly(A) lengths and to represent the distribution of poly(A) tail reads for that region. To avoid any errors in poly(A) tail inferences caused by a shift in APA usage, we removed PAS isoforms subject to potential bias from APA by processing our 3'-seq data as follows. First, we defined tandem PAS clusters of size = 300 nt by clustering experimentally determined PAS isoforms within ≤ 300 nt, and testing whether APA usage was significantly different for all isoforms within each cluster between any two time points of the macrophage LPS activation time-course (Chi-squared test with FDR <0.1). We only considered PAS clusters with

consistent APA isoform usage across the time-course for further analysis. Next, if a PAS cluster contained multiple PAS, the PAS with the most read counts was defined as the major PAS isoform, whereas PAS isoforms with fewer read counts were considered as minor PAS isoforms (n=2,979). To avoid redundant use of a collection of TED-seq reads for the PAS in the same cluster, minor PAS isoforms were removed from the tail length analysis, leaving 6,269 isoforms (5,079 genes). For this set of transcript isoforms, differences in poly(A) tail length for a given transcript isoform between two biological conditions were compared using the Kolmogorov–Smirnov test and the p-value adjusted by FDR with the criteria of $FDR < 0.01$, and $|\Delta PAL|$ (difference in mean poly(A) tail length) ≥ 10 nucleotides.

Quantification and Statistical Analysis

Statistical Methods

Data presented as mean \pm SD or mean \pm 95% CI (as indicated). Statistical significance was calculated with one of Student's t-test, Kolmogorov–Smirnov test, Mann-Whitney test (Wilcoxon signed-rank test) with the significance denoted as follows, $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), unless noted otherwise. The type of statistical test, and sample numbers, and statistical significance (P value or FDR) are described in the figures or legends. All graphs and statistical tests were performed using R.

Data and Code Availability

Raw and processed data of TED-seq, PRO-seq, and 3'-seq are available at the GEO accession number GSE161188.

Figures

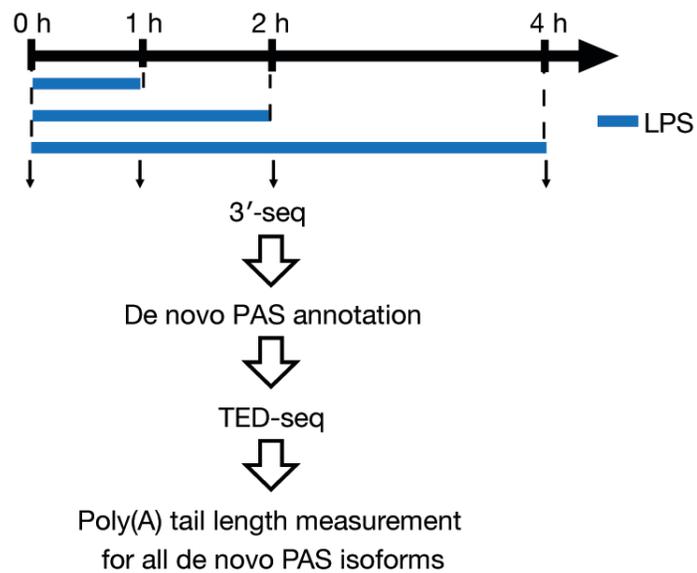


Figure 1. Genome-wide, transcript isoform-specific determination of poly(A) tail lengths. Schematic of macrophage activation time-course, and the sequencing strategy. Blue bars indicate the duration of LPS treatments, and the labeled time-points indicate when the RNAs are harvested. PAS: 3' cleavage and polyadenylation site.

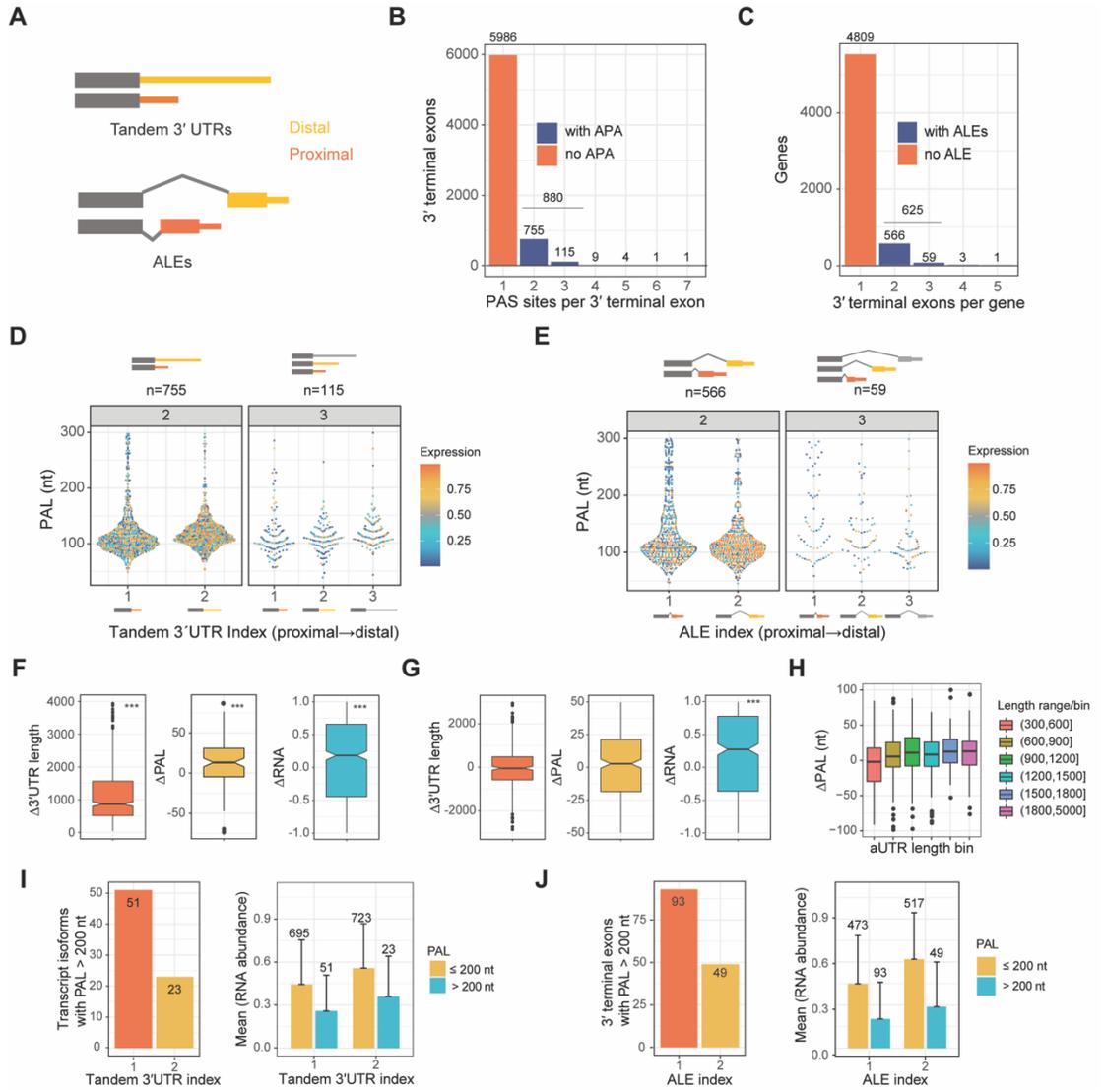


Figure 2. Characterization of poly(A) tail lengths of alternative 3'UTR isoforms in unstimulated THP-1 cells.

(A) Schematics of different mechanisms that generate alternative 3'UTR. ALE: alternative last exon.

(B) Histogram of the number of APA isoforms (tandem 3' UTRs) per 3' terminal exons in unstimulated THP-1 cells, analyzed for 6,881 terminal exons in 5,438 genes.

(C) Histogram of number of 3' terminal exons per gene in unstimulated THP-1 cells, in 5,438 genes.

(D) Comparison of the mean poly(A) tail lengths between sets of proximal and distal tandem 3'UTR isoforms. Terminal exons containing 2 or 3 PAS are examined separately (left and right panels, respectively). Each point represents the poly(A) tail length of a PAS isoform on the y-axis, color-coded by RNA abundance relative to the overall expression of all PAS isoforms in the corresponding terminal exon (Expression heat-map) . Difference in tail lengths for each set was assessed by a two-tailed t-test ($P < 10^{-7}$ comparing proximal and distal sets for exons with 2 PAS; $P < 10^{-4}$ comparing the most proximal (tandem 3'UTR index = 1) and the most distal (index = 3) isoforms for exons with 3 PAS).

(E) Comparison of weighted mean poly(A) tail lengths between sets of ALE isoforms, examining genes expressing 2 or 3 ALEs (left and right panels, respectively), otherwise as described in panel D ($P > 0.1$ comparing proximal and distal sets for exons with 2 ALEs; $P > 0.2$ comparing the most proximal (index = 1) and the most distal (index = 3) sets for exons with 3 ALEs).

(F) Paired analysis of tandem 3'UTR isoforms, comparing changes in 3'UTR length, poly(A) tail length and RNA abundance (left, middle and right panels, respectively) between distal and proximal APA pairs derived from a common terminal exon,

plotting the average of distal-proximal differences. 3' terminal exons with 2 PAS were analyzed. *** denotes $P < 0.001$ (two-tailed Student's t-test).

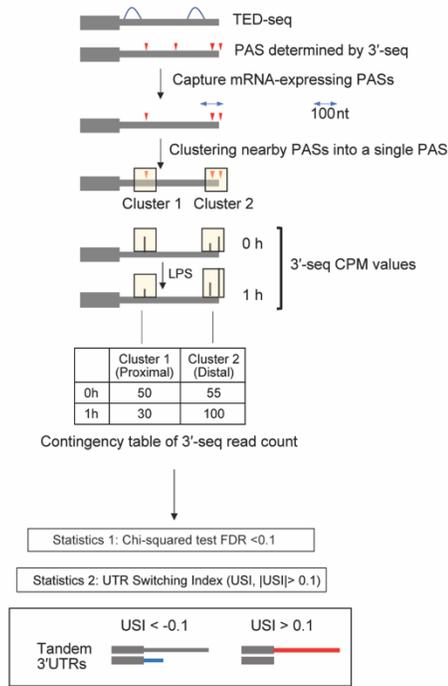
(G) Paired analysis of ALE 3'UTR isoforms, otherwise as described in panel F. Genes expressing two 3' terminal exons were considered. For each 3' terminal exon, 3' UTR length and poly(A) tail length of all PAS isoforms derived from it were weighted by RNA abundance; weighted mean of 3' UTR length and poly(A) tail length respectively were compared between distal and proximal ALE pairs.

(H) Distribution of Δ PAL (distal-proximal) with respect to alternative 3'UTR length (aUTR length). 3' terminal exons with two APA isoforms were binned into 6 groups based on their alternative 3'UTR lengths (x-axis; see color-coded key for 3'UTR length ranges). Transcript isoforms with mean tail lengths exceeding 200 nt were excluded in panels D through H, reserved for long-tail analyses.

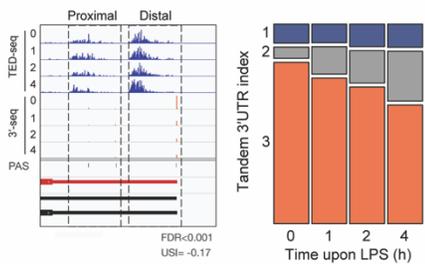
(I) (left) Number of long tailed isoforms (PAL > 200 nt) that are proximal (index = 1) or distal (index = 2) tandem isoforms in exons with 2 PAS (x-axis). (right) Average RNA abundance of long tailed (orange; mean PAL > 200 nt) and shorter tailed (blue; mean PAL \leq 200 nt) isoforms, for proximal and distal tandem isoforms (index = 1 and 2, respectively; as defined in panel D) in exons with 2 PAS. Error bars are the standard deviation of the mean, and the number of transcript isoforms are indicated above the error bars.

(J) (left) Number of long tailed isoforms (PAL > 200 nt) in each set of ALE isoforms (x-axis; as defined in panel E). (right) Average RNA abundance of long tailed and shorter tailed transcript isoforms; otherwise as described in panel I. Boxplot keys: box shows the interquartile range (IQR), center line is the median, whiskers add/subtract $1.5 \times$ IQR to the 75/25 percentile respectively, dots are individual data points outside the whiskers, and notch displays a confidence interval around the median.

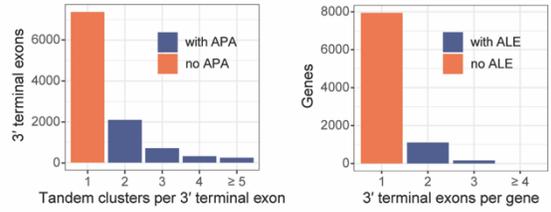
A



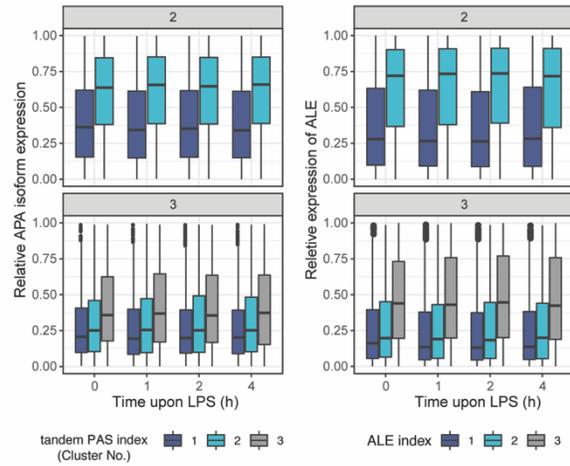
D



B



C



E

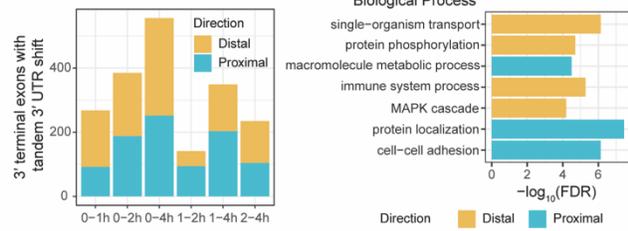


Figure 3. Alternative 3'UTR regulation upon LPS stimulation.

(A) Schematic of the pipeline used to assess changes in APA isoform usage. Tandem PAS clusters indicated in yellow boxes, considering each cluster as a distinct isoform. 3'-seq read counts per PAS were averaged from two biological replicates for each time point. Sum of the 3'-seq read counts of all PAS isoforms in a PAS cluster represents the usage of the most dominant isoform within the cluster. Polychoric correlation coefficient from the contingency table is used to indicate UTR switching index (USI); positive and negative USI values indicate downstream and upstream shifts of UTR usage, respectively.

(B) Number of tandem PAS clusters per 3' terminal exon (left) and ALEs per gene (right), analyzed for 10,717 terminal exons in 9,226 genes.

(C) Boxplots depicting relative expression of ALE isoforms and tandem PAS isoforms (left and right, respectively) for genes with 2 (top) or 3 (bottom) isoforms, analyzed over the activation time-course (x-axis). Relative expression level (y-axis) is the 3'-seq count of a PAS cluster normalized by the sum of 3'-seq counts in all isoforms of the corresponding gene. ALE isoforms are indexed (1, 2 or 3, color-coded) based on increasing 3' terminal distances from the exon upstream to the terminal exon.

Similarly, tandem isoforms are indexed by increasing distances from the stop codon.

(D) (left) Genome browser track of DAZAP1, exemplifying a switch to the proximal tandem 3'UTR isoform upon activation (FDR<0.001, USI=-0.17). TED-seq (top four tracks) and 3'-seq (next four tracks), from indicated points over the activation time-course, with the height of the orange bar indicating 3'-seq read density at each 3'UTR isoform. (right) Change in relative expression levels of the distal (index 3, orange) versus the proximal (indices 1 and 2) isoforms upon activation, as defined in panel C.

(E) (left) Counts of genes with changes in APA usage within 3' terminal exons (FDR<0.1 and |USI| < 0.1). (right) Gene ontologies enriched in genes with shifts in APA usage.

Supplemental figures

FIGURE S1

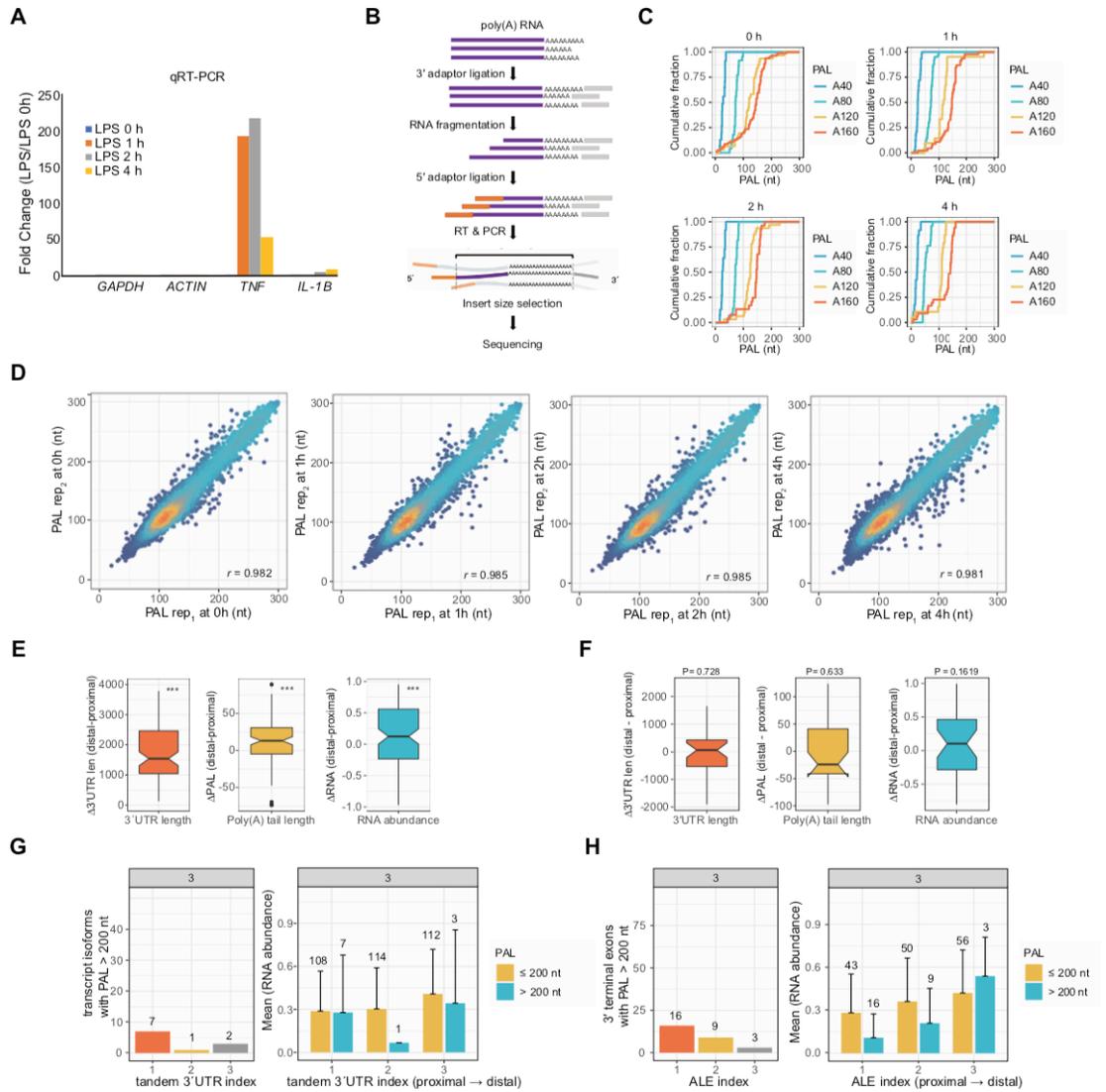


Figure S1. Reproducibility of TED-seq and 3'-seq, and Association of Poly(A) Tail Length with 3'UTR Isoform Usage.

A. Validation of macrophage activation using qRT-PCR. In each sample, the data values were normalized to GAPDH expression. For each gene, plotted is the Fold change of gene expression (Post-stimulation/unstimulated).

B. Schematic of TED-seq library preparation (modified from Woo et al., 2018).

C. The tail length distributions of 4 distinct synthetic poly(A) standards, as measured by TED-seq. A set of four RNA standards with different tail lengths (Table S2) were added to a pool of poly(A)-selected RNAs. Plotted is the cumulative distribution of poly(A) tail length for each standard analyzed by TED-seq.

D. Comparison of poly(A) tail length measured by TED-seq between two biological replicates. Each point indicates mean poly(A) tail length of an individual PAS isoform (mapped reads ≥ 50 counts), and colored based on the density on the scatter plot (orange: high density, blue: low density). Pearson correlation coefficient between biological replicates is shown. Comparison of mean poly(A) tail length between the most distal and the most proximal tandem 3'UTR isoforms from each 3' terminal exon with three PAS sites, otherwise plotted in the same way as the right panel of Figure 2E. *** indicates p-value < 0.001 using Student's t test.

E. Comparison of mean poly(A) tail length between the most distal and the most proximal ALE isoforms from each gene with three ALE isoforms, plotted in the same way as the Figure 2F right panel. No significant difference ($P > 0.1$) was observed in neither of 3'UTR length, mRNA abundance, nor poly(A) tail length.

F. Left, the number of transcript isoforms with mean poly(A) tail length (PAL) > 200 nt in each set of three distinct APA isoform classes (proximal, middle, and distal), examined for 3' terminal exons possessing 3 PAS sites. Right, for each APA isoform

class, comparison of RNA abundance between transcript isoforms with mean poly(A) tail length > 200 nt versus those with mean poly(A) tail length ≤ 200 nt.

G. Left, the number of transcript isoforms with weighted mean poly(A) tail length (PAL) > 200 nt in each set of three distinct ALE isoform classes (proximal, middle, and distal), examined for genes expressing 3 alternative last exons. Right, for each ALE isoform class, summary of RNA abundance between a set of 3' terminal exons with PAL > 200 nt versus those with PAL ≤ 200 nt. The error bar is the standard deviation of mRNA abundances for the given set of 3' terminal exons.

Figure S2

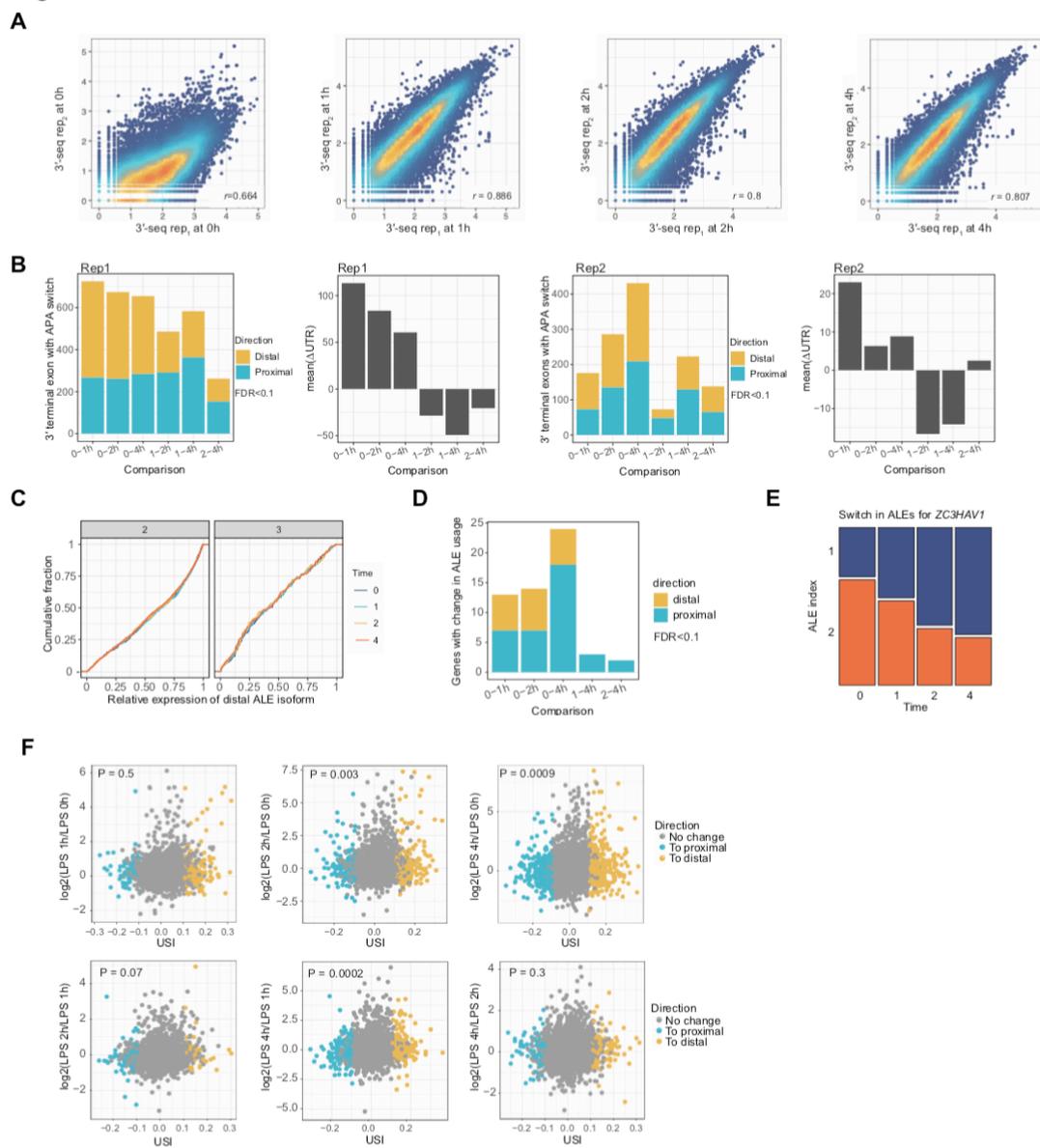


Figure S2. Analysis of 3'UTR Switch and Measurements of Poly(A) Tail Length Change.

A. Scatterplots of 3'-seq between two biological replicates. Each point indicate 3'-seq read count at a PAS isoform after CPM normalization ($n = 23,467$). Pearson correlation coefficient between biological replicates is shown at the bottom of the scatterplot. Local point density is color-coded from dark-blue to orange (low to high). Axes are \log_2 transformed after adding a pseudocount of 1 to the CPM values.

B. Change in APA usage between the indicated time points, tested in individual sets of 3'-seq biological replicates. Left panel of each replicate set, the number of 3' terminal exons with shifts in APA isoform usages, color-coded by the switching direction. Right panel of each replicate set, Changes in the weighted 3'UTR length (Δ UTR) between two time points for the 3' terminal exons with two or more PAS. Switching to distal APA isoform is shown as an increase in Δ UTR (3' UTR lengthening), and switching to proximal APA isoform shown as a decrease in Δ UTR (3'UTR shortening). Barplots on the right panels are the mean of Δ UTR values for 3' terminal exons with significant shift in APA usage ($FDR < 0.1$, and $|USI| > 0.1$).

C. The relative mRNA abundance of the distal ALE isoforms during LPS stimulation time course. Genes with two (left) and three (right) ALE isoforms were used for this analysis. Plotted are the cumulative distributions of the relative expression of the most distal ALE isoform from each gene (x-axis). No significant differences.

D. The number of genes with changes in ALE isoform usage (y-axis) between the indicated time points (x-axis). Genes with distal isoform switch are defined by Chi-squared test $FDR < 0.1$ and $USI > 0$, whereas genes with proximal switch are defined by Chi-squared test $FDR < 0.1$ and $USI > 0$. USI are calculated in the same way as the APA switch analysis. Total of 32 genes are identified to have a switch in relative ALE expression upon LPS.

E. Relative usage of individual ALE isoforms in ZC3HAV1 gene (y axis: ALE index color labeled and the height indicate the proportion of the given isoform relative to all isoforms) during macrophage activation (x-axis) as an example of proximal ALE switch upon LPS stimulation. Mean values from 3'-seq biological replicates are used in this analysis.

F. Scatterplots of tandem 3'UTR switch index (USI; x-axis) and RNA level change for all time- point comparisons. Each point is a 3'UTR exon, color -coded by the direction of the UTR isoform switch.

REFERENCES

- Alasoo K, Martinez FO, Hale C, Gordon S, Powrie F, Dougan G, Mukhopadhyay S, Gaffney DJ. 2015. Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription. *Sci Rep* **5**: 12524.
- Arora A, Goering R, Lo HYG, Lo J, Moffatt C, Taliaferro JM. 2022. The Role of Alternative Polyadenylation in the Regulation of Subcellular RNA Localization. *Frontiers in Genetics* **12**.
<https://www.frontiersin.org/article/10.3389/fgene.2021.818668> (Accessed April 15, 2022).
- Chen M, Lyu G, Han M, Nie H, Shen T, Chen W, Niu Y, Song Y, Li X, Li H, et al. 2018. 3' UTR lengthening as a novel mechanism in regulating cellular senescence. *Genome Res* **28**: 285–294.
- Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell* **43**: 853–866.
- Domingues RG, Lago-Baldaia I, Pereira-Castro I, Fachini JM, Oliveira L, Drpic D, Lopes N, Henriques T, Neilson JR, Carmo AM, et al. 2016. CD5 expression is regulated during human T-cell activation by alternative polyadenylation, PTBP1, and miR-204. *Eur J Immunol* **46**: 1490–1503.
- Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, Xu A. 2011. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* **21**: 741–747.
- Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**: 812–824.
- Gennarino VA, Alcott CE, Chen C-A, Chaudhury A, Gillentine MA, Rosenfeld JA, Parikh S, Wheless JW, Roeder ER, Horovitz DD, et al. NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *eLife* **4**: e10782.
- Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature Methods* **10**: 133–139.
- Jan CH, Friedman RC, Ruby JG, Bartel DP. 2011. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* **469**: 97–101.

- Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *PNAS* **106**: 7028–7033.
- Jia X, Yuan S, Wang Y, Fu Y, Ge Y, Ge Y, Lan X, Feng Y, Qiu F, Li P, et al. 2017. The role of alternative polyadenylation in the antiviral innate immune response. *Nat Commun* **8**.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333124/> (Accessed July 31, 2020).
- Kalam H, Fontana MF, Kumar D. 2017. Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection. *PLOS Pathogens* **13**: e1006236.
- Kalsotra A, Cooper TA. 2011. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12**: 715–729.
- Katz Y, Wang ET, Airoidi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009–1015.
- Lee JY, Yeh I, Park JY, Tian B. 2007. PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res* **35**: D165–168.
- Legnini I, Alles J, Karaiskos N, Ayoub S, Rajewsky N. 2019. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nature Methods* **16**: 879–886.
- Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, et al. 2015. Systematic Profiling of Poly(A)⁺ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. *PLOS Genetics* **11**: e1005166.
- Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. 2013. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* **27**: 2380–2396.
- Lima SA, Chipman LB, Nicholson AL, Chen Y-H, Yee BA, Yeo GW, Collier J, Pasquinelli AE. 2017. Short Poly(A) Tails are a Conserved Feature of Highly Expressed Genes. *Nat Struct Mol Biol* **24**: 1057–1063.
- Liu X, Freitas J, Zheng D, Oliveira MS, Hoque M, Martins T, Henriques T, Tian B, Moreira A. 2017. Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*. *RNA* **23**: 1807–1816.

- Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu A-B, Li W, Wagner EJ. 2014. CFIm25 links Alternative Polyadenylation to Glioblastoma Tumor Suppression. *Nature* **510**: 412–416.
- Mayr C. 2016. Evolution and Biological Roles of Alternative 3'UTRs. *Trends in Cell Biology* **26**: 227–237.
- Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**: 673–684.
- Mishima Y, Tomari Y. 2016. Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell* **61**: 874–885.
- Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**: 409–418.
- Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, Chen J, Rowley JD, Wang SM. 2002. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proceedings of the National Academy of Sciences* **99**: 6152–6156.
- Pai AA, Baharian G, Sabourin AP, Brinkworth JF, Nédélec Y, Foley JW, Grenier J-C, Siddle KJ, Dumaine A, Yotova V, et al. 2016. Widespread Shortening of 3' Untranslated Regions and Increased Exon Inclusion Are Evolutionarily Conserved Features of Innate Immune Responses to Infection. *PLOS Genetics* **12**: e1006338.
- Ren F, Zhang N, Zhang L, Miller E, Pu JJ. 2020. Alternative Polyadenylation: a new frontier in post transcriptional regulation. *Biomarker Research* **8**: 67.
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**: 1643–1647.
- Starr T, Bauler TJ, Malik-Kale P, Steele-Mortimer O. 2018. The phorbol 12-myristate-13-acetate differentiation protocol is critical to the interaction of THP-1 macrophages with Salmonella Typhimurium. *PLOS ONE* **13**: e0193601.
- Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS, et al. 2016. Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Molecular Cell* **61**: 821–833.
- Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**: 312–320.

- Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**: 18–30.
- Tian B, Pan Z, Lee JY. 2007. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res* **17**: 156–165.
- Woo YM, Kwak Y, Namkoong S, Kristjánisdóttir K, Lee SH, Lee JH, Kwak H. 2018. TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Reports* **24**: 3630-3641.e7.
- Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W. 2014. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun* **5**: 5274.
- Zheng D, Wang R, Ding Q, Wang T, Xie B, Wei L, Zhong Z, Tian B. 2018. Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nature Communications* **9**: 2268.
- Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, Engelman AN, Xie X, Hertel KJ, Shi Y. 2018. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Molecular Cell* **69**: 62-74.e4.

