Functions and Mechanisms of mTORC1-Mediated Post-Transcriptional Regulations

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Hsin-Sung Yeh

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisor: Jeongsik Yong, Ph.D

February 2021

Acknowledgements

This work is the result of the combined efforts of many brilliant and patient people who helped me along the way of my education. First of all, I would like to thank my advisor, Dr. Jeongsik Yong. I really appreciate his full support of not only my training and growth as a scientist but also my career choice and life decisions. I particularly am grateful that he constantly pushes me to think harder and work harder to get at aspects of scientific research that is truly novel and exciting.

I am also grateful for the collaborative efforts by my lab mates in the Yong lab, including Meeyeon Park, Sze Cheng, and Jae-Woong Chang. I am also thankful for the collaboration from others, including Luke Erber, Ebbing de Jong, Wei Zhang, Todd Markowski, and Kenneth Beckman.

Finally, I would like to thank my family, particularly my wife, Tiffany, for their constant support and encouragement, without which I would not have the endurance to finish this thesis work.

Abstract

Post-transcriptional regulations of mRNA transcripts such as alternative splicing and alternative polyadenylation can affect the expression of genes without changing the transcript levels. These events have significant physiological impacts on various biological systems. Nevertheless, how cellular signaling pathways control these post-transcriptional processes in cells has not been very well explored. The mammalian target of rapamycin complex 1 (mTORC1) pathway plays a key role in sensing cellular nutrient and energy status and regulating the proliferation and growth of cells. Dysregulation of mTORC1 pathway is associated with a number of pathological conditions, including cancers. It has been well-known that mTORC1 controls its downstream pathways through translational and/or transcriptional regulation of the expression of key downstream effectors. Yet recent studies have also suggested that mTORC1 can control downstream pathways via post-transcriptional regulations. In this work, I discuss the roles of post-transcriptional processes in gene expression regulations and argue that post-transcriptional regulation is an additional layer of gene expression control by mTORC1 to steer cellular biology. Particularly, I demonstrate this by showing how mTORC1 promotes the transcriptome-wide 3'-UTR shortening, as well as how it controls the expression profile of the functionally distinct isoforms of the splicing factor gene *U2AF1* in cells via alternative splicing to modulate the proteome. These emphasize the importance of studying post-transcriptional events in transcriptome datasets for gaining a fuller understanding of gene expression regulations in the biological systems of interest.

Table of Contents

ACKNOWLEDGEMENT……..……...i
ABSTRACT……...……...……....ii
LIST OF TABLES…….....v
LIST OF FIGURES…….……vi
CHAPTER 1: **The Roles of Post-Transcriptional Regulations in Gene Expression and How to Study Them**…….....…….....…….……1
Introduction to post-transcriptional regulations…….....……...2
Splicing and alternative splicing.....2
Polyadenylation and alternative polyadenylation.....6
Introduction to methods in studying alternative polyadenylation.....10
A historic overview.....10
Alternative polyadenylation studies in the era of second-generation sequencing technologies.....12
Alternative polyadenylation studies beyond the second-generation sequencing era.....19
Current objectives…….....……...……21
mTOR signaling pathway.....22
Overview of mTORC1's regulations on various biological processes.....22
References..……..……..……27
CHAPTER 2: Alternative Polyadenylation in the 3'-UTR is a New Layer of Gene Expression Control by mTORC1.....…….…….38
Summary……..……...39
Introduction……..……..40
Materials and methods…….....…….....43
Results……...……...54
Discussion……....83
Acknowledgements……..90
Supplementary materials.....91
References……..109
CHAPTER 3: mTORC1 Controls Alternative Splicing Events by Modulating the Splicing of *U2af1* Tandem Exons…….....……116
Summary……..……..117
Introduction……..……..118
Materials and methods…….....…….....121
Results……....131
Discussion…….....158
Acknowledgements……..……....161
Supplementary materials 162
References……....181
CHAPTER 4: Conclusions and Perspectives…….....189
Functional difference between U2AF1 isoforms and its potential roles in cancer biology.....190

The role of post-transcriptional regulations in mTORC1 biology and beyond...194
Post-transcriptional regulations beyond mTORC1 biology.....197
References.....201
COMPLETE BIBLIOGRAPHY…….....204

List of Tables

Chapter 1:
The Roles of Post-Transcriptional Regulations in Gene Expression and How to Study Them

Table 1: Currently available algorithms to analyze the variations of 3'-UTR length using RNA-seq data.....14

Chapter 2:
Alternative Polyadenylation in the 3'-UTR is a New Layer of Gene Expression Control by mTORC1

Supplementary Table 1: Tophat alignment statistics.....107
Supplementary Table 2: 3'-UTR shortened transcripts in the mTORC1-activated transcriptome.....107
Supplementary Table 3: The catalog of proteins identified in 2D-LC-MS/MS.....107
Supplementary Table 4: The catalogue of differentially expressed genes in WT and $TSC1^{-/-}$ MEFs.....107
Supplementary Table 5: The gene list of 3'-UTR APA by RNA-Seq and PAS-Seq.....107
Supplementary Table 6: The list of genes showing 3'-UTR APA by RNA-Seq, PAS-Seq and IntMAP.....107
Supplementary Table 7: The enrichment analysis of 3'-UTR APA profiled by IntMAP or 3 unions of APA methods.....107
Supplementary Table 8: The GO-term analysis using 280 genes belonging to IntMAP analysis only.....108

Chapter 3:
mTORC1 Controls Alternative Splicing Events by Modulating the Splicing of *U2af1* Tandem Exons

Supplementary Table 1. Alternative splicing events upon U2af1 knockdown in U2Af1a-only cell.....180
Supplementary Table 2. Alternative splicing events upon U2af1 knockdown in U2Af1b-only cell.....180
Supplementary Table 3. Differential alternative splicing events between U2Af1a only and U2AF1b-only cells.....180
Supplementary Table 4. List of identified U2AF1 isoform-specific interactors..180

List of Figures

Chapter 1:

The Roles of Post-Transcriptional Regulations in Gene Expression and How to Study Them

Figure 1: Overview of post-transcriptional regulations in eukaryotic cells.....….3
Figure 2: A work flow for 3′end-seq and PacBio Iso-seq.....16
Figure 3: Illustration of mTORC1's translational and transcriptional controls over various metabolic pathways and physiological outcomes.....24

Chapter 2:

Alternative Polyadenylation in the 3'-UTR is a New Layer of Gene Expression Control by mTORC1

Figure 1: mTOR activation leads to genome-wide 3'-UTR shortening.....56
Figure 2: 3'-UTR shortening is a downstream target of mTORC1.....59
Figure 3: 3'-UTR shortening due to mTOR activation targets specific cellular pathways including ubiquitin-mediated proteolysis.....63
Figure 4: 3'-UTR shortening is an mTORC1-activated molecular signature for protein synthesis.....68
Figure 5: Profiling of 3'-UTR APA events in the mTOR-activated transcriptome using Poly(A) Site sequencing (PAS-Seq).....72
Figure 6: Development of a novel algorithm IntMAP by integrating RNA-Seq and PAS-Seq dataset.....75
Figure 7: Identification of novel 3'-UTR APA events by IntMAP.....78
Figure 8: mTOR-coordinated cellular stress response network identified by IntMAP.....80
Supplementary Fig. 1: Genome-wide 3'UTR shortening by mTOR activation...91
Supplementary Fig. 2: The mTOR pathway targets 3'UTR shortening.....93
Supplementary Fig. 3: Differential gene expression analysis of mTOR-activated transcriptome.....97
Supplementary Fig. 4: 3'UTR shortening functions in the up-regulation of E3 ubiquitin ligase complexes.....100
Supplementary Fig. 5: Examples of RNA-Seq and PAS-Seq alignments for genes showing 3'-UTR APA events identified by IntMAP.....104
Supplementary Fig. 6: The full list of KEGG pathways enriched in the APA events by the union of RNA-Seq, PAS-Seq and IntMAP analyses.....105
Supplementary Fig. 7: Expression of Cepbg gene.....106

Chapter 3:

mTORC1 Controls Alternative Splicing Events by Modulating the Splicing of *U2af1* Tandem Exons

Figure 1: Cellular mTOR activity affects the expression profile of *U2af1* isoform.....132
Figure 2: U2AF1 isoforms display distinctive alternative splicing profiles.....138

Figure 3: Stoichiometry of U2AF1 isoforms determines alternative 3'-splice site.....144
Figure 4: Overlapping but distinct interactome profiles of U2AF1 isoforms represents refined functional differences.....150
Figure 5: U2AF1 isoform-regulated alternative splicing in 5'-UTR modulates translation.....153
Supplementary Fig. 1.....162
Supplementary Fig. 2.....166
Supplementary Fig. 3.....169
Supplementary Fig. 4.....171
Supplementary Fig. 5.....174

Chapter 4:
Conclusions and Perspectives
Figure 1: Interrogation of *U2AF1* isoform expressions in TCGA RNA-seq data collections.....192
Figure 2: Illustration of how mTORC1-mediated post-transcriptional regulations play a role in controlling various cellular pathways and physiological outcomes..195

CHAPTER ONE

The Roles of Post-Transcriptional Regulations in Gene Expression and How to Study Them

This chapter contains mini review articles previously published.

Hsin-Sung Yeh, Wei Zhang, and Jeongsik Yong (2017) *BMB Reports* 50(4):201–207; Hsin-Sung Yeh and Jeongsik Yong (2020) *Journal of Lipid and Atherosclerosis* 9(1):8-22.

Hsin-Sung Yeh wrote this chapter in its entirety.

Introduction to post-transcriptional regulations

Post-transcriptional regulations refer to the processes that RNA transcripts are subjected to between transcription and translation, namely 5'-capping, splicing, polyadenylation, RNA modification, etc. These processes greatly impact the expression of genes (Fig. 1A). Particularly, recent studies have convincingly demonstrated that post-transcriptional regulations play critical roles in controlling cellular biology and are associated with various diseases[1–4].

Splicing and Alternative Splicing

Splicing is a critical mRNA maturation process in eukaryotic cells in which intervening sequences (introns) are cut out from the nascent transcript and exon sequences are pieced together by the spliceosome to form uninterrupted coding DNA sequences (CDS) and UTR sequences for proper protein translation. This process occurs co-transcriptionally and is dependent upon the actions and activities of RNA polymerase II (Pol II) (Fig. 1A)[5–7]. In alternative splicing (AS), certain exons, or part of certain exons, are alternatively included in the mature transcripts in different biological contexts. AS occurs in about 95% of multi-exon genes in human transcriptome[8]. The regulation of AS is dependent upon the coordination of relevant *trans*-acting RNA binding proteins (RBPs) and the *cis*-acting elements surrounding the alternative exons. A series of RBPs such as SR (serine arginine)

proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) that can affect the molecular actions of the spliceosomal complexes play important roles in the regulation of

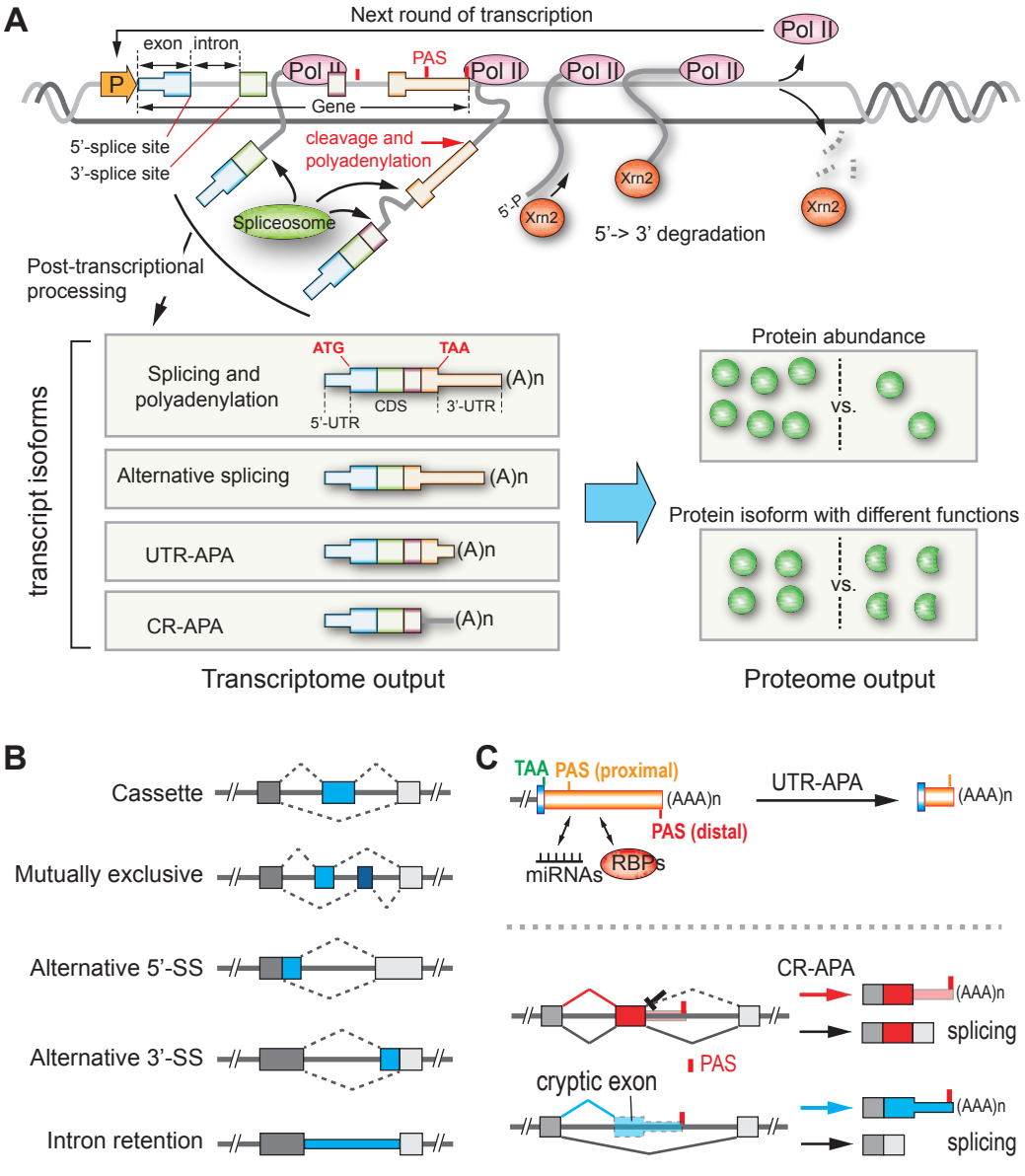


Figure 1. Overview of post-transcriptional regulations in eukaryotic cells. (A) Co-transcriptional events and transcriptional termination. Post-transcriptional processing, *i.e.* splicing, polyadenylation, alternative splicing, and alternative polyadenylation occur co-transcriptionally. These post-transcriptional events can produce transcript isoforms from genes and contribute to the diversity and dynamics of the transcriptome and the resulting proteome. P, promoter; Pol II, RNA polymerase II; PAS, poly(A) signal; 5'-P, 5' phosphate group; Xrn2, 5'-3' exoribonuclease 2. **(B)** The five different types of alternative splicing events. **(C)** The two types of Alternative polyadenylation events. (upper) 3'-UTRs serve as binding platforms of various regulatory RBPs and miRNAs. Upon UTR-APA, since most of the alternative PASs are proximal, 3'-UTRs are often shortened, resulting in the production of transcripts that can escape the regulation of those regulatory factors. (lower) The two types of CR-APA.

general RNA splicing and AS[9,10]. There are generally five basic types of AS: cassette type (or exon inclusion/skipping type) occurs when the 3'-splice site of the downstream intron is used for splicing, skipping over an entire exon; mutually exclusive type occurs when two cassette exons are mutually exclusively selected for splicing; alternative 5'-splice site type occurs when a different 5'-splice site is defined and used for splicing; alternative 3'-splice site type occurs when a different 3'-splice site is recognized; finally, intron retention occurs when neither the 5'-splice site nor the 3'-splice site of an intron is used for splicing, leaving that particular intron in the final mRNA transcript (Fig. 1B)[11,12]. Since AS alters the sequence of the resulting mature mRNA, it can impact the expression and function of the gene. For example, given that 5'-UTRs often contain sequence elements or secondary structures that can affect ribosome loading, AS of exons in the 5'-UTRs may alter the translation efficiency of the transcripts[13,14]. Moreover, AS of exons in the CDS may alter the function of certain domains or affect the folding of proteins, leading to changes in the characteristics of the protein products translated from the transcripts. These AS events increase the capacity of functional proteome[12]. And more importantly, AS events can have significant physiological impacts without necessarily affecting the transcript expression level of genes. Indeed, multiple biological contexts and diseases, e.g. cancers, have been associated with transcriptome-wide AS events[1, 2]. These demonstrate the importance of studying the roles of alternative splicing events in cellular gene expression regulations.

Polyadenylation and Alternative Polyadenylation

The termination of the transcription activity by RNA polymerase II (Pol II) involves a sequence of molecular events. Towards the end of transcription, the nascent transcript undergoes endonucleolytic cleavage to be released and polyadenylated while Pol II continues with transcription. Then, Pol II is released from the DNA for recycling and allow for the next rounds of transcription (Fig. 1A). For cleavage and polyadenylation, namely the maturation of the 3' end of mRNA molecules (sometimes simply referred to as polyadenylation), Pol II first reaches and transcribes sequence elements that recruit the formation of the 3'-end processing complex, which include a poly-A signal (PAS, most commonly AAUAAA, AUUAAA, and several other variants), and often a U-rich auxiliary upstream element (USE) and an U-rich or AU-rich downstream element (DSE). These signal the recruitment of *trans*-acting factors such as the PAS-binding cleavage and polyadenylation specificity factors (CPSFs) and the DSE-binding cleavage stimulation factors (CSTFs) to bind to the nascent transcript to catalyze an endonucleolytic reaction at a CA dinucleotide that is usually 15–30 nucleotides downstream of the PAS. Then, poly-A polymerase adds a stretch of untemplated adenosines, the poly-A tail, to the 3'-end of the transcript from the cleavage site. The poly-A tail is needed for downstream metabolism of the mature mRNA transcript including nuclear export of mRNA, translation, localization, and stability[15–17]. As for the release and the termination of the transcription reaction of Pol II, 5' to 3' exonucleases are recruited to attack the 5'-end generated by the cleavage during polyadenylation process, which is unprotected by a 5'-cap. The exonucleases (e.g. Xrn2, 5'-3' exoribonuclease 2) then chase down Pol II along their

substrate and finally displace Pol II from the transcription bubble to terminate the transcription reaction[18–20].

Thus, termination of transcription reactions is dependent upon the occurrence and "strength" of the *cis*-acting elements that signal for polyadenylation (the USE-PAS-DSE pattern). Interestingly, at least 70% of human genes are predicted to possess two or more such *cis*-acting elements[21]. And, when more than one PASs in a gene are capable to be utilized for 3'-end processing, alternative polyadenylation (APA) occurs. The regulation of APA, like AS, is determined by the coordination of various *trans*-acting factors and *cis*-acting elements surrounding the PAS and alternative PAS[3, 4]. The expression levels and activities of a number of RBPs have been shown to be able to affect APA at a transcriptome-wide level due to their roles in interacting with 3'-end processing factors and/or the *cis*-acting elements near PASs; these factors include the components of the 3'-end processing complex such as CPSFs and CSTFs, as well as other polyadenylation-associated RBPs such as cytoplasmic polyadenylation element binding protein 1 (CPEB1) and poly(A) binding protein nuclear 1 (PABPN1)[22–27].

Generally, there are two types of APA: UTR-APA and CR-APA (coding region APA)[28]. 3'-UTR of genes is often longer than CDS and contains binding sites for microRNAs and regulatory RNA-binding proteins (RBPs). It also contains alternative PASs. Most of the alternative PASs in 3'-UTRs are upstream of the canonical or annotated PASs (in this scenario, the alternative PASs are termed proximal PASs and the canonical PASs are termed distal PASs, based on their relative distance to the stop codon). Thus, when these alternative PASs are utilized for polyadenylation, namely when UTR-APA

occurs, in most cases, 3'-UTRs are shortened. This phenomenon is referred to as 3'-UTR shortening. As UTR-APA alters the availability of these regulatory *cis*-acting elements on the mature transcripts, it can affect the behaviors of mRNA transcripts, including translation efficiency, localization, and stability, without changing the coding capacity of the transcripts (Fig. 1C)[3,4,21,29,30]. For example, PAX3, one of the master transcription regulators of the myogenic transcriptional network, has a miR-206 binding site in its 3'-UTR. It has been observed that among different muscle cell types, the ratios of PAX3 UTR-APA isoforms (thus the ratio of transcripts being able to be regulated by miR-206) differ. This results in varying degrees of protein translation efficiency for *PAX3* transcripts in the different muscle cell types, which can help explain the varying differentiation patterns in these distinct muscle cell types[31]. It has also been shown that in neurons, where accurate localization of gene expression is crucial for proper cellular functions, the isoforms of hundreds of genes are differentially localized based on the UTR-APA events in their 3'-UTRs[32]. These highlight the crucial role of UTR-APA events in determining the fate of mRNA transcripts. Furthermore, the 3'-UTR of an mRNA transcript has also been shown to serve as a molecular scaffold for protein-protein interactions, particularly, immediate interactions between the nascent protein synthesized from the mRNA transcript and its binding partners. For example, the 3'-UTR of the membrane protein CD47 can recruit protein complexes including ELAV like RNA binding protein 1 (ELAVL1, or HuR) and SET nuclear proto-oncogene (SET), allowing immediate interaction of these proteins with the nascent CD47 protein. This molecular event leads to the efficient translocation of CD47 to the plasma membrane. Upon UTR-APA, *CD47* transcript can no longer recruit the

binding partners for plasma membrane localization. The CD47 protein produced from *CD47* transcript with UTR-APA localizes to endothelium reticulum, instead[33,34]. As for CR-APA, it occurs when alternative PASs in the upstream intronic regions are utilized for polyadenylation. It is thus also sometimes referred to as intronic APA. Once a transcript is truncated, the transcript would lose the coding capacity of a chunk of polypeptide on the C-terminal end. The resulting protein product may thus function differently. Moreover, the truncated transcript would be differentially regulated compared to the full-length counterpart as it would possess a completely different 3'-UTR that originates from the intron region downstream of the alternative PAS. There are two types of CR-APA; the mechanism of both types of CR-APA involves the interplay between splicing and polyadenylation[3,28,35]. The first type of CR-APA occurs when the splicing of a PAS-containing intron is inhibited, and the 3'-end processing complex outcompetes the splicing machinery, leading to truncation of the transcript at that intron. The second type of CR-APA occurs when a cryptic exon that is followed by a PAS is utilized for splicing. Due to the presence of the PAS, this "alternative splicing" event leads to the truncation of the mRNA transcript (Fig. 1C). While the studies in the field focusing on CR-APA is relatively few, one of the first and most famous identified APA events belongs to this category. In resting B lymphocytes, IGHM gene (IgM) is transcribed to full-length, which includes the last two exons that encode a transmembrane domain, leading to the expression of membrane-bound antibody chain. And upon B lymphocyte activation, CR-APA of *IGHM* transcript occurs, causing the two exons encoding the transmembrane domain to be excluded from the final mRNA transcript. This leads to the production of secreted form of

IgM antibody chain[36]. Another example of CR-APA is the MAPK interacting serine/threonine kinase 1 (MKNK1) gene. The CR-APA event of MKNK1 gene results in the loss of 89 amino acids at the C-terminal end of the kinase and the addition of 12 amino acids encoded by the intronic sequence. It has been shown that this truncated version of MKNK1 has a higher kinase activity compared to its full-length counterpart[37]. Moreover, intriguingly, a recent study found that in chronic lymphocytic leukemia cells, a number of tumor suppressor genes undergo CR-APA and consequently lose their tumor-suppressing activities. Some of the truncated proteins even seemed to act in an oncogenic manner[38]. Together, all these previous findings show that APA can drastically affect gene expression and steer cellular biology independent of the transcription regulation of genes.

Introduction to Methods in Studying Alternative Polyadenylation

Having established the importance of understanding alternative post-transcriptional events in cells, below I discuss the methods that have been used to study them. In this section, methods for studying APA events are highlighted. A bioinformatics tool developed by our group for studying transcriptome-wide AS events will be presented and discussed in Chapter 3.

A Historic Overview

Some of the earliest APA discoveries were reported in the 1980s. For example, a CR-APA event was observed in IgM gene, and DHFR gene showed UTR-APA[36,39–41]. These cases of alternative processing events were first revealed by the discrepancies of sizes of the same gene in northern-blotting, and western-blotting, in the case of CR-APA. R-loop mapping and restriction mapping were then used to confirm that the differences reside in the 3' end structure of the transcripts. In the following decade, dozens of APA events were discovered by the similar approach, albeit at a one-gene-at-a-time pace.

As technologies in molecular biology matured and sequencing data accumulated, APA studies were introduced to a more global scale in the 2000s. The first large-scale APA surveys were done by analyzing Expressed Sequence Tag (EST) data of human, mouse, and rat. To search for poly(A) sites in the genomes, Tian *et al*.[42], as well as Yan and Marr[43], first aligned ESTs to the genomes, then singled out 3' end ESTs by looking for stretches of As and Ts at 5' or 3' termini of unaligned EST sequences. These 3' end ESTs were then validated by searching for the presence of consensus PAS sequence patterns. Their analyses showed that a great proportion of genes (~50% in human and ~30% in murine) have APA. Moreover, many of the APA events are conserved between human and mouse; indicating that APA is a widely-employed gene regulation strategy in cellular biology.

Analyzing EST data revealed the presence of APA in genes at a transcriptome-wide scale. However, the dynamics of global APA regulations remained elusive until microarray-based approaches were used to study global APA pattern changes[30,44,45]. In these studies, probes on the microarrays were designed to be APA sensitive. For each APA regulated gene, there are two or more probes specific to only the full-length transcript, and

two or more probes specific to both the full-length and the shorter APA product on the microarray chip. After applying fluorescently-labelled nucleic acid library to the chip for hybridization, the ratio of the signals from these probes can then be calculated to measure the APA status of the gene. Microarray data obtained from two different cellular conditions can then be compared to study the APA dynamic changes and its physiological implications. Surprisingly, by adopting this approach it was shown that highly proliferating cells tend to have more 3' UTR shortening in their transcriptomes; while generating induced pluripotent stem cells, global 3' UTR lengthening is observed[30,44]. Microarray is a powerful tool to obtain a global picture on APA events in the transcriptome. Nevertheless, it suffers from several drawbacks. First of all, microarray cannot detect novel APA events, for APA-sensitive probe sets can only be designed if an APA events are previously known. Second, it cannot precisely pinpoint where the poly(A) site is, which may be important for studying the physiological functions of APA events. Moreover, if a gene has two or more alternative PAS, probe design and quantification can become quite complicated and challenging[46].

Alternative Polyadenylation Studies in the Second-Generation Sequencing Era

The advent of second-generation sequencing technologies enabled researchers to rapidly obtain a large amount of sequence information at single nucleotide resolution. Technologies such as RNA-seq quickly became commonly used for surveying the transcriptome of various cell types and tissues in different organisms[47–49]. In RNA-seq, poly(A) tail-containing RNA is first isolated from total RNA. They are then either primed

with oligo d(T) primer or random hexamers for cDNA synthesis followed by fragmentation. (Alternatively, the poly(A)-containing RNA pool is first fragmented followed by random hexamer priming to generate cDNA pool.) The cDNA pool is then amplified and constructed into library, which can be sequenced by various sequencing platforms, most commonly the Illumina sequencing technology. After mapping the short sequence fragments, or reads, to the corresponding genome, the reads can be piled up for visualization of gene expression profile of the cell or tissue.

The highly quantitative nature of RNA-seq makes it suitable for APA pattern analysis. This may be achieved in a similar way as the APA calculation done in microarray approaches, namely, by taking the ratio of the read density of the long form-only regions and the read density of the regions common to both long and short transcripts. However, since many genes contain isoforms with complicated and overlapping structures, using RNA-seq reads for APA analysis on certain genes can still be challenging. Fortunately, many sophisticated bioinformatics tools have been developed to more accurately analyze APA patterns in transcriptomes.

For instance, a probabilistic transcript quantification method named Kallisto was developed to estimate the expression levels of annotated transcript isoforms[50]. APA dynamics of a gene can then be measured by comparing the expression ratios of its short isoforms over the long isoforms between two biological conditions. This is the general method for CR-APA analysis.

For UTR-APA identification and measurement, many algorithms were written as listed in Table 1. In general, 3'UTR length changes are measured by modeling the RNA-seq read density changes near the 3' end of mRNA transcripts.

Algorithm	Reference	Description
DaPars	[51]	It first models the RNA-seq-read densities of both tumor and normal as a linear combination of both proximal and distal polyA sites. It then uses a linear regression model to identify the location of the *de novo* proximal polyA site, followed by quantification of the changes in APA between tumor and normal.
ChangePoint	[52]	It is based on a generalized likelihood ratio statistic for identifying 3'UTR length change in the analysis of RNA-seq data. A directional multiple test procedure is then developed to identifying APA events between two samples.
Roar	[39]	It is based on Fisher test to detect disequilibriums in the number of RNA-seq reads mapped to the 3'UTRs. Read counts and lengths of fragments are then used to calculate the prevalence of the short isoform over the long one in two biological conditions to identify APA events.
3USS	[53]	A web-server developed with the aim of giving experimentalists the possibility to identify alternative 3'UTRs between two samples by RNA-seq data analysis.
IsoSCM	[54]	A method for transcript assembly that incorporates change point analysis by a Bayesian framework to improve the 3'UTR annotation process with RNA-seq data.
KLEAT	[55]	An analysis tool that uses *de novo* assembly of RNA-seq data to characterize cleavage sites on 3'UTRs through direct observation of poly(A) tails.
GETUTR	[56]	It first makes a density function of RNA-seq reads aligned to the 3'UTRs using kernel density estimation. A smoothing step is then applied to maintain the biological changes of the 3'UTR. The goal of the method is to estimate the 3'UTR landscape based on these smoothed RNA-seq signal.

Table 1. Currently available algorithms to analyze the variations of 3′-UTR length using RNA-seq data.

Indeed, with the aid of these bioinformatics tools, RNA-seq can be a powerful tool to study the alternative processing of mRNAs. However, when profiling APA patterns, especially when handling genes with multiple isoforms, often times the reads mapped to regions that differentiate isoforms constitute only a relatively small portion of the total reads mapped to the gene, and even less so for 3' end junction reads, making it rather challenging to confidently calculate the expression ratios of different isoform. Moreover, RNA-seq is not particularly accurate when it comes to identifying poly(A) sites, making novel APA isoform identification rather difficult. Therefore, several methods have been developed to address these issues by enriching for 3' end reads in high-throughput sequencing experiments[46,57,58].

The most common way to enrich for 3' end reads (adopted in PAS-seq[59], A-seq[22], 3SEQ[60], SAPAS[61], etc.) is to first fragment the poly(A) tail-containing RNA pool, followed by reverse transcription using oligo d(T) priming (Fig. 2A). The cDNA pool, which should only contain 3' end junction fragments, is then amplified and sequenced. Alternatively, an oligo d(T) primed cDNA library can be sequenced using oligo d(T) sequencing primer directly. All the sequencing reads should therefore be 3' end junction reads (PolyA-seq[62]). Moreover, direct RNA sequencing technology by Helicos Biosciences has also been used for sequencing the 3' ends of poly(A) tail-containing RNAs[63]. In this method, mRNA molecules are hybridized to a "lawn" of oligo d(T) primers attached to the flow cell and are sequenced directly by synthesis. Compared to PAS-seq and equivalents, Helicos platform is more quantitative as no amplification step is involved; it requires less starting

material. However, Helicos platform suffers from higher error rate, shorter read lengths, lower throughput, and the lack of multiplexing capability[46,49].

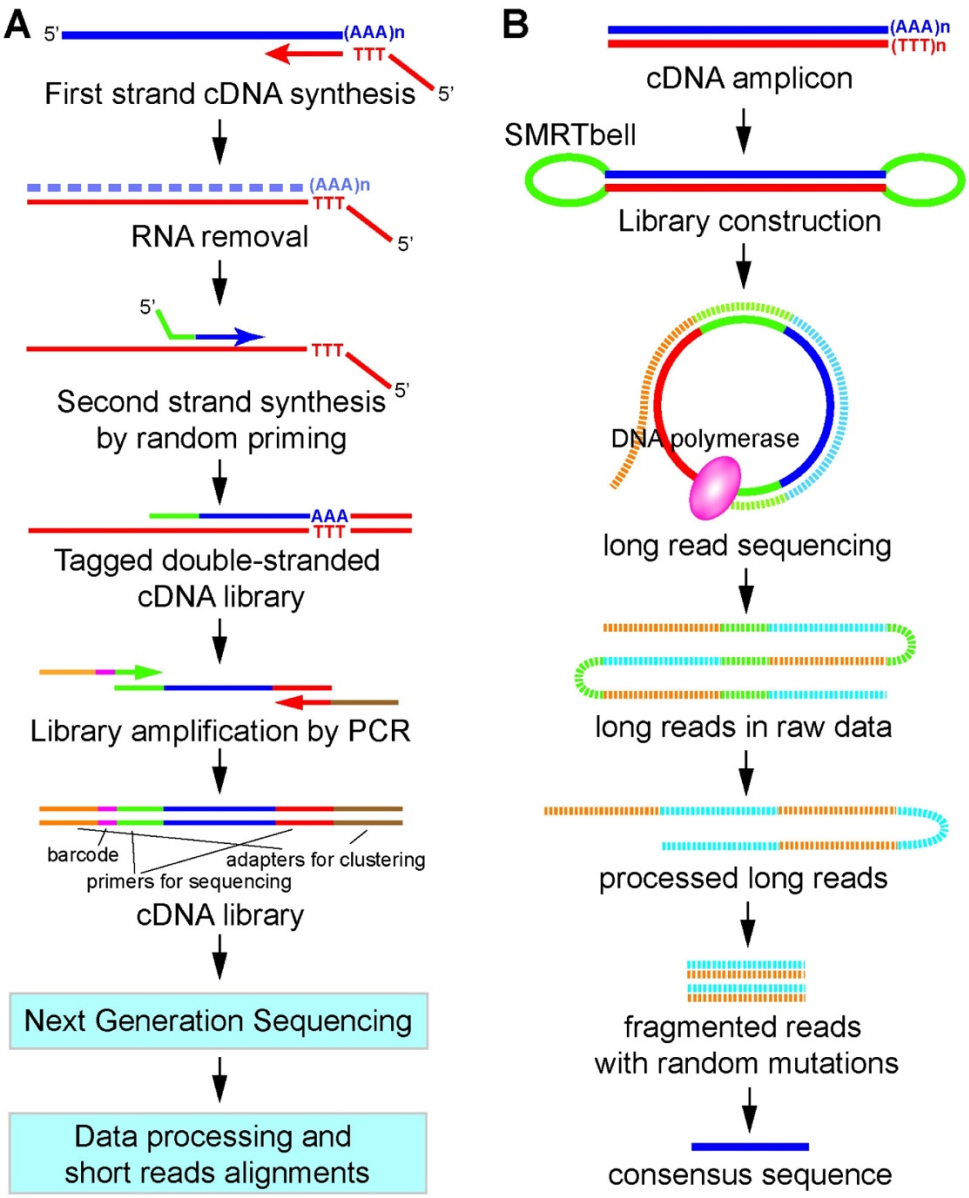


Figure 2. A work flow for 3′end-seq and PacBio Iso-seq. (A) Multiple versions of global profiling method for 3′-end sequence of mRNAs are available. An example of 3′end-seq method is shown. A 3′end-seq cDNA library is produced by a series of

molecular biology work integrating first strand cDNA synthesis and PCR. Short sequence reads of polyadenylation site are cataloged by conducting RNA-seq using the cDNA library and trimming/aligning sequencing data. (B) The SMRT bell cDNA library for PacBio-seq can be produced from cDNA amplicon which is made by reverse transcription. Concatemerized long reads of insert in SMRT bell cDNA library are produced as raw data. Processed long reads (by eliminating SMRT bell sequences) are aligned to generate a consensus sequence of long reads.

All of the above-mentioned methods use oligo d(T) for priming at some points during the procedures. Internal priming (stretches of As in the middle of transcripts being falsely recognized as poly(A) tails by the oligo d(T) primer), and thus false identification of 3' end junctions, is therefore a major issue for these methods[64]. In an effort to lower the false discovery rate, Jan *et al.* developed a modified version of 3' end sequencing method that avoided the use of oligo d(T) priming, named 3P-Seq[64]. After isolating poly(A) tail-containing RNAs, the first step of 3P-seq is to add a biotinylated double-stranded adapter to the 3' end of the poly(A) tail through splint-ligation, which eliminates the possibility of internal priming. The mRNAs are then partially digested, and the 3' end fragments are captured by streptavidin. cDNA synthesis is primed with the adapter itself and reverse-transcribed with dTTP as the only deoxynucleoside triphosphate present, limiting the reverse transcription to the poly(A) site. The RNA fragments immediately upstream of poly(A) tails can then be released and processed for sequencing by RNase H digestion. Since RNase H would only digest RNA strand that is hybridized with a DNA strand, in this case the poly(A) tail region, the RNA fragments released after RNase H digestion should most likely come from poly(A) tail-containing fragments. This method indeed eliminates a great number of falsely identified 3' ends, yet it is more labor intensive and involves more enzymatic reactions, which may introduce biases in terms of the quantification of the signals[65].

3' end sequencing data provides sophisticated knowledge for pinpointing annotated as well as unannotated poly(A) sites in the transcriptome that is under interrogation. As mentioned above, although RNA-seq is highly quantitative, yet it does not provide enough

information to accurately identify poly(A) sites. Therefore, by incorporating 3' end-seq data with RNA-seq data, the quality of APA profiling can be greatly improved. Briefly, by analyzing 3' end-seq data, potential poly(A) sites and thus isoform structures in the transcriptome can be defined and reported. The expression levels of the isoforms can then be estimated by a maximum-likelihood method that best explains the observed RNA-seq read profiles. Finally, the APA events can be measured by the expression ratios of the isoforms in the gene between two biological conditions.

Alternative Polyadenylation Studies Beyond the Second-Generation Sequencing Era

Second-generation sequencing technologies have revolutionized the research involving transcriptome characterization. However, when it comes to expression profiling of mRNA isoforms, these methods still suffer from their limitation in read lengths. Due to the relatively short read lengths (~100 bp), compared to the lengths of most transcripts, full-length transcript isoforms must be reconstructed via various computational methods. Yet the performances of the reconstruction methods have been shown to be unsatisfactory[66]. Alternative sequencing platforms have been developed to achieve longer read lengths.

For example, the SMRT (single molecule, real-time) sequencing technology by PacBio has achieved average read length of > 10,000 bp[67]. In SMRT technology, DNA polymerase is immobilized to the bottom of a specialized light detecting well called zero-mode waveguide (ZMW). ZMW is designed to only be light sensitive at the bottom of the well, where sequencing by synthesis is performed by the DNA polymerase. A movie that

contains the sequencing information can then be recorded as a single DNA molecule is replicated by the DNA polymerase in full-length. PacBio has also developed a protocol specifically for transcript isoform characterization called Iso-Seq (Fig. 2B). It has been successfully adopted in characterizing the transcriptomes of human and herpesvirus[68,69]. The transcriptome dynamics during lineage commitment of blood cell and the progression of brain tumor have also been characterized by Iso-Seq[70,71]. Last but not least, it was recently used to profile the APA events in sorghum transcriptome[72]. In all of these studies, due to the long read-lengths, novel isoforms have been identified and characterized with high confidence.

However, SMRT sequencing still suffers from certain drawbacks. For instance, longer transcripts still cannot be sequenced in full-length in high quality. This is partly due to the limitations in library preparation and the limitation of read length (or movie time), and also the fact that shorter cDNA molecules (~1.5 kb) are more favored by the sequencing platform. Moreover, since shorter transcripts are more favored than longer transcripts during sequencing, the quantitative performance of Iso-Seq is severely affected[67].

To harness the quantitative power of the short-read second generation sequencing and the isoform characterization ability of long-read PacBio sequencing, hybrid sequencing has been developed. By integrating the long read data from PacBio and the short read data from Illumina, Au *et al.* significantly reduced the error rate of long reads in PacBio sequencing[73]. When applying the hybrid sequencing to human embryonic stem cell transcriptome, they demonstrated the higher sensitivity and accuracy of isoform

characterization, as well as a better ability to identify novel isoforms, over traditional methods solely using short-read second generation sequencing method[74].

In summary, with the advance of modern technologies and the rapid accumulation of the knowledge on the physiological outcomes of APA events in the field, in the near future, transcriptome-wide APA analysis can potentially become as routinely and easily performed as it currently is with gene expression profiling by RNA-seq; more importantly, valuable insights on various biological processes and the pathogenesis of diseases can be obtained by APA analysis.

In my research, we adopt second-generation sequencing technologies, namely RNA-seq, to study the dynamics of post-transcriptional regulations in cells under various conditions. We have also worked to improve upon the above-mentioned inherent limitations and drawbacks of RNA-seq when applied to APA (Chapter 2) and AS (Chapter 3) profiling.

Current Objectives

So far, I have presented the critical roles post-transcriptional regulation events play in controlling cellular biology and in the pathogenesis of varying diseases; I have also discussed the methods that are available for researchers to study them. Here is the gap in knowledge in the field: our knowledge on the mechanistic integration of the control of post-transcriptional regulations to cellular signaling pathways is still largely lacking. To begin to undertake the exploration of the integration between cellular signaling pathways and the

regulation of post-transcriptional regulations, our lab decided to start by focusing on a well-known signaling pathway: mammalian target of rapamycin signaling pathway.

mTOR Signaling Pathway

Mammalian target of rapamycin (mTOR) is a serine/threonine kinase that forms two functional complexes known as mTOR Complex 1 (mTORC1) and mTOR Complex 2 (mTORC2). Raptor and Rictor are the specific components for mTORC1 and mTORC2, respectively. Although mTORC2 has been functionally associated with cell survival, metabolic control and cytoskeleton organization, its regulations and functions still await more thorough studies to be revealed[75]. On the other hand, mTORC1 has been extensively studied in the past few decades due to its important roles in regulating various anabolic and catabolic processes, as well as its involvement in the pathogenesis of cardiovascular diseases, diabetes, and many types of cancers[75–77]. As a master regulator of various metabolic processes, mTORC1 is sensitive to growth factor signaling, nutrient and oxygen availability, and intracellular energy status.

Overview of mTORC1's Regulations on Various Biological Processes

It is generally understood that the activated mTORC1 promotes key anabolic processes including protein synthesis, lipid synthesis, nucleotide synthesis, biogenesis of organelles, and that it inhibits certain catabolic processes such as autophagy. As a result, activation of mTORC1 leads to cellular growth and proliferation[78]. On the other hand, decreased mTORC1 activity, *e.g.* when cells experience nutrient deprivation, can activate autophagy

and lower mitochondrial membrane potential in cells so that cellular energy may be conserved[79,80]. Moreover, activation of mTORC1 has also been shown to promote angiogenesis and inflammation in certain tissues[81–83]. It is thus not surprising that mTORC1 inhibition has been proposed and used as a treatment for several types of cancers and atherosclerosis[84,85].

mTORC1 is negatively regulated by tuberous sclerosis complexes (TSC1/2), which are downstream of and can be regulated by several major cellular signaling pathways including PI3K/PKB (phosphoinositide-3-kinase/protein kinase B) signaling pathway, AMPK (AMP-activated protein kinase) signaling pathway, MAPK/ERK (mitogen-activated protein kinase/extracellular-signal-regulated kinase) pathway, etc[75].

Many studies have shown that mTORC1 regulates its various downstream pathways through translational and transcriptional controls. One of the most well-known functions of mTORC1 is the direct phosphorylation of ribosomal protein S6 kinase beta-1 (S6K1) and eukaryotic initiation factor 4E-binding protein 1 (4E-BP1) (Fig. 3). These lead to the increase of protein synthesis efficiency, particularly for transcripts containing 5'-terminal oligopyrimidine tract (TOP) or 5'-pyrimidine-rich translational element (PRTE) in their 5'-untranslated regions (UTRs)[86–88]. Moreover, mTORC1 phosphorylates protein phosphatase 2A (PP2A) and transcription initiation factor 1A (TIF-1A) to increase the transcription of ribosomal RNAs, enabling efficient cellular protein synthesis (Fig. 3)[89]. One of the genes whose protein production is elevated by activated mTORC1 is the activating transcription factor 4 (ATF4). Through the transcription regulation of ATF4 on several genes involved in *de novo* purine synthesis, mTORC1 can promote purine synthesis

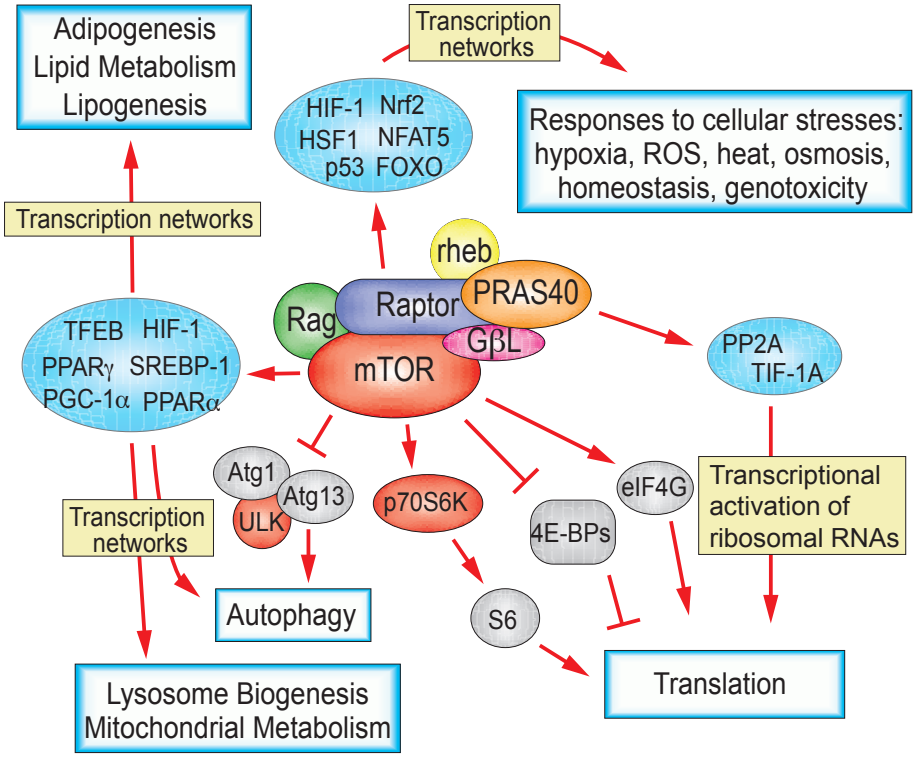


Figure 3. Illustration of mTORC1's translational and transcriptional controls over various metabolic pathways and physiological outcomes. The activation of mTORC1 not only leads to the upregulation of cellular translation activity, mTORC1 also regulates various metabolic pathways through controlling transcription networks.

in cells[90]. mTORC1 also positively regulate lipid and sterol biosynthesis by activating transcription factors such as sterol regulatory element binding protein 1 (SREBP1), peroxisome proliferator-activated receptor γ (PPARγ), and PPARγ coactivator 1 (PGC1-α), which control the transcription of genes involved in *de novo* lipid synthesis as well as lipid and cholesterol homeostasis[91]. Through PGC1-α, mTORC1 also controls mitochondrial biogenesis and oxidative metabolism (Fig. 3)[92]. On top of these, it has been shown that mTORC1 can promote angiogenesis by increasing the translation activity of hypoxia-inducible factor 1 subunit alpha (HIF-1α) gene, a transcription factor that regulates the expression of several angiogenic growth factors including vascular endothelial growth factor (VEGF)[81]. Furthermore, through regulating the activities of inflammatory transcription factors such as nuclear factor kappa B (NF-κB) and signal transducer and activator of transcription-3 (STAT3), mTORC1 can also control pro- and anti-inflammatory responses in blood cells, depending on cellular contexts[83]. Again, as these examples demonstrate, it has been well-established that mTORC1 regulates a variety of cellular processes through controlling the translation and transcription activities of its downstream effector genes (Fig. 3).

Nonetheless, whether mTORC1 also regulates cellular processes through controlling the post-transcriptional events in cells has not been explored directly in the field. And this is where our research focuses on. In the following chapters, I present the works we have done in pioneering the mechanistic integration between this well-known cellular signaling pathway to the controls of post-transcriptional regulations in cells. Specifically, in Chapter 2, I present our work in showing mTORC1's regulation of

transriptome-wide 3'-UTR shortening events, leading to the upregulation of the protein production of the targeted genes; in Chapter 3, I present our work on mTORC1's role in regulating AS events in the transcriptome through its modulation of U2AF1 isoform expression. In these chapters, I also discuss the functional outcomes of these mTORC1-mediated post-transcriptional regulations in cellular biology.

References

1. El Marabti, E. & Younis, I. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front Mol Biosci* **5**, 80–80 (2018).
2. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development* **24**, 2343–2364 (2010).
3. Tian, B. & Manley, J. L. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**, 312–320 (2013).
4. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**, 853–866 (2011).
5. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology* **18**, 637 (2017).
6. Hicks, M. J., Yang, C.-R., Kotlajich, M. V. & Hertel, K. J. Linking Splicing to Pol II Transcription Stabilizes Pre-mRNAs and Influences Splicing Patterns. *PLOS Biology* **4**, e147 (2006).
7. David, C. J., Boyne, A. R., Millhouse, S. R. & Manley, J. L. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. *Genes & Development* **25**, 972–983 (2011).
8. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* **40**, 1413 (2008).

9. Huelga, S. C. *et al.* Integrative Genome-wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins. *Cell Reports* **1**, 167–178 (2012).

10. Zhou, Z. & Fu, X.-D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122**, 191–207 (2013).

11. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences* **111**, E5593 (2014).

12. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed Rep* **3**, 152–158 (2015).

13. Rosenstiel, P. *et al.* Functional characterization of two novel 5' untranslated exons reveals a complex regulation of NOD2 protein expression. *BMC Genomics* **8**, 472 (2007).

14. Wang, G., Guo, X. & Floros, J. Differences in the translation efficiency and mRNA stability mediated by 5′-UTR splice variants of human SP-A1 and SP-A2 genes. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **289**, L497–L508 (2005).

15. Hsin-Sung, Y. & and Jeongsik, Y. Alternative Polyadenylation of mRNAs: 3′-Untranslated Region Matters in Gene Expression. *Mol. Cells* **39**, 281–285 (4).

16. Zhang, X., Virtanen, A. & Kleiman, F. E. To polyadenylate or to deadenylate: that is the question. *Cell Cycle* **9**, 4437–4449 (2010).

17. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev* **25**, 1770–1782 (2011).

18. Kornblihtt, A. R. Shortcuts to the end. *Nature Structural & Molecular Biology* **11**, 1156–1157 (2004).

19. West, S., Gromak, N. & Proudfoot, N. J. Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525 (2004).

20. Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517–522 (2004).

21. Elkon, R., Ugalde, A. P. & Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics* **14**, 496 (2013).

22. Martin, G., Gruber, A. R., Keller, W. & Zavolan, M. Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. *Cell Reports* **1**, 753–763 (2012).

23. Masamha, C. P. *et al.* CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **510**, 412 (2014).

24. Yang, Q., Gilmartin, G. M. & Doublié, S. Structural basis of UGUA recognition by the Nudix protein CFI_m25 and implications for a regulatory role in mRNA 3′ processing. *Proceedings of the National Academy of Sciences* **107**, 10062 (2010).

25. Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation. *Cell* **87**, 941–952 (1996).

26. Bava, F.-A. *et al.* CPEB1 coordinates alternative 3′-UTR formation with translational regulation. *Nature* **495**, 121 (2013).

27. Jenal, M. *et al.* The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. *Cell* **149**, 538–553 (2012).

28. Hsin-Sung, Y. *et al.* Analyses of alternative polyadenylation: from old school biochemistry to high-throughput technologies. *BMB Reports* **50**, 201–207 (4).

29. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA Translation and Stability by microRNAs. *Annual Review of Biochemistry* **79**, 351–379 (2010).

30. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).

31. Boutet, S. C. *et al.* Alternative Polyadenylation Mediates MicroRNA Regulation of Muscle Stem Cell Function. *Cell Stem Cell* **10**, 327–336 (2012).

32. Ciolli Mattioli, C. *et al.* Alternative 3′ UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. *Nucleic Acids Research* **47**, 2560–2573 (2018).

33. Berkovits, B. D. & Mayr, C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363 (2015).

34. Mayr, C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* **26**, 227–237 (2016).

35. Hoque, M. *et al.* Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nature Methods* **10**, 133 (2012).

36. Rogers, J. *et al.* Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin μ chain. *Cell* **20**, 303–312 (1980).

37. O'Loghlen, A. *et al.* Identification and molecular characterization of Mnk1b, a splice variant of human MAP kinase-interacting kinase Mnk1. *Experimental Cell Research* **299**, 343–355 (2004).

38. Lee, S.-H. *et al.* Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* **561**, 127–131 (2018).

39. Grassi, E., Mariella, E., Lembo, A., Molineris, I. & Provero, P. Roar: detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinformatics* **17**, 423 (2016).

40. Alt, F. W. *et al.* Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. *Cell* **20**, 293–301 (1980).

41. Early, P. *et al.* Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* **20**, 313–319 (1980).

42. Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research* **33**, 201–212 (2005).

43. Yan, J. & Marr, T. G. Computational analysis of 3′-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome Research* **15**, 369–375 (2005).

44. Ji, Z. & Tian, B. Reprogramming of 3′ Untranslated Regions of mRNAs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. *PLOS ONE* **4**, e8419 (2009).

45. Flavell, S. W. *et al.* Genome-Wide Analysis of MEF2 Transcriptional Program Reveals Synaptic Target Genes and Neuronal Activity-Dependent Polyadenylation Site Selection. *Neuron* **60**, 1022–1038 (2008).

46. Shi, Y. Alternative polyadenylation: New insights from global analyses. *RNA* **18**, 2105–2117 (2012).

47. Nagalakshmi, U. *et al.* The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* **320**, 1344 (2008).

48. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

49. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* **12**, 87–98 (2011).

50. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525–527 (2016).

51. Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nature Communications* **5**, 5274 (2014).

52. Wang, W., Wei, Z. & Li, H. A change-point model for identifying 3′UTR switching by next-generation RNA sequencing. *Bioinformatics* **30**, 2162–2170 (2014).

53. Le Pera, L., Mazzapioda, M. & Tramontano, A. 3USS: a web server for detecting alternative 3′UTRs from RNA-seq experiments. *Bioinformatics* **31**, 1845–1847 (2015).

54. Shenker, S., Miura, P., Sanfilippo, P. & Lai, E. C. IsoSCM: improved and alternative 3′ UTR annotation using multiple change-point inference. *RNA* **21**, 14–27 (2015).

55. Birol, I. *et al.* Kleat: cleavage site analysis of transcriptomes. *Pac Symp Biocomput* 347–358 (2015).

56. Kim, M., You, B.-H. & Nam, J.-W. Global estimation of the 3′ untranslated region landscape using RNA sequencing. *Methods* **83**, 111–117 (2015).

57. Mangone, M. *et al.* The Landscape of C. elegans 3′UTRs. *Science* **329**, 432 (2010).

58. Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H. & Fu, X.-D. A multiplex RNA-seq strategy to profile poly(A+) RNA: Application to analysis of transcription response and 3′ end formation. *Genomics* **98**, 266–271 (2011).

59. Shepard, P. J. *et al.* Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**, 761–772 (2011).

60. Beck, A. H. *et al.* 3′-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples. *PLOS ONE* **5**, e8768 (2010).

61. Fu, Y. *et al.* Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Research* **21**, 741–747 (2011).

62. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Research* **22**, 1173–1183 (2012).

63. Ozsolak, F. *et al.* Direct RNA sequencing. *Nature* **461**, 814–818 (2009).

64. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. *Nature* **469**, 97–101 (2011).

65. Hafner, M. *et al.* RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697–1712 (2011).

66. Steijger, T. *et al.* Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods* **10**, 1177–1184 (2013).

67. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* **13**, 278–289 (2015).

68. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology* **31**, 1009–1014 (2013).

69. O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Research* **44**, e145–e145 (2016).

70. Chen, L. *et al.* Transcriptional diversity during lineage commitment of human blood progenitors. *Science* **345**, 1251033 (2014).

71. Singh, N. *et al.* IsoSeq analysis and functional annotation of the infratentorial ependymoma tumor tissue on PacBio RSII platform. *Meta Gene* **7**, 70–75 (2016).

72. Abdel-Ghany, S. E. *et al.* A survey of the sorghum transcriptome using single-molecule long reads. *Nature Communications* **7**, 11706 (2016).

73. Au, K. F., Underwood, J. G., Lee, L. & Wong, W. H. Improving PacBio Long Read Accuracy by Short Read Alignment. *PLOS ONE* **7**, e46679 (2012).

74. Au, K. F. *et al.* Characterization of the human ESC transcriptome by hybrid sequencing. *Proc Natl Acad Sci USA* **110**, E4821 (2013).

75. Laplante, M. & Sabatini, D. M. mTOR signaling at a glance. *Journal of Cell Science* **122**, 3589 (2009).

76. Laplante, M. & Sabatini, D. M. mTOR Signaling in Growth Control and Disease. *Cell* **149**, 274–293 (2012).

77. Ma, X. M. & Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nature Reviews Molecular Cell Biology* **10**, 307 (2009).

78. Guertin, D. A. & Sabatini, D. M. Defining the Role of mTOR in Cancer. *Cancer Cell* **12**, 9–22 (2007).

79. Hardie, D. G. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nature Reviews Molecular Cell Biology* **8**, 774 (2007).

80. Codogno, P. & Meijer, A. J. Autophagy and signaling: their role in cell survival and cell death. *Cell Death & Differentiation* **12**, 1509–1518 (2005).

81. Karar, J. & Maity, A. PI3K/AKT/mTOR Pathway in Angiogenesis. *Front Mol Neurosci* **4**, 51–51 (2011).

82. Liu, J., Ma, K., Gao, M., Zhang, X. & Liu, B. The activation of mTOR pathway induced by inflammation accelerates the progression of atherosclerosis in hemodialysis patients. *International Journal of Cardiology* **152**, S5–S6 (2011).

83. Jones, R. G. & Pearce, E. J. MenTORing Immunity: mTOR Signaling in the Development and Function of Tissue-Resident Immune Cells. *Immunity* **46**, 730–742 (2017).

84. Kurdi, A., Martinet, W. & De Meyer, G. R. Y. mTOR Inhibition and Cardiovascular Diseases: Dyslipidemia and Atherosclerosis. *Transplantation* **102**, (2018).

85. Xie, J., Wang, X. & Proud, C. G. mTOR inhibitors in cancer therapy. *F1000Res* **5**, F1000 Faculty Rev-2078 (2016).

86. Richter, J. D. & Sonenberg, N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433**, 477–480 (2005).

87. Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109 (2012).

88. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55 (2012).

89. Mayer, C., Zhao, J., Yuan, X. & Grummt, I. mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. *Genes & Development* **18**, 423–434 (2004).

90. Ben-Sahra, I., Hoxhaj, G., Ricoult, S. J. H., Asara, J. M. & Manning, B. D. mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. *Science* **351**, 728 (2016).

91. Düvel, K. *et al.* Activation of a metabolic gene regulatory network downstream of mTOR complex 1. *Mol Cell* **39**, 171–183 (2010).

92. Cunningham, J. T. *et al.* mTOR controls mitochondrial oxidative function through a YY1–PGC-1α transcriptional complex. *Nature* **450**, 736 (2007).

CHAPTER TWO

Alternative Polyadenylation in the 3'-UTR is a New Layer of Gene Expression Control by mTORC1

This chapter contains original research articles previously published.

Jae-Woong Chang[#], Wei Zhang[#], Hsin-Sung Yeh, Ebbing P. de Jong, Semo Jun, Kwan-Hyun Kim, Sun S. Bae, Kenneth Beckman, Tae Hyun Hwang, Kye-Seong Kim, Do-Hyung Kim, Timothy J. Griffin, Rui Kuang, and Jeongsik Yong (2015) *Nature Communications* 15(6):7218;

[#]These authors contribute equally to this work

Jae-Woong Chang[#], Wei Zhang[#], Hsin-Sung Yeh, Meeyeon Park, Chengguo Yao, Yongsheng Shi, Rui Kuang, and Jeongsik Yong (2018) *Nucleic Acids Research* 46(12): 5996–6008.

[#]These authors contribute equally to this work

Hsin-Sung Yeh and Jae-Woong Chang performed the experiments in Figures 2, 4, and 8.

Summary

Mammalian target of rapamycin complex 1 (mTORC1) enhances translation from a subset of messenger RNAs containing distinct 5'-untranslated region (UTR) sequence features. Here we identify 3'-UTR shortening of mRNAs as an additional molecular signature of mTORC1 activation and show that 3'-UTR shortening enhances the translation of specific mRNAs. Taking the advantage of RNA-Seq technology, by using genetic or chemical modulations of mTORC1 activity in cells or mouse tissues, we show that cellular mTORC1 activity is crucial for 3'-UTR shortening. Although long 3'-UTR-containing transcripts minimally contribute to translation, 3'-UTR-shortened transcripts efficiently form polysomes in the mTORC1-activated cells, leading to increased protein production. Strikingly, select E2 and E3 components of ubiquitin ligase complexes are enriched by this mechanism, resulting in elevated levels of protein ubiquitination on mTORC1 activation. Moreover, we developed a novel bioinformatics algorithm, IntMAP, which integrates RNA-Seq and PolyA Site (PAS)-Seq data for a comprehensive characterization of APA events. By applying IntMAP to the datasets, we further identified novel APA events that could otherwise not be identified by either profiling method alone. Several transcription factors including Cebpg (CCAAT/enhancer binding protein gamma, an ER stress response regulator) were among the newly discovered APA transcripts. This further links mTORC1-driven 3'-UTR shortening to the regulation of ER stress response in cells. Together, these findings identify previously uncharacterized roles for mTORC1 in the selective regulation of protein synthesis by modulating 3'-UTR length of mRNAs.

Introduction

As discussed in Chapter 1, alternative polyadenylation (APA) provides an added layer of complexity in regulating gene expression at the posttranscriptional level. Many publically available algorithms have been developed for the identification and quantification of 3'-UTR APA on a genome-wide scale[1]. If alternative poly(A) sites are known, 3'-UTR APA changes can be quantitatively profiled by comparing the RNA-seq read coverage before and after these sites. However, unannotated or novel alternative poly(A) sites cannot be accurately identified by analyzing RNA-seq data. To this end, various versions of 3'-end-targeted sequencing methods have been used to map out the location of poly(A) sites and to quantitate transcripts with different 3'-UTR lengths[1–14]. Yet, the quantitative power of 3'-end-targeted sequencing methods is limited by several factors such as false priming on A-rich regions or inconsistent efficiency in library preparation[6,15,16]. Regardless of these short-comings, these high-profiling technologies focusing on 3'-UTRs of mRNAs still provided high-resolution snap shots of alternatively polyadenylated mRNA isoforms in various tissues and cells across many species[2,14,17–20]. An important insight that emerged from these studies is that 3'-UTR length undergoes dynamic changes under pathogenic conditions such as cancer and in diverse biological processes such as cell proliferation, differentiation and development[2,17,21–24]. Although the information on alternative polyadenylation sites in transcriptomes across different species and tissues is rapidly accumulating, it is not clear what cellular mechanism(s) controls the switches between proximal and distal polyadenylation sites and how this process is regulated.

The mTORC1 pathway is crucial for regulating cell proliferation/growth and its dysregulation causes many human diseases[25]. Recent studies identified *cis*-acting elements in mRNAs such as 5'-terminal oligopyrimidine tract (TOP) or 5'-pyrimidine-rich translational element (PRTE) that render the association of a transcript with polysomes. mRNAs containing these elements in their 5'-UTRs encode proteins for cellular pathways including translation, cell invasion and metastasis, suggesting their relevance in cancer pathogenesis[26,27]. Recently, mTORC1 was also shown to play a role in the regulation of proteasome activity by upregulating a transcription factor Nrf-1[28]. Although these studies were mainly focusing on the role of mTOR in the synthesis of proteins, lipids and nucleic acids through transcriptional networks, whether mTORC1 is involved in other cellular processes by modulating gene expression at post-transcriptional level is relatively unclear.

Here, we used isogenic non-cancerous mouse embryonic fibroblast (MEF) cell lines to understand changes of molecular features on dysregulated activation of mTORC1. We employed RNA sequencing (RNA-seq) and two-dimensional liquid chromatography tandem mass spectrometry (2D LC–MS/MS) approaches to investigate the changes at high resolution and found an unexpected link between mTORC1 and ubiquitin-mediated proteolysis pathway through 3'-UTR shortening. These findings expand our understanding of mTORC1 to regulation of RNA processing and protein degradation pathways.

Furthermore, to better understand the dynamics of 3'-UTR length in the mTORC1-hyperactivated transcriptome and to improve upon the above-mentioned challenges of the current algorithms for studying 3'-UTR APA events, we developed an algorithm named IntMAP (Integrative Model for Alternative Polyadenylation). This method combines the

RNA-Seq data as well as the 3'-end-seq data to take advantage of the quantitative power of RNA-Seq and the precise mapping of poly(A) site usage by 3'-end-seq, allowing accurate profiling of 3'-UTR APA events. Interestingly, we were able to identify novel genes and pathways that are targeted by mTORC1-mediated 3'-UTR APA using IntMAP. Among these genes is the C/EBPγ (CCAAT/enhancer binding protein gamma) gene, a C/EBP family member that acts as a negative regulator of other C/EBP members' transcriptional activities by forming heterodimers[29,30]. Studies have shown that C/EBPγ promotes cell proliferation and inhibits senescence by interacting with C/EBPβ[31]. C/EBPγ is also crucial for cellular redox homeostasis and attenuates cellular stress by forming heterodimer with ATF4, a bZIP transcription factor[32]. Thus, with IntMAP, we further show that mTORC1 activation up-regulates the protein expression of C/EBPγ via 3'-UTR APA to enable cells to respond to ER stress, making an unexpected link between mTOR signaling and C/EBPγ pathway.

Materials and Methods

Cell lines and cell culture

Wild-type (WT) and Tsc1$^{-/-}$ MEF cells were obtained from Dr. Kwiatkowski at Harvard University. WT and Tsc1$^{-/-}$ MEFs were cultured in DMEM (Gibco, USA) supplemented with 10% (v/v) FBS (Gibco, USA), 100μg/ml streptomycin, and 100U/ml penicillin at 37˚C.

RNA-seq and alignments

To evaluate transcriptome-features under mTORC1 hyper-activation at nucleotide-wise resolution, we performed RNA-seq analysis of poly(A+) RNAs isolated from WT and TSC1$^{-/-}$ MEFs. In total 63,742,790 paired end-reads for WT and 74,251,891 paired-end reads for TSC1$^{-/-}$ MEFs were produced from Hi-Seq pipeline with length of 50 bps of each end. The short reads were aligned to the mm10 reference genome by TopHat [43] with up to 2 mismatches allowed. The unmapped reads were first trimmed to remove poly-A/T tails (repeats of [A/N]s or [T/N]s) from read ends/starts, and then, aligned to the reference genome. Note that we only retained the reads with at least 30 bps in both ends after trimming. Finally, 87.1 % of short reads from WT and 87.5 % of sequence reads from TSC1$^{-/-}$ MEFs were mapped to the reference genome by TopHat for APA analysis in the study.

APA analysis

To detect the potential alternative PAS of a transcript between WT and TSC1$^{-/-}$ MEFs, we evaluated candidate PAS motifs (AATAAA, ATTAAA, AGTAAA, CATAAA, TATAAA, GATAAA, ACTAAA, AATACA, AATATA, AAGAAA, AATAGA, AATGAA, TTTAAA, AAAATA, TATATA, AGATAA, ATTACA, AGAATA) [44,45] in the 3'UTR of the transcript by contrasting the short-read coverage up/down-stream of the site across WT and TSC1$^{-/-}$ samples with chi-square test. Specifically, we first scanned the 3'UTR of a transcript (by mm10 annotation) to identify PAS motifs as candidates of alternative PAS. For each candidate PAS, we calculated the mean coverage up-stream of the site (N and M) and down-stream of the site (n and m) with (N, n) denoting the coverage in WT and (M, m) denoting the coverage in TSC1$^{-/-}$. In the calculation, the up-stream region starts at the beginning of the last coding exon adjacent to the 3'UTR of the transcript and ends at the beginning of the PAS motif site. Then a canonical 2X2 chi-square test was applied to report a p-value for each candidate site. The candidate PAS with the most significant p-value <= 0.05 was considered for further analysis. Note that the chi-square test will report shortening events in both WT (when N/n>M/m) and TSC1$^{-/-}$ (when N/n<M/m). Out of the 5,160 transcripts, 846 (16.4%) show p-value <= 0.05 in TSC1$^{-/-}$ MEFs and 69 (1.3%) show p-value <= 0.05 in WT MEFs.

Scatter plot for differential expression and APA analysis

To select candidate transcripts with sufficient signal for reliable differential expression analysis and 3'UTR shortening identification, we first analyzed the short-read alignments of the RNA-seq data against mouse mm10 reference genome using Cufflink [12].

In the alignments, 14,378 and 14,175 transcripts are considered 'expressed' in WT and $TSC1^{-/-}$ cell lines, respectively, with a FPKM (Fragments Per Kilobase of transcript per Million mapped reads) cutoff=0.17. The union of the two sets gives 15,340 transcripts that are expressed in at least one of the cell lines. We further filtered out the transcripts with positional short-read coverage <= 25 in the entire 3'UTR in both cell lines. In addition, transcripts with 3'UTR overlapping exons in the strand in opposite direction were removed to avoid mingled short-read signals that might lead to inaccurate 3'UTR shortening identification. Finally, to allow precise PAS analysis, only transcripts with at least two occurrences of the eighteen PAS motifs in the 3'UTR are retained in the study. The entire pruning procedure left 5,160 transcripts for further analysis.

Measurement of relative shortening index (RSI)

A numerical presentation of 3'UTR shortening was developed by calculating the RSI of a given transcript. A relative expression of total or longer 3'UTR-containing transcripts was measured by normalizing to total amount of RNAs used in RT-qPCR analysis. The following equation was used to determine the RSI.

LI = [normalized expression of longer 3'UTR-containing transcript]/[normalized expression of total (long+short) transcript]

$RSI = -Log_2(LI/[LI \text{ in reference cell line}])$, so RSI=0 for a reference cell line.

If RSI>0 in a target cell line, then there is a 3'UTR shortening. If RSI<0 in a target cell line, then there is a 3'UTR lengthening. The RSI contains the information about the changes in the proportion of a longer 3'UTR-containing transcript in a given cellular

context compared to a reference cell. For example, a value of 1 in the RSI of a transcript indicates that the proportion of the longer 3'UTR containing transcript of the total (long+short) transcript decreases by 50% compared to that of the reference cell line, indicating an enrichment of 3'UTR-shortened transcript (i.e. 3'UTR shortening).

TMT labeling and proteomics analysis

Total cell lysates of WT and TSC1$^{-/-}$ MEFs were prepared by re-suspending the cells in 2% w/v SDS solution in 50 mM Tris pH 7.4 and heating to 95°C for 5 min, followed by 5 passes through 25 gauge needle. Cell debris was removed by centrifugation at 16,100 × g for 10 min. The resulting proteins were quantified by BCA assay and 100 µg digested using the FASP protocol [46]. Peptides were cleaned up using Waters Sep-Pak SPE cartridges (WAT054925 Waters Corp,, Milford, MA) and labeled using isobarically-labeled TMT 6-plex reagents (90061, Thermo Scientific) following the manufacturer's protocol. The labeled peptides were pooled and again cleaned up by Sep-Pak SPE. Peptides were fractionated by high-pH reversed phase chromatography into 40 fractions and recombined to give 12 fractions for analysis by LC-ESI-MS/MS [47]. Approximately 1.0-1.5 µg of each recombined fraction was analyzed on a Velos Orbitrap mass spectrometer using HCD dissociation as per Lin-Mosier *et al* [48]. Three repeats of independent experiments were carried out for statistical confidence.

Protein identification and quantification

MS data were searched against a Uniprot mouse database (downloaded July 2013) containing common protein contaminants plus all reversed sequences (102,424 total

sequences) using X!Tandem. Search parameters included semi-tryptic cleavage with one missed cleavage allowed, 0.01 Da monoisotopic precursor mass error, ± 0.5 Da product mass error, fixed modifications of carbamidomethylation at C, TMT label at K, and N-termini, variable modification of oxidation at M and P. All search results were loaded into Scaffold Q+ (Proteome Software Inc.) as a Mudpit sample and filtered to 99% protein confidence, 95% peptide confidence with 6 ppm precursor mass error and 1 non-tryptic terminus and 2 peptides per protein resulting in 2778 proteins identified at 0% protein and peptide FDR. The fold change ratios were exported from Scaffold Q+ for further bioinformatics analysis.

Western blotting and antibodies

The following antibodies were used in this study; anti-mTOR (cat. 2972, Cell Signaling, 1:500), anti-TSC1 (cat. 6935, Cell Signaling, 1:500), anti-ACTIN (cat. A2668, Sigma, 1:2,000), anti-Phospho-S6 ribosomal protein (cat. 2211, Cell Signaling, 1:1,000), anti-UBIQUITIN (cat. 13-1600, Life technologies, 1:500), anti-Ube2b (cat. 4944, Cell Signaling, 1:1,000), anti-Ube2i (cat. 4786, Cell Signaling, 1:1,000), anti-Rbx1 (cat. 11922, Cell Signaling, 1:500), and anti-tubulin (cat. sc-53646, Santa Cruz Biotechnology, 1:2,000). anti-hnRNP A1 (4B10 from Abcam), anti-CPSF6 (ab175237 from Abcam), anti-C/EBPγ (sc-517003, Santa Cruz Biotechnology), anti-Lamin A/C (#4777 from Cell Signaling).

Polysome isolation and analysis

Isolation of polysome fractions from total cell lysates using sucrose gradient was carried out as previously described[3]. Briefly, Cells were lysed in the polysome buffer (20 mM Tris/7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 ug/ml cycloheximide and 1% Triton X-100). Cell extracts were loaded onto sucrose gradient (5-45%). Fractionation was done by centrifugation at 36,000 rpm for 2 hrs at 4 C. Eleven fractions were collected for the analysis. The percentage of mRNAs in each fraction was calculated as described previously[18]. 5% (v/v) of total RNAs in each fraction were used for RT-qPCR. Relative enrichment of total or long transcripts was measured by normalizing to BC200 RNA, which were added to each fraction for recovery and normalization control. The signals of total transcripts for a gene in all fractions of $TSC1^{-/-}$ were combined and set as 100%. Then the percentage of a total transcript in each fraction was calculated and the distribution of total transcripts for a gene was shown by the fraction. In case of long 3'UTR transcripts, the signals of long transcripts in all fractions were combined and the relative amount of long transcripts was scaled down based on the relative differences between total and long transcript signals in input. The percentage and distribution of long transcripts in each fraction were calculated and presented. The same procedure was applied to the rest of transcripts (long and total in WT and $TSC1^{-/-}$ treated with Torin1) by considering total transcripts in $TSC1^{-/-}$ as 100%.

SiRNAs and siRNA-mediated knockdowns

Cells were transfected with siRNA oligos synthesized by IDT for each gene using RNAi Max (Thermo Fisher Scientific) according to the manufacturer's protocol. The

following sequences were used for the knockdown (KD) of *Cebpg* transcript: 5'- GAT ACA CTG CAA AGA GTA AAC CAG C-3' (Cebpg siRNA).

Real-time quantitative PCR (RT-qPCR) analysis and primer sequences

Total cellular RNAs were isolated by Trizol method according to manufacturer's protocol. Reverse transcription reaction was performed using Oligo-d(T) or random hexamer priming and superscript III (Thermo Fischer Scientific) according to the manufacturer's protocol. SYBR Green was used to detect and quantitate PCR products in real-time reactions via the comparative Ct method. We normalized the Ct values to total RNAs for quantitation of total or long transcripts. The PCR primers for quantitative analyses are as follows: Cth forward 5'-AGCAATCATGACCCATGCCT-3', Cth reverse 5'-CTCTAGGCCCACAGAAAGTCG-3', GPt2 forward 5'-GCAGCGAGAAGGCACTTAC-3', GPt2 reverse 5'-TGGAGCACGGTCTTCAGTTTA-3', Mthfd2 forward 5'-ATATCACTCCCGTCCCTGGT-3', Mthfd2 reverse 5'-TCTTAGCACTTTCTTTGCGGC-3', Slc1a5 forward 5'-GAGCCCGAATTGATCCAGGT-3', Slc1a5 reverse 5'-TGGTACTGTTTCAGGAGGGGA-3', Slc7a11 forward 5'-CACCGGGGTCGGTTTTCTTA-3', Slc7a11 reverse 5'-TCGTCTGAACCACTTGGGTTT-3', Mtdh total forward 5'-GACCAAGTCTGAAACTAACTGGGA-3', Mtdh total reverse 5'-TTCACGTTTCCCGTCTGG-3', Mtdh long forward 5'-TGTCAACTAGGAAAGCTAAAATGGT-3', Mtdh long reverse 5'-

GAGGAAAGCTGTCCATTAATAAGGC-3', Appl1 total forward 5'-TCATTTCCCTGGGATGTGGC-3', Appl1 total reverse 5'-GCTGAAGCACACTACTGTAAAGC-3', Appl1 long forward 5'-TTTCTGTGTAGTCCTGGGAGC-3', Appl1 long reverse 5'-GACAGAGGCAAGCGGGTATG-3', Cebpg total forward 5'-ACACTGCAAAGAGTAAACCAGC-3', Cebpg total reverse 5'-GTGCGCATGCTCAAGAAACA-3', Cebpg long forward 5'-TGTAGAGTGCTCCTGATGCC-3', Cebpg long reverse 5'-GGCAGATCTGATTAGCTGTGGA-3'.

Short read alignments and peak identification

To better evaluate the transcriptome features, particularly APA, under mTORC1-hyperactivation at single-nucleotide resolution, we further performed Poly (A)-site sequencing (PAS-Seq) analyses of poly(A+) RNAs isolated from WT and $Tsc1^{-/-}$ MEFs. In the PAS-Seq analysis, the reads from WT and $Tsc1^{-/-}$ were preprocessed to trim A's off the 3'-ends and then filtered by removing the reads of low-quality 3'-ends (Phred score < 30) and shorter than 25 bps. The remaining reads were aligned to the mm10 reference genome by Bowtie without allowing any mismatches. In total, 6,186,893 aligned reads were aligned for WT and 5,382,111 reads were aligned for $Tsc1^{-/-}$. All aligned reads from PAS-Seq were pooled together in order to identify peaks and the corresponding cleavage sites in the reference genome by the read coverage signals. In each read alignment 'hill', the location with the highest read coverage between two zero coverage positions was

considered as the peak of the 'hill'. The 3'-end of the peak is chosen as the potential corresponding cleavage sites where the read coverage at the peak quantifies the cleavage at the site. To remove false positives, only peaks with at least four supporting reads were analyzed.

IntMAP algorithm

To detect alternative PAS in the 3'-UTR of a gene, we propose an integrative model for alternative polyadenylation, IntMAP, by combining the RNA-seq read alignments and PAS-seq peak calling. IntMAP is built on a constrained probabilistic model with the probabilistic modeling of the RNA-seq read mapping uncertainty for estimating the abundance of all the short and long transcripts of a gene and the constraint on the abundance by the PAS peak callings.

Let $\boldsymbol{T}$ denote the set of the transcripts with different 3'UTR lengths in a gene with T_i being the *i*th transcript in $\boldsymbol{T}$, and a set of reads $\boldsymbol{r}$ aligned to the 3'UTRs of the gene. Note that the complete list of the transcripts in a gene is combined from all the transcript variants with different cleavage sites predicted by the PAS-seq data or annotated in the reference genome. The probability of a read being generated by the transcript in $\boldsymbol{T}$ is modeled by a categorical distribution specified by parameter p_i, where $\sum_{i=1}^{|\boldsymbol{T}|} p_i = 1$ and $0 \leq p_i \leq 1$. We consider the likelihood that each read in $\boldsymbol{r}$ is sampled from one of the 3'-UTRs to which the read aligns. Specifically, for each read r_j aligned to the 3'-UTRs in T_i, the probability of obtaining r_j by sampling from T_i, namely $Pr(r_j|T_i)$ is $q_{ijk} = \frac{1}{l_i - l_r + 1}$ (46–48), where l_i and l_r are the length of the 3'-UTR in T_i and the length of the read, respectively. Assuming

each read is independently sampled from one transcript, the uncommitted likelihood function to estimate the parameters $\boldsymbol{P}$ from the observed read alignments against the 3'-UTRs in $\boldsymbol{T}$ is

$$\mathcal{L}(\boldsymbol{P};\boldsymbol{r}) = P\,r(\boldsymbol{r}|\boldsymbol{P}) = \prod_{j=1}^{|\boldsymbol{r}|} P\,r\left(r_j\middle|\boldsymbol{P}\right) = \prod_{j=1}^{|\boldsymbol{r}|}\sum_{i=1}^{|\boldsymbol{T}|} P\,r(T_i|\boldsymbol{P})\,P\,r\left(r_j\middle|T_i\right) =$$

$$\prod_{j=1}^{|\boldsymbol{r}|}\sum_{i=1}^{|\boldsymbol{T}|} p_i q_{ij},$$

where $\boldsymbol{P}$ is the probability of a read being generated by the transcript $\boldsymbol{T}$, specifically, $\boldsymbol{P} = [p_1, \ldots, p_{|\mathbf{T}|}]$. The likelihood function is concave and an Expectation Maximization (EM) algorithm is applied to obtain the optimal $\boldsymbol{P}$. With $\boldsymbol{P}$ estimated, the abundance of the *i*th transcript is derived as $\frac{p_i|\boldsymbol{r}|}{l_i}$, where l_i is the length of transcript T_i.

In the PAS-Seq analysis, the read coverage (the height of the peak in Fig. 6A) also provides the measure of abundances of the transcript variants at 3'-UTR. IntMAP adopts the following constrained log-likelihood function,

$$\mathcal{L}_{pen}(\boldsymbol{P};\boldsymbol{r}) = \log(\mathcal{L}(\boldsymbol{P};\boldsymbol{r})) - \lambda \left\|\frac{\boldsymbol{P}|\boldsymbol{r}|}{\boldsymbol{l}} - \alpha\boldsymbol{E}\right\|^2, \quad \text{(a)}$$

where the vector $\boldsymbol{E}$ contains the expression values for the transcripts provided by PAS-seq, $\frac{\boldsymbol{P}|\boldsymbol{r}|}{\boldsymbol{l}} = (\frac{p_1|\boldsymbol{r}|}{l_1}, \ldots, \frac{p_i|\boldsymbol{r}|}{l_i}, \ldots, \frac{p_{|T|}|\boldsymbol{r}|}{l_{|T|}})$ represents the transcripts' expressions and α is a scaling factor between the expression values learned from RNA-seq data and PAS-seq data. There are two terms in the penalized log-likelihood function (equation (a)). The first term is the likelihood of the observed the RNA-seq read alignment as in the original probabilistic model. The second term, $\left\|\frac{\boldsymbol{P}|\boldsymbol{r}|}{\boldsymbol{l}} - \alpha\boldsymbol{E}\right\|^2$, encourages the consistency between the transcript expressions learned from RNA-seq data and PAS-seq data. The parameter λ balances the

two terms, where larger λ weights PAS-seq data more. By subtracting the second convex term, the penalized log-likelihood function is still concave and a similar EM algorithm is applied to obtain the optimal $\boldsymbol{P}$ with the CVX package. The scaling factor $\alpha = \frac{\boldsymbol{P}|\boldsymbol{r}|}{\boldsymbol{lE}}$ is updated with current $\boldsymbol{P}$ in each iteration in the EM algorithm. Then the chi-squared test is applied to the abundances of transcripts with different 3'-UTR lengths in WT and Tsc1$^{-/-}$ to detect 3'-UTR APA events.

Accession code

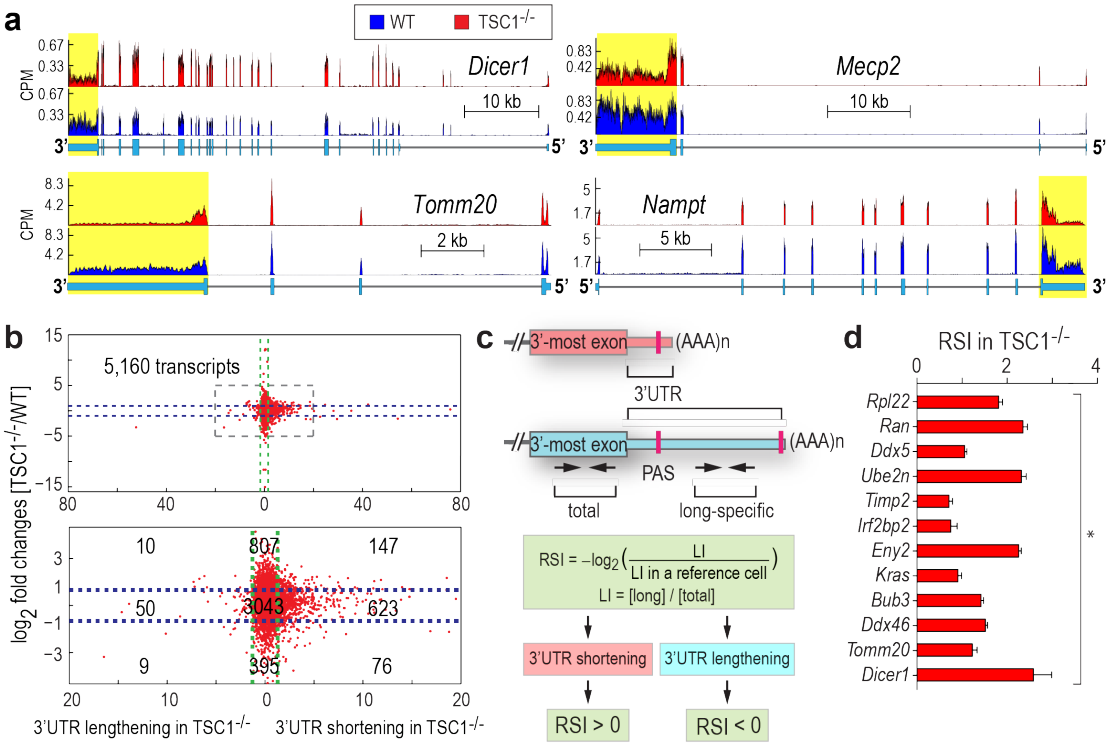
The accession numbers for the RNA-seq data in this chapter are SRP056624 and SRP056624. The accession number for the PAS-Seq data in this chapter is PRJNA436720. The accession number for the proteome data in this chapter is PXD002006.

Results

mTORC1 Activation Leads to 3'-UTR Shortening of mRNAs

To explore mTORC1 function in gene expression at a single nucleotide resolution, we performed RNA-seq experiments using Tsc1 knockout ($TSC1^{-/-}$) and wild-type (WT) MEF cells (Supplementary Table 1)[33]. Knockout of Tsc1, a negative regulator of mTORC1, leads to uncontrolled mTORC1 hyperactivation compared with WT[33,34]. One of the striking features in our data set was that many transcripts in $TSC1^{-/-}$ showed an abrupt signal drop only for a segment of the 3'-most exon of an annotated gene compared with WT (Fig. 1A and Supplementary Fig. 1A). For example, the read signal for *Dicer1* in $TSC1^{-/-}$ dropped after the termination codon in the 3'-most exon, although the signal from upstream exons increased (Fig. 1A). In some cases, upstream exons showed either similar (for example, *Mecp2* and *Tomm20*) or decreased (for example, *Anxa7* and *Timp2*) signal, although we observed the same pattern of signal drop in the 3'-most exon from $TSC1^{-/-}$ (Fig. 1A and Supplementary Fig. 1A). Further sequence analysis revealed that canonical or non-canonical PAS(s) exists around the regions showing the signal drop. This indicates that the synthesis of these transcripts terminated early in the 3'-most exon using the proximal PASs for polyadenylation, suggesting a predominant production of mRNA isoforms with a shorter 3'-UTR in the mTOR-activated transcriptome (Fig. 1A and Supplementary Fig. 1A, yellow box, and Supplementary Table 2). For some transcripts such as *Tomm20* and *Nampt*, 3'-UTR-shortened transcripts were already present in WT where the mTOR activity is low but not entirely absent (Fig. 1A and Supplementary Fig. 1A). These 3'-UTR-shortened transcripts increased significantly in $TSC1^{-/-}$ (Fig. 1A and Supplementary Fig. 1A),

indicating that individual transcripts differ in the regulation of their 3'-UTR length in response to cellular mTORC1 activity. As the signals from upstream exons reflecting the amount of transcripts varied among the 3'-UTR-shortened transcripts, we examined whether 3'-UTR shortening in the mTORC1-activated transcriptome correlates to differential gene expression. To this end, we enriched 5,160 transcripts in our data set that are eligible for combined analysis of 3'-UTR shortening and differential expression. Next, each transcript was plotted by fold changes in the differential gene expression (y axis in Fig. 1B) and the significance of 3'-UTR shortening (x axis in Fig. 1B). This approach identified 846 3'-UTR-shortened transcripts (about 16.4%) out of 5,160 transcripts in $TSC1^{-/-}$ (Fig. 1B). Although 26.3% (223/846) of the 3'-UTR-shortened transcripts either increased (147/846) or decreased (76/846) their expression level, a significant proportion (73.7%) of them in the mTOR-activated transcriptome remained unchanged (Fig. 1B), indicating no strong correlation between the differential gene expression and the 3'-UTR shortening in the mTORC1-activated transcriptome. Of note, only a small percentage (1.3%) of transcripts showed 3'-UTR shortening in WT over $TSC1^{-/-}$ MEFs. To confirm the RNA-seq data, we developed a method to determine 3'-UTR shortening or lengthening by calculating 'relative shortening index (RSI)' (Fig. 1C). Twelve genes, covering a wide range of P-values, were randomly selected from the 3'-UTR shortening data set; all showed the RSI >0 in $TSCI^{-/-}$ (Fig. 1D and see also Supplementary Fig. 1D–E for alternative presentations of the data using different experimental and calculation methods), validating our RNA-seq data analysis. Together, these data strongly suggest that mTOR activation in



cells leads to a preferred usage of proximal PAS in the 3'-most exon of mRNAs and results in transcriptome-wide 3'-UTR shortening.

Figure 1. mTORC1 activation leads to genome-wide 3'-UTR shortening. (A) RNA-seq reads from WT and TSC1$^{-/-}$ are aligned to mouse genome mm10 RefSeq. Representative examples of transcripts with 3'-UTR shortening are presented. Annotated gene structures are at the bottom of the alignment. The yellow boxes highlight the aligned reads in 3'-UTRs. (B) Scatter plot of RNA-seq data. Red dots represent individual transcripts in the analysis. Horizontal blue-dashed lines represent the cutoff values for twofold changes in differential gene expression. Vertical green-dashed lines represent the cutoff values for −log10 (P-value) of 3'-UTR shortening (1.3 corresponds to P=0.05) in TSC1$^{-/-}$ and WT, which was determined by χ^2-test. (C) A schematic presenting primer

sets for RT–qPCR and the RSI determination. Pairs of primers were used to detect a total (short+long) or a long-specific transcript. The RSI was calculated to determine the 3'-UTR shortening in a target cell line by RT–qPCR. (D) Validation of RNA-seq data. Error bars represent s.e. from three repeats of experiments. Student's t-tests are done for statistical significance. *P<0.0025.

3'-UTR Shortening is a Downstream Target of mTORC1

To determine whether 3'-UTR shortening due to APA is a previously uncharacterized cellular target downstream of mTOR pathway, we established a stable mTOR knockdown cell line $TSC1^{-/-}$ MEFs ($TSC1^{-/-}$ mTOR kd; Supplementary Fig. 2A). The tested transcripts showed the RSI <0 in $TSC1^{-/-}$ mTOR kd MEFs as compared with a control knockdown cell line, indicating the enrichment of 3'-UTR-lengthened transcripts in mTOR-deficient cells (Fig. 2A and Supplementary Fig. 2B), thus supporting the idea that mTOR functions in 3'-UTR length regulation. Consistently, the same results were observed in a HEK293 stable cell line with mTOR knockdown or in a bladder cancer cell line HCV29 (where TSC1 is mutated and as a consequence, mTORC1 activity is upregulated) with TSC1 added back (HCV29 Add back)[35,36]. These results suggest that the function of mTORC1 in 3'-UTR shortening is a general phenomenon and evolutionarily conserved between human and mouse (Supplementary Fig. 2C–E).

Proliferative cells are known to carry short 3'-UTRs in their transcriptome and terminally differentiated tissues are known to produce transcripts with long 3'-UTRs[2,17,18]. We asked whether mTORC1 activation could be an underlying reason that explains these observations. We used a mouse model with skeletal muscle-specific Tsc1 knockout (SM $TSC1^{-/-}$) to render mTORC1 hyperactivated only in skeletal muscle (Fig. 2B and Supplementary Fig. 2F)[37] and compared the changes in 3'-UTR length. Control brain tissues from both WT and SM $TSC1^{-/-}$ did not show significant changes in 3'-UTR length (Fig. 2B). In contrast, the RSI of the tested transcripts in the skeletal muscle from SM $TSC1^{-/-}$ mice was >0 as compared with that from WT, providing evidence that the

mTORC1 activation is sufficient to drive 3'-UTR shortening in terminally differentiated skeletal muscles (Fig. 2B and Supplementary Fig. 2M).

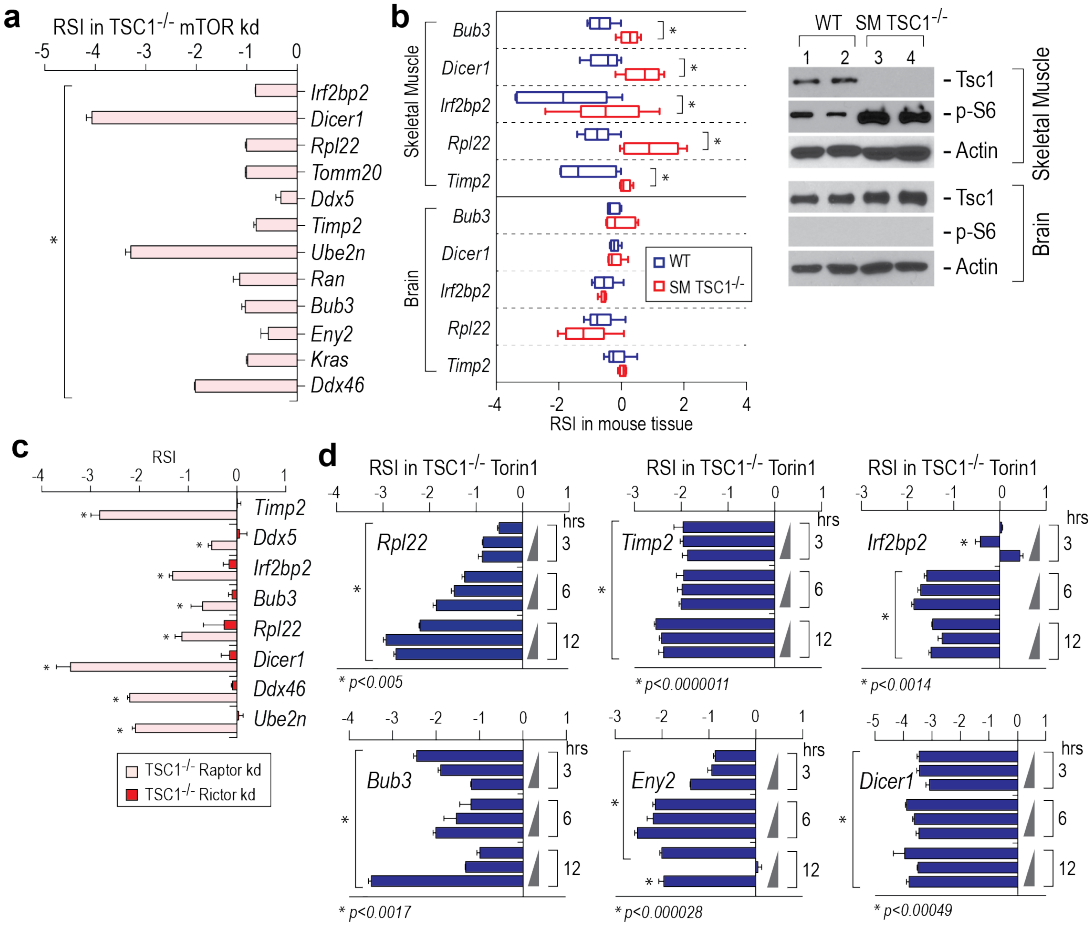


Figure 2. 3'-UTR shortening is a downstream target of mTORC1. (A) mTOR knockdown (kd) in TSC1$^{-/-}$ recovers the 3'-UTR length. The RSI was measured using total RNAs isolated from TSC1$^{-/-}$ MEFs with mTOR kd. *P<0.015. (B) Activation of mTORC1 in terminally differentiated skeletal muscle leads to 3'-UTR shortening. Total RNAs from skeletal muscles in WT or skeletal muscle-specific knockout of Tsc1 (SM TSC1$^{-/-}$) mice were used for the RSI measurement. The brain was used as an additional tissue control. Western blotting on tissue extracts from two randomly chosen mice was

done. p-S6 denotes phosphorylated S6, a downstream target of activated mTOR kinase. *P<0.016. (C) mTORC1 but not mTORC2 is crucial for 3'-UTR shortening. mTORC1 or mTORC2 was specifically deactivated by targeting Raptor or Rictor, respectively, using short hairpin RNAs in $TSC1^{-/-}$ MEFs. The RSI was measured using RT–qPCR. *P<0.05. (D) A selective inhibitor of mTOR, Torin1, alters 3'-UTR length in mRNAs. $TSC1^{-/-}$ MEFs were treated with Torin1 at various doses (10, 50 and 250 nM, presented as incremental triangles) and time courses (3, 6 and 12 h). Changes in the 3'-UTR length were determined by measuring the RSI. *The conditions that accumulate the long 3'-UTR-containing transcripts with statistical significance.

To address which mTOR complex regulates 3'-UTR shortening, we established stable cell lines using short hairpin RNA that specifically knocks down Raptor (a component of mTORC1) or Rictor (a component of mTORC2) in $TSC1^{-/-}$ MEFs (Supplementary Fig. 2G). The knockdown of Raptor but not Rictor resulted in the RSI <0 when compared with control knockdown cells, suggesting that mTORC1 plays an important role in 3'-UTR shortening (Fig. 2C and Supplementary Fig. 2F). mTOR is a key therapeutic target for many human disease treatments[25] and several versions of selective mTOR inhibitors have been developed including Torin1[38]. Thus, we examined whether pharmacological inhibition of mTOR via Torin1 could modulate APA. Torin1 treatment of $TSC1^{-/-}$ or HEK293 cells resulted in the RSI <0 in most of the tested transcripts in all conditions and the lengthening of 3'-UTR occurred as early as 3 h after the treatment, indicating that the cellular APA pattern changes drastically on the inhibition of mTOR (Fig. 2D and Supplementary Fig. 2I–J, N). In contrast, the treatment of rapamycin, which does not fully inhibit mTORC1[27,39], showed minimal effect, if any, on 3'-UTR lengthening (Supplementary Fig. 2J–L). As Torin1 arrests cells in G1/S phase[38,39], the lengthening of the 3'-UTR on Torin1 treatment could come from the inhibition of cell proliferation. To rule out this possibility, we used another class of small molecules such as hydroxyurea or mimosine, which arrest cells in G1/S without downregulating mTORC1 activity (Supplementary Fig. 2O)[40,41]. Hydroxyurea or mimosine treatment for 12 h in WT MEFs showed the RSI >0, indicating that 3'-UTR shortening occurred even though cells stop proliferating (Supplementary Fig. 2P). However, the same treatment of WT MEFs with mTOR knockdown abolished 3'-UTR shortening (Supplementary Fig. 2Q). Thus, the 3'-

UTR lengthening after Torin1 treatment is not caused by the inhibition of cell proliferation but rather from the inactivation of mTORC1 (Supplementary Fig. 2R–S). Taken together, we conclude that the mTORC1 pathway is an upstream regulator for APA process and determines the 3'-UTR length in the transcriptome independent of cell proliferation status.

3'-UTR Shortening Activates Ubiquitin-Mediated Proteolysis

Activation of mTORC1 increases global protein synthesis by controlling multiple downstream events such as ribosome biogenesis and cap-dependent translation initiation and elongation[25,42]. Especially, mTORC1 promotes the translation of a subset of mRNAs carrying 5'-UTR sequences such as 5'TOP, 5'TOP-like motif and 5'PRTE[26,27]. Similar to 5'-UTR, 3'-UTR in mRNA also plays an important role in the regulation of gene expression. In particular, 3'-UTR shortening in a transcript has been shown to increase protein production[2,17,43,44]. Therefore, we asked whether the mTORC1-activated 3'-UTR shortening contributes to mTORC1-mediated upregulation of protein synthesis and influences mTORC1-related biology. To this end, we first conducted quantitative proteomic studies using tandem mass tag (TMT)-labelled total cell lysates prepared from WT and $TSC1^{-/-}$ MEFs, to quantitatively profile the changes in the cellular proteome due to mTORC1 activation. We identified a total of 2,754 proteins that were found in either cell line by two or more unique peptides via 2D LC–MS/MS (Supplementary Table 3). The KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis on the catalogue of proteins with >20% increase (1,014 proteins) in abundance ($TSC1^{-/-}$ compared with WT) shows that multiple cellular pathways are activated in $TSC1^{-/-}$ (Fig. 3A). We also

performed the KEGG pathway analysis using the catalogue of 846 3'-UTR-shortened transcripts and the differentially expressed transcripts in TSC1$^{-/-}$ MEFs (Supplementary Fig. 3A and Supplementary Table 4). By comparing the enriched pathways from these three data sets, we found that multiple pathways in the mTOR-activated proteome such as spliceosome (mmu03040) and RNA degradation (mmu03018) are upregulated by both 3'-UTR shortening and differential gene expression, whereas other enriched pathways such as DNA replication (mmu03030) and pyrimidine/purine metabolism (mmu00240/mmu00230) are attributable solely to the differential gene expression (Fig. 3A and Supplementary Fig. 3A–B).

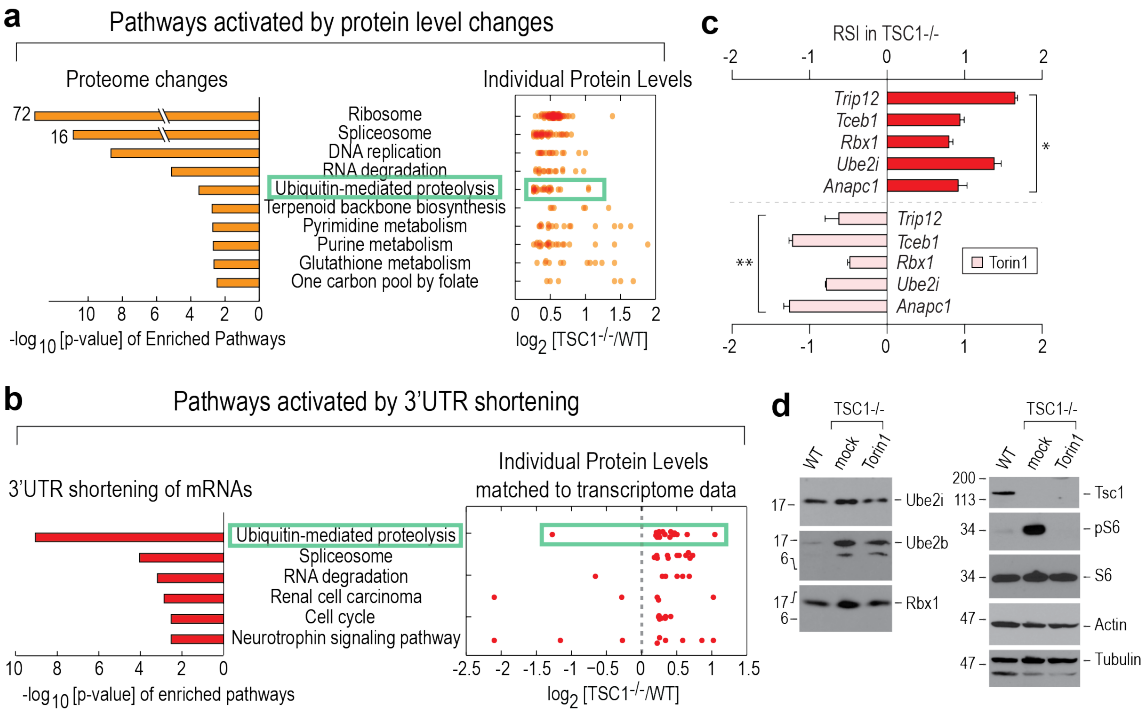


Figure 3. 3'-UTR shortening due to mTORC1 activation targets specific cellular pathways including ubiquitin-mediated proteolysis. (A) Analysis of mTORC1-activated proteome. The KEGG pathway enrichment analysis was performed using the

catalogue of identified proteins from 2D LC–MS/MS. The left panel shows the enriched pathways in −log10(P-value). The right box shows the distribution of individual proteins in each KEGG pathway index shown in the left panel. Proteins showing more than 1.2 fold changes in $TSC1^{-/-}$ compared with WT MEFs are plotted. Ubiquitin-mediated proteolysis pathway is marked with a light green box. (B) Analysis of enriched KEGG pathways by 3'-UTR shortening in $TSC1^{-/-}$ MEFs. The mTORC1-activated transcriptome is described in Fig. 1. The KEGG pathways enriched in $TSC1^{-/-}$ MEFs by 3'-UTR shortening are shown in –log10(P-value). Fold changes of individual proteins in each pathway index are plotted in the right box. Ubiquitin-mediated proteolysis pathway is marked with a light green box. (C) The RSI was measured for the transcripts enriched in ubiquitin-mediated proteolysis pathway in TSC1−/− MEFs. The lengthening of 3'-UTR in $TSC1^{-/-}$ MEFs treated with Torin1 at 50 nM for 24 h was shown by the RSI. *$P<1.5 \times 10-6$, **$P<6.3 \times 10-6$. (d) Western blot analysis of E2 and E3 enzymes showing the 3'-UTR shortening in the RNA-seq experiments. Cells were treated with Torin1 for 24 h at 50 nM for 3'-UTR lengthening. pS6 denotes phosphorylated S6.

Intriguingly, among those enriched pathways in the quantitative proteome analysis, ribosome- (mmu03010) and ubiquitin-mediated proteolysis pathways (mmu04120) did not appear in the differential gene expression data set (Fig. 3A and Supplementary Fig. 3B–C). This indicates that these two pathways are most likely to be activated by other mTORC1-mediated regulatory mechanisms rather than transcriptional regulation. It is known that the ribosome pathway is activated by mTORC1 through the translational regulation of mTORC1-responsive 5'-UTR cis-elements such as 5'TOP and 5'PRTE in the mRNAs[26,27,45]. On the other hand, mTORC1-responsive 5'-UTR sequence elements do not exist in most transcripts from the ubiquitin-mediated proteolysis pathway, supporting the idea that 3'-UTR shortening could be an explanation for the activation of ubiquitin-mediated proteolysis pathway in the mTORC1-activated proteome (the light green box in Fig. 3A–B and Supplementary Fig. 3D). All transcripts from the ubiquitin-mediated proteolysis pathway containing short 3'-UTR encode the components of E2 ubiquitin-conjugating enzymes and E3 ubiquitin ligases such as Anapc1, Rbx1, Trip12, Ube2i and Tceb1. Consistent with our findings in this study, these transcripts carry an mTORC1-dependent short 3'-UTR in $TSC1^{-/-}$ MEFs as determined by the RSI in the presence or the absence of Torin1 treatment (Fig. 3C). As shown by western blotting (Fig. 3D), the expression of Ube2i, Ube2b and Rbx1 proteins matched the progression of 3'-UTR shortening in these transcripts, supporting our conclusions from quantitative proteomics studies and 3'-UTR-shortening analysis. Together, these results suggest that 3'-UTR shortening by mTORC1 activation plays an important role in altering gene expression. These results also identify the ubiquitin-mediated proteolysis pathway as an additional

cellular target of mTORC1. Thus, mTORC1-driven 3'-UTR shortening might explain part of cellular phenotypic changes in TSC1$^{-/-}$ over WT MEFs.

3'-UTR Shortening Contributes to mTORC1-Promoted Translation

To examine whether 3'-UTR shortening in the group of transcripts of the ubiquitin-mediated proteolysis pathway is the major contributor to the increase in protein production, we looked at the effect of mTORC1-mediated 3'-UTR shortening on the association with actively translating ribosomes using polysome fractionation and reverse transcription–quantitative PCR (RT–qPCR). We fractionated cytoplasmic extracts on sucrose gradients to separate polysomes from 40S, 60S and 80S (monosome) ribosomes (Fig. 4A). Using RT–qPCR on RNAs purified from each fraction, we measured the percentage of total (long+short) and long 3'-UTR transcripts in each tested cell line. Interestingly, an overall distribution of long 3'-UTR transcripts across the separated fractions was polarized into light (fractions 2 and 3) and heavy fractions (fractions 6, 7 and 8; *Rbx1*, *Ube2i*, *Anapc1* and *Ube2n*) or was biased to light fractions (fractions 2 and 3; *Trip12* and *Tceb1*) (Fig. 4B and Supplementary Fig. 4A). In cases of *Rpl22* and *Bub3*, long 3'-UTR transcripts were distributed almost evenly across the separated fractions (Supplementary Fig. 4A). All these results indicate that a significant portion of long 3'-UTR transcripts are not translated in cells or do not associate with ribosomes efficiently. In contrast, the overall distribution of total (long+short) transcripts was biased towards heavy polysome fractions (fractions 7 through 11) in all cell lines (Fig. 4B and Supplementary Fig. 4A). The association of total transcripts with polysomes was more prominent in TSC1$^{-/-}$ compared with WT or TSC1$^{-/-}$

treated with Torin1. As total transcripts contain both long and short 3'-UTR transcripts, we compared the distribution of short 3'-UTR transcripts with that of total. To do this, we first conducted rapid amplification of complementary DNA 3' ends to locate the precise proximal poly(A) sites of short 3'-UTR transcripts (Supplementary Fig. 4E) and used oligo d(T)30 primers with 3'-end degenerated C (for *Rbx1*) or G (for *Anapc1*) for amplification. RT–PCR and agarose gel-based detection of total, short and long 3'-UTR transcripts for Rbx1 and Anapc1 in each fraction from $TSC1^{-/-}$ revealed that most short 3'-UTR transcripts formed polysomes, whereas a significant fraction of long 3'-UTR transcripts do not (Supplementary Fig. 4B). These results are consistent with the observations made by RT–qPCR. Importantly, the distribution of total transcripts mirrors short 3'-UTR transcripts more than long 3'-UTR transcripts (Supplementary Fig. 4B). Thus, highly enriched total transcript signals in the polysome fractions of $TSC1^{-/-}$ are likely to be due to the presence of short 3'-UTR transcripts. Considering the biggest differences between total and long transcript signals in the polysome fractions, short 3'-UTR transcripts are the major form of mRNAs in polysomes and determine protein synthesis in $TSC1^{-/-}$ MEFs. Contrary to this, in WT MEFs the overall total transcript signals from the polysome fractions are reduced and the proportion of long 3'-UTR transcripts is close to that of total transcripts in polysomes (Fig. 4B and Supplementary Fig. 4A). Intriguingly, the distribution and percentage of long 3'-UTR-containing transcripts showed minimal, if any, changes in different cellular contexts (WT, $TSC1^{-/-}$ and $TSC1^{-/-}$ treated with Torin1; Fig. 4B and Supplementary Fig. 4A). These results propose that long 3'-UTR transcripts participate in the basal level of protein synthesis regardless of cellular contexts and play a

bigger role in translation in WT compared with TSC1$^{-/-}$. Together, these results provide evidence that short 3'-UTR transcripts define part of proteome changes on mTORC1 activation. We also noted that in the presence of 50 nM Torin1 for 24 h, formation of polysomes with total transcripts decreased, while formation of polysomes with long 3'-UTR transcripts stayed constant (Fig. 4B and Supplementary Fig. 4A). This suggests that the formation of polysomes in TSC1$^{-/-}$ is largely dependent on the level of short 3'-UTR transcripts.

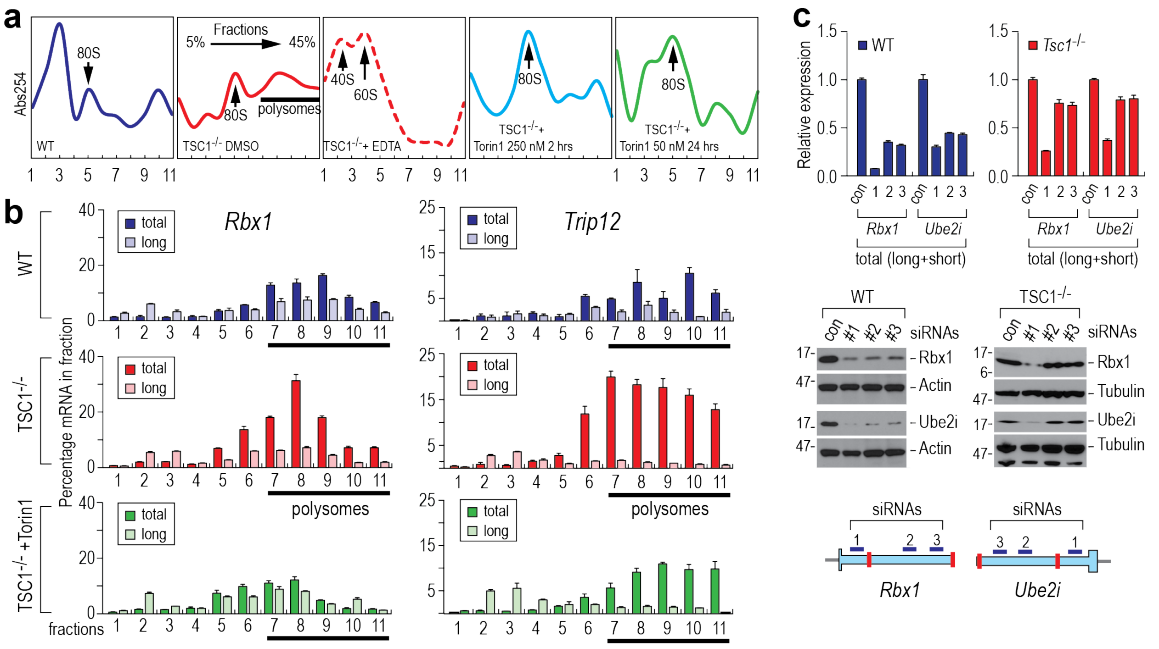


Figure 4. 3'-UTR shortening is an mTORC1-activated molecular signature for protein synthesis. (A) Polysome fractionation using 5%–45% sucrose gradient. Cytoplasmic extracts were prepared from WT, TSC1$^{-/-}$ MEFs in the presence or the absence of 50 mM EDTA, or in the presence of 50 nM or 250 nM Torin1. Eleven fractions were collected and subjected to absorbance measurement at 254 nm. The arrows mark 40S, 60S subunits and 80S monosomes. The black thick line indicates polysome fractions

(fractions 7 through 11). (B) Polysome analyses of selected total (long+short) and long transcripts in each fraction from sucrose gradient. Percentage of total and long transcripts for *Rbx1* or *Trip12* in each fraction was measured using RT–qPCR. (C) Long 3'-UTR-containing transcripts minimally contribute to protein synthesis on mTORC1 activation. siRNAs targeting three distinct regions in the 3'-UTR of *Rbx1* and *Ube2i* mRNAs were designed (right bottom panel). siRNA#1 targets the region between the termination codon and the first PAS. siRNAs#2 and 3 specifically targets long 3'-UTR-containing transcripts only. Knockdown of total or long transcripts was carried out using these siRNAs in WT and TSC1$^{-/-}$ MEFs. The amounts of remaining *Rbx1* and *Ube2i* total (long+short) transcripts were measured using RT–qPCR (left panel). Western blotting was performed using the cell lysates from RNA interference knockdown experiments (right top panel).

To further investigate the effects of 3'-UTR shortening on protein synthesis in WT and TSC1$^{-/-}$ MEFs, we designed small interfering RNAs (siRNAs) that target either total (siRNA 1) or long (siRNAs 2 and 3) 3'-UTR transcripts of Rbx1 and Ube2i specifically and measured their effects on gene expression by RT–qPCR and western blotting (Fig. 4C). Notably, siRNAs 2 and 3 minimally decreased total (long+short) mRNAs of Rbx1 and Ube2i in TSC1$^{-/-}$ (~20%–25% knockdown), whereas siRNA 1 could knock down those total mRNAs efficiently (~60%–70%). On the other hand, siRNAs 2 and 3 increased the knockdown efficiency dramatically in WT (50–70%) compared with TSC1$^{-/-}$ (Fig. 4C). As all siRNAs used in the experiment knocked down long 3'-UTR transcripts with high efficiency (70%–95%; Supplementary Fig. 4F), the discrepancy in the efficiency of siRNAs 2 and 3 in WT and TSC1$^{-/-}$ is likely to be due to the differences in the relative amounts of long 3′-UTR transcripts in these cells. The siRNA 1 significantly affected the protein levels of both Rbx1 and Ube2i in WT and TSC1$^{-/-}$. In contrast, the siRNAs 2 and 3 affected target protein levels only in WT (Fig. 4C). Taken together with polysome profiling experiments, these results suggest that most of the polysomes are formed with short 3'-UTR transcripts, which plays a major role in protein production on mTORC1 activation. Meanwhile, relatively smaller portion of the polysomes formed with long 3'-UTR transcripts, which minimally contributes to protein production. Thus, we conclude that the overexpression of selected E2 and E3 ligase complexes is due to mTORC1-promoted 3'-UTR shortening of those mRNAs.

A More Comprehensive Profiling of 3'-UTR Length Dynamics by an Integration Model, IntMAP

It has been shown that the outcome of APA analysis on 3'-UTR could vary by the choice of profiling methods[6]. Thus, to investigate whether other APA profiling methods could expand our understanding of mTORC1-controlled biological pathways, we used the PAS-Seq method to map out experimentally proven polyadenylation sites and to quantitate APA events in the WT and $TSC1^{-/-}$ MEF transcriptomes. Compared to 846 short 3'-UTR transcripts identified by the RNA-Seq method, we found 843 short 3'-UTR transcripts in the $TSC1^{-/-}$ transcriptome using the PAS-Seq method (Fig. 5A and Supplementary Table 5). Consistent with the previous report[6], our APA profiling using two separate methods also showed a limited overlap of 305 genes (36%) (Fig. 5A). Although the number of overlapping short 3'-UTR transcripts between the two datasets is not considerably large, many KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways including protein processing in the endoplasmic reticulum (mmu04141) and ubiquitin-mediated proteolysis (mmu04120) exist in both datasets with a significant P-value (Fig. 5B). In comparison, KEGG pathways that are unique to either dataset tend to have P-values that are less significant. (Fig. 5B and Supplementary Table 5). By pooling the two datasets together, we were able to identify 3'-UTR APA events in 1,384 genes in the $TSC1^{-/-}$ transcriptome (Fig. 5A). Interestingly, the combined data enriched distinct KEGG pathways such as RNA transport (mmu03013) and adherens junction (mmu04520) which did not appear in either dataset (Fig. 5B and Supplementary Table 5). Thus, two profiling methods increase the

total number of genes with 3'-UTR APA events and broaden the list of biological pathways that are modulated by 3'-UTR APA in TSC1$^{-/-}$ MEFs.

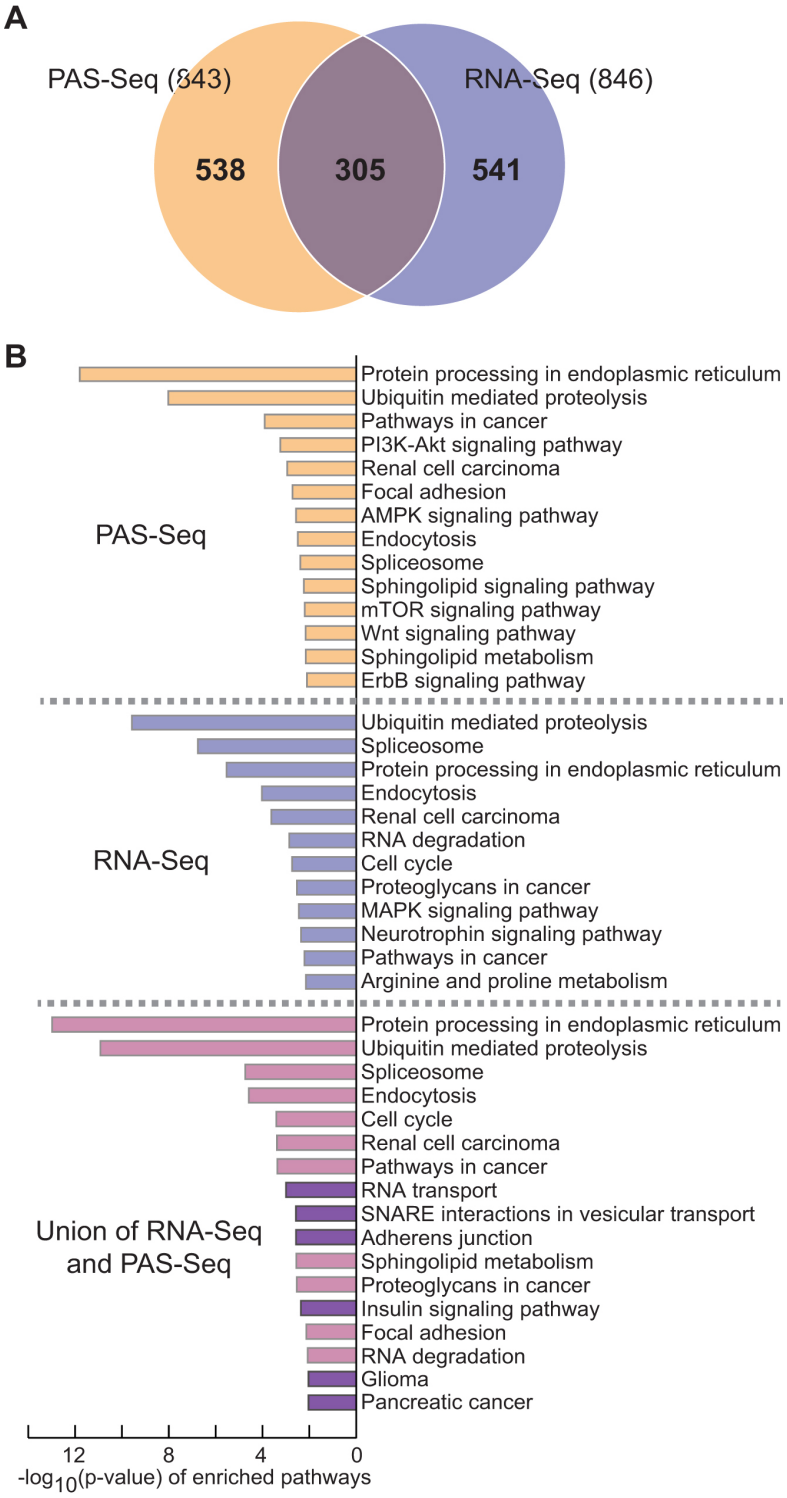
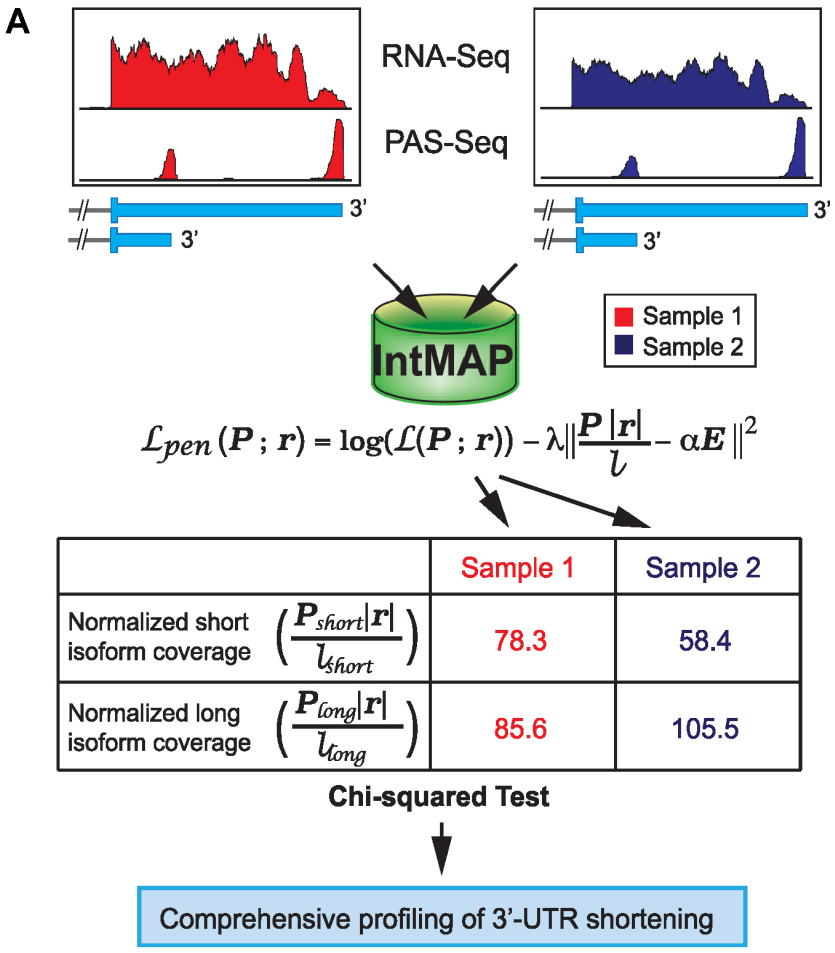


Figure 5. Profiling of 3'-UTR APA events in the mTORC1-activated transcriptome using Poly(A) Site sequencing (PAS-Seq). (A) A Venn diagram of 3'-UTR shortened transcripts in TSC1$^{-/-}$ MEFs using two independent profiling methods. (B) KEGG pathway analysis of 3'-UTR shortened transcripts identified by PAS-Seq, RNA-Seq and the union of the two methods. The KEGG pathways only enriched after combining two profiling methods are highlighted in dark violet.

Even though the union of two datasets shows an advantage in expanding the APA profile of a transcriptome over using single APA profiling method, this approach does not fully complement limitations inherited from each profiling method. Thus, we developed a novel algorithm named IntMAP (Integrative Model for Alternative Polyadenylation), which integrates RNA-Seq and PAS-Seq data for exhaustive analysis of 3'-UTR APA events (Fig. 6A). In IntMAP, first the positions of multiple polyadenylation sites in a 3'-UTR of a gene are defined and the 3'-UTR isoforms of the gene are accordingly deduced. Then the quantitative information of RNA-Seq and PAS-Seq data is integrated to calculate the expression level of inferred 3'-UTR isoforms. Two elements in IntMAP work systemically to help the quantitation of isoform expression. The first element promotes the isoform expression to comply with the observed read counts from RNA-Seq data. The second element encourages the consistency between the isoform expression learned from RNA-Seq and PAS-Seq data. After the quantitation by IntMAP, the calculated expression level of different 3'-UTR isoforms is applied to the chi-squared test to determine the 3'-UTR shortening of a gene in a given biological context compared to control (Fig. 6A). Next, we applied IntMAP to our RNA-Seq and PAS-Seq data and found 975 genes with 3'-UTR APA events (Fig. 6B and Supplementary Table 6). Among 975 genes, 592 and 370 genes overlapped with the PAS-Seq and RNA-Seq analyses, respectively. Of note, IntMAP could not entirely encompass the genes with 3'-UTR APA found in the PAS-Seq or RNA-Seq analysis (Fig. 6B). Importantly, IntMAP identified 280 new 3′-UTR APA events, which were not reported in the union of two data sets (Fig. 6B and Supplementary Table 6). By pooling genes showing APA altogether from RNA-Seq, PAS-Seq, and

IntMAP analyses, we could determine 1658 APA events in the TSC1$^{-/-}$ transcriptome (Fig. 6B).



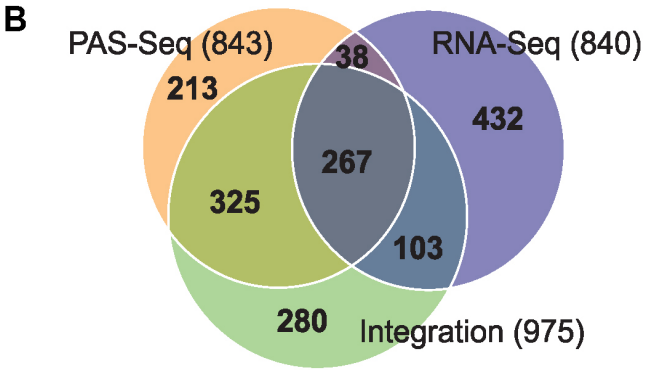


Figure 6. Development of a novel algorithm IntMAP by integrating RNA-Seq and PAS-Seq dataset. (A) A schematic of the experimental workflow and algorithm to integrate RNA-Seq and PAS-Seq data. An algorithm integrating two datasets was developed for comprehensive profiling of 3'-UTR APA events. (B) A Venn diagram of 3'-UTR shortened transcripts in $TSC1^{-/-}$ MEFs using IntMAP. The number of genes producing 3'-UTR shortened transcripts in each section of Venn diagram is indicated.

Examples of RNA-Seq and PAS-Seq read alignments from newly identified 280 genes show a marginal pattern for 3'-UTR shortening which was not significant enough to be determined as APA events in both analyses (Fig.7A and Supplementary Fig. 5). But the RSI of those genes measured by qRT-PCR was positive, indicating the production of short 3'-UTR transcripts from those genes (Fig. 7B). The KEGG pathway analysis using 975 genes identified by IntMAP showed distinct enriched pathways such as FoxO signaling pathway (mmu04068) and p53 signaling pathway (mmu04115) as well as pathways identified by RNA-Seq or PAS-Seq (Fig. 7C and Supplementary Table 7). Thus, the newly developed IntMAP could comprehensively catalog 3'-UTR APA events in the $TSC1^{-/-}$ transcriptome and reveal additional biological pathways that might be regulated by APA in $TSC1^{-/-}$ MEFs. Using the list of genes combined altogether, we were able to expand the repertoire of 3'-UTR APA-modulated cellular pathways upon the activation of mTORC1 (Fig. 7D, Supplementary Fig. 6, and Supplementary Table 7).

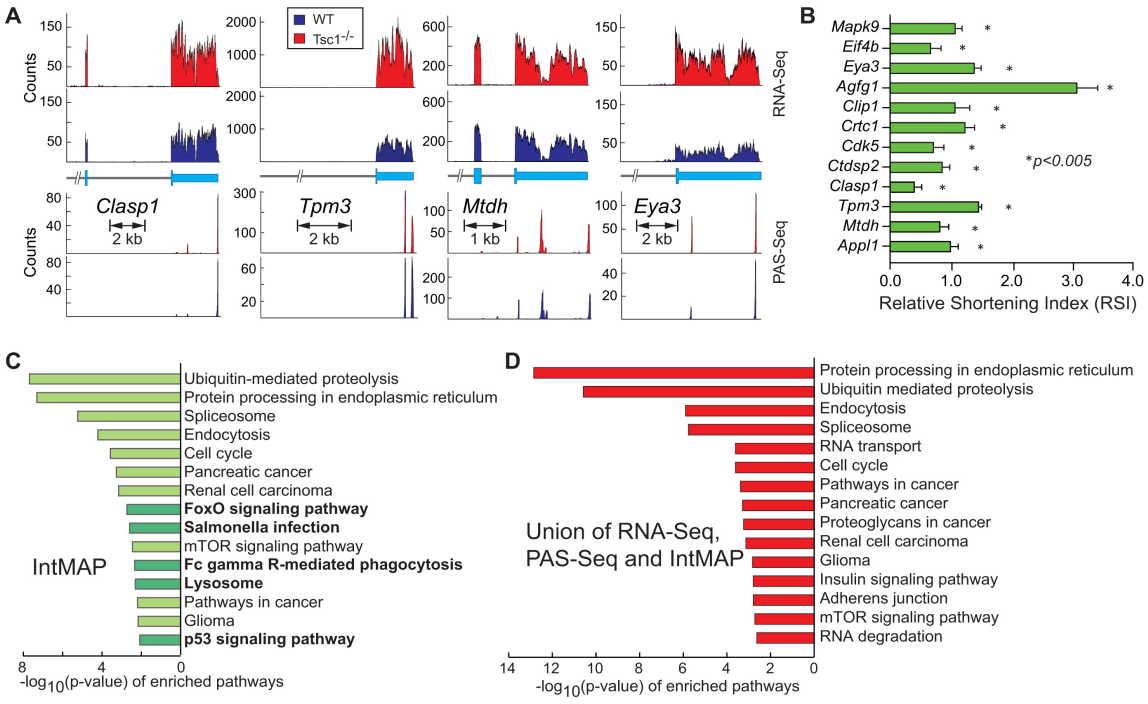


Figure 7. Identification of novel 3'-UTR APA events by IntMAP. (A) Examples of RNA-Seq and PAS-Seq alignments for genes showing 3'-UTR APA events identified by IntMAP. Read alignments of the 3'-most exon or last two exons of RNA- and PAS-Seq are shown. The orientation of genes is presented from 5' (left) to 3' (right). (B) The relative shortening index (RSI) of genes based on the comparison between WT and TSC1$^{-/-}$ MEFs was measured using qPCR to validate the analysis done by IntMAP. qPCR results are from four technical replicates. Data represent the mean ± SEM. Two-tailed student t-tests were performed for statistical significance. (C) KEGG pathway analysis using 975 genes identified by IntMAP is shown. The KEGG pathways only enriched in IntMAP analysis are highlighted in dark green. (D) KEGG pathway analysis of 3'-UTR APA events in the union of RNA-Seq, PAS-Seq and IntMAP analyses is shown. The complete list of KEGG pathways is shown in Supplementary Figure 6.

Alternative Polyadenylation as a Molecular Link between Cellular Stress Response and mTORC1 Activation

To determine whether the APA events solely identified by IntMAP have significant physiological impacts, we set out to examine the newly identified 280 APA events more closely. We first performed the gene ontology (GO) analysis to search for enriched molecular functions in 280 genes. Because mTORC1 has been known to activate many cellular pathways through the modulation of transcriptional networks, we particularly looked for transcription-related functions in the enriched GO analysis[46,47]. From this approach, we found that several transcription factors including *Appl1*, *Mtdh* and *Cebpg* showed 3'-UTR APA in the TCS1$^{-/-}$ transcriptome (Supplementary Table 8). Consistent with IntMAP-identified APA events, the RSI of those transcription factors showed 3'-UTR shortening although the APA event was not apparent in the RNA-Seq read alignments and PAS-Seq counts (Fig. 7A–B and 8A, Supplementary Fig. 5). Notably, several of these transcription factors have not been linked to the mTORC1 signaling pathway extensively, including C/EBPγ. C/EBPγ is known to function in the integrated cellular stress responses to redox imbalance and nutrient deprivation stress by forming a heterodimer with ATF4[32]. In TSC1$^{-/-}$ MEFs, the expression of C/EBPγ protein is highly upregulated compared to WT MEFs (Fig. 8B and Supplemental Fig. 7B). Of note, as opposed to the drastic increase in the protein level, the transcript level of *Cebpg* only increased by ~2-fold from WT to Tsc1$^{-/-}$ MEFs (Fig. 8A–B and Supplementary Fig. 7B). The fact that the protein level change cannot be fully explained by the transcript level change indicates that APA plays an important role in the regulation of C/EBPγ expression in WT and Tsc1$^{-/-}$ MEFs.

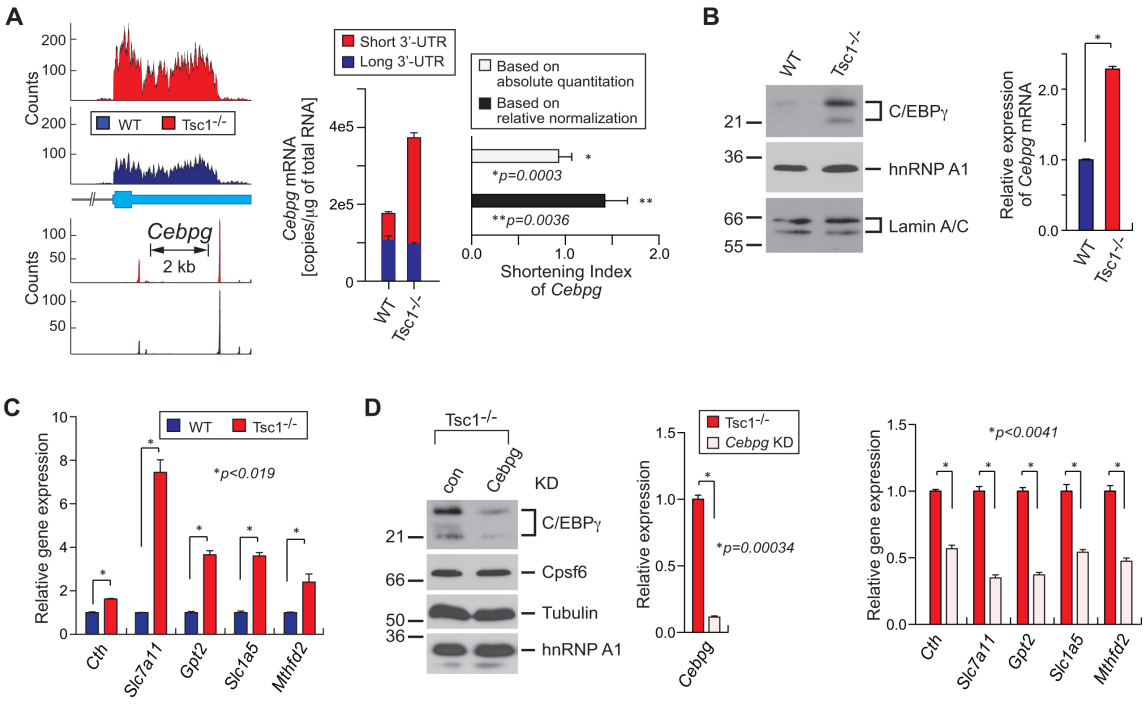


Figure 8. mTORC1-coordinated cellular stress response network identified by IntMAP. (A) RNA- and PAS-Seq read alignments of Cebpg are shown. An absolute quantitation of long 3'-UTR and total (long+short 3'-UTR) *Cebpg* mRNAs was made using qPCR. The amounts of short versus long 3'-UTR transcripts are presented by copies/µg of total RNAs. The shortening index of *Cebpg* in the TSC1$^{-/-}$ transcriptome was calculated based on absolute or normalized quantitation. qPCR results are from 4 technical replicates. Data represent the mean ± SEM. Two-tailed Student's t-tests were performed for statistical significance. (B) Western blotting for Cebpg, hnRNP A1 and Lamin A/C using WT and TSC1$^{-/-}$ cell extracts. Relative amounts of *Cebpg* mRNA were measured by qPCR. qPCR results are from four technical replicates. Data represent the mean ± SEM. Two-tailed student t-tests were performed for statistical significance. *P = 7.97e–007. (C) Relative gene expression of Cebpg target genes in WT and TSC1$^{-/-}$

MEFs was measured using qPCR. qPCR results are from 4 technical replicates. Data represent the mean ± SEM. Two-tailed Student's t-tests were performed for statistical significance. (D) RNAi knockdown (KD) of *Cebpg* in Tsc1$^{-/-}$ MEFs and the analysis of Cebpg target genes. Western blotting was conducted for Cebpg, Cpsf6, Tubulin, and hnRNP A1 using con or *Cebpg* KD Tsc1$^{-/-}$ cell extracts. Relative amounts of *Cebpg* mRNA in the KD cells compared to control cells were measured using qPCR. Expression of Cebpg target genes in the KD and control cells were quantitated using qPCR. qPCR results are from four technical replicates. Data represent the mean ± SEM. Two-tailed Student's t-tests were performed for statistical significance.

Because C/EBPγ is known to activate oxidative stress response-related transcription networks with ATF4[32], we examined the expression level of C/EBPγ downstream target genes using qPCR in $TSC1^{-/-}$ and WT MEFs. As anticipated, all tested target genes of C/EBPγ were significantly up-regulated in $TSC1^{-/-}$ compared to WT MEFs (Fig. 8C). To validate whether the up-regulation of these genes is specifically mediated by the increase of C/EBPγ in $TSC1^{-/-}$ MEFs, we specifically knocked down C/EBPγ using RNAi (Fig. 8D and Supplementary Fig. 7A) and measured its effect on target gene expression. When C/EBPγ was down-regulated, the expression of all tested genes was decreased (Fig. 8D), indicating that the transcriptional activation of oxidative stress response-related genes in $TSC1^{-/-}$ MEFs is mediated by the upregulation of C/EBPγ. Together, these demonstrate that, by integrating the APA profiling of two different sequencing technologies by IntMAP, we were able to further unveil regulatory connections between mTORC1-mediated 3'-UTR shortening to previously unassociated cellular functions.

Discussion

In this study, we used genetically well-defined MEFs to investigate the molecular signatures of mTORC1-activated transcriptome and discovered widespread 3'-UTR shortening due to dysregulated mTORC1 activation. Although a precise mechanism(s) of how mTORC1 activation leads to the 3'-UTR shortening in selected transcripts is unknown, we found that almost all known 3'-end processing factors alter their expression on changes in cellular mTORC1 activity in $TSC1^{-/-}$ compared with WT MEFs, suggesting that mTORC1-mediated 3'-UTR shortening occurs by multiple factors (Supplementary Table 4). Analysis on pathways enriched in 3'-UTR-shortened transcripts and quantitative proteomics on mTORC1 activation identified ubiquitin-mediated proteolysis as an additional target pathway of mTORC1. Considering the well-documented function of mTORC1 in the activation of cellular anabolic metabolism for rapid cell proliferation[48], the newly discovered function of mTORC1 in ubiquitin-mediated proteolysis through 3'-UTR shortening is surprising. A recent study suggests a role of mTOR in the activation of proteasome through the modulation of a transcriptional network[28]. This study argued that the promotion of protein degradation pathway on mTOR activation is required for a continuous supply of amino acids to cellular systems, to maintain the steady-state protein synthesis. Our data set also indicates a marginal increase in the proteasome activity through a transcriptional upregulation of several proteasomal subunits (Supplementary Fig. 4D) and an increase in the polyubiquitination of proteins on mTORC1 activation (Supplementary Fig. 4I). Moreover, our data demonstrate that mTORC1-promoted 3'-UTR shortening leads to the overexpression of selected E2 and E3 components in ubiquitin ligase complexes (Figs. 3B, D and 4B),

which is known to increase polyubiquitination of their substrates[49–55]. Therefore, it is possible that the enrichment of polyubiquitination in cellular proteins on mTORC1 activation could come from selective polyubiquitination of those E2 and E3 substrates. Together, our study proposes the molecular mechanism of how mTORC1 pathway selects proteins to degrade through 3'-UTR shortening of a subset of mRNAs. The E2 and E3 enzymes upregulated by mTORC1-driven 3'-UTR shortening mostly target cell cycle regulators, tumor suppressors and pro-apoptotic proteins for ubiquitin–proteasome system[50,56]. For instance, Rbx1, Trip12 or Anapc1/5 are all components of E3 ligase complexes that selectively polyubiquitinate Arf and Cyclins, and Birc6 E3 ligase targets Caspase 3/7 for degradation[50,56,57]. For rapidly proliferating cells, a timely removal of cell cycle regulators such as Arf and Cyclins is a key step for rapid progression of the cell cycle[50]. Therefore, our findings are particularly important, because unlike a previous argument, it explains how upregulated mTORC1 and consequent proteasome activation recycles proteins that are only needed to foster a cellular environment favorable to rapid cell proliferation. Thus, we suggest that this selective proteolysis not only provides a surplus of amino acids to cellular systems but also makes cells proliferate rapidly by efficient modulation of the cellular levels of cell cycle regulators.

Although a consensus on the role of mTORC1 is an increase in global protein synthesis, many studies suggest that mTORC1-modulated promotion of protein synthesis requires numerous trans- and cis-acting elements. In particular, 5'TOP and 5'PRTE sequence features found in the 5'-UTR in a group of mRNAs render the activation of cellular pathways such as translation, cellular invasion and metastasis[26,27]. One of the

distinctive behaviors of this group of mRNAs is their acute dissociation from actively translating polysomes in response to mTOR inhibition[26,27]. Interestingly, *Rpl22* mRNA contains both 5'TOP and 3'-UTR-shortening molecular features. Compared with mRNAs containing 5'TOP feature, the formation of polysomes with this particular mRNA is not entirely dependent on 5'TOP but is rather regulated by 3'-UTR shortening, because the long 3'-UTR *Rpl22* mRNA was not enriched in the polysome fractions in TSC1$^{-/-}$ MEFs (Supplementary Fig. 4A, C). As most of the 3'-UTR-shortened transcripts for ubiquitin-mediated proteolysis do not contain mTOR-responsive sequence features in their 5'-UTR (Supplementary Fig. 3D), 3'-UTR shortening due to mTOR activation could function in protein synthesis independent of 5'-UTR structure of an mRNA and it presents an additional molecular signature targeted by the mTORC1 pathway.

Profiling the changes of 3'-UTR length is a critical aspect in the understanding of comprehensive gene expression. Multiple workflows adopting high profiling technologies and bioinformatics tools have been developed to catalog the dynamics of 3'-UTR length on a global scale[1]. However, an exhaustive survey of 3'-UTR APA is still far from complete due to inherent weaknesses of currently available profiling methods. For example, several versions of profiling methods were designed to specifically sequence the fragmented poly(A) tail region of mRNAs[14,16,18,19,58–60]. Despite the advantage of acquiring high resolution peaks of polyadenylation sites transcriptome-wide, these methods suffer from false signals caused by internal priming on A-rich sequences instead of poly(A) tails and frequent unmanageable production of very short reads during the library preparation[1,15]. In addition, a variable efficiency of library preparations between

samples often complicates the quantitation of APA events. Alternatively, widely conducted RNA-Seq experiments can be used to assess APA events. In this case, a customized or current bioinformatics algorithm is needed to register the pattern of APA events in the RNA-Seq data[1]. Depending on the parameters and considerations implanted in an algorithm, the outcome of the analysis could vary. For instance, the definition of the actionable proximal PAS in a 3'-UTR could vary by the algorithms, and the resolution on the usage of multiple proximal PASs in a 3'-UTR could be limited by the nature of RNA-Seq read alignments.

Given that APA is a critical molecular process that coordinates mTORC1 downstream pathways, the resolution and depth of APA survey is critical in investigating mTORC1-regulated pathways. Accordingly, we looked for any combinations of sequencing technologies and bioinformatics tools to help improve the APA survey in the $TSC1^{-/-}$ transcriptome. One simple solution to this quest was simply combining RNA-Seq and PAS-Seq data. As shown in Figure 5, this approach expanded the number of APA events in the $TSC1^{-/-}$ transcriptome compared to the APA analysis using only a single method. However, this union method gives equal weight to each profiling method regardless of the quality of the data. In addition, one method does not complement the data analysis of the other method in this approach. In contrast, newly developed IntMAP formulates one unified machine learning framework to integrate the quantitative information from RNA-Seq and PAS-Seq to estimate the expression level of different 3'-UTR isoforms based on a global optimization strategy. IntMAP could amplify the weak signals of 3'-UTR shortening from both data platforms to avoid false negative genes.

Thus, this method improves isoform quantification and APA profiling compared to RNA-Seq or PAS-Seq method alone, or the simple union of both methods. In addition, the model can be tailored by the parameter λ to balance the contributions of RNA-Seq and PAS-Seq data. As such, the optimization of λ is critical, and the model can be trained by qRT-PCR validation experiments on a small set of genes (Fig. 6A). Intriguingly, we noticed that the APA events characterized by IntMAP do not completely encompass the RNA-Seq or PAS-Seq data (Fig. 6B). This may reflect the ability of IntMAP to filter out false positives. Some APA events falsely identified by the RNA-Seq method may not contain experimentally proven actionable PASs and thus are eliminated by IntMAP due to the lack of APA evidence in PAS-seq data. Other APA events falsely identified by the PAS-Seq method may come from the internal priming on A-rich regions, which can be eliminated by IntMAP based on insignificant changes in read density before and after the putative poly(A) sites. On top of this, each APA profiling method uses different cut-off values for the statistical significance of data analysis, which could cause an incomplete overlap among the datasets.

Characterization of the 3'-UTR APA events in the $TSC1^{-/-}$ transcriptome by the RNA-Seq method highlighted the mTORC1-regulated ubiquitin-mediated proteolysis[61]. Consistent with these findings, the ubiquitin-mediated proteolysis was ranked second among the enriched pathways in the APA events profiled by IntMAP (Fig. 7D). Compared to the 27 genes previously identified, we found that 41 genes were enriched in the ubiquitin-mediated proteolysis pathway by APA with IntMAP (Supplementary Table 7). These results again provide a link between the mTORC1 signaling pathway and

ubiquitin-mediated proteolysis. Importantly, in contrast to the previous findings of ubiquitin-mediated proteolysis mainly targeting tumor suppressors or cell cycle regulators, the newly identified APA genes of E2 and E3 ligases regulate cellular pathways of hypoxia, p53, and other signaling cascades (Supplementary Table 7). Thus, IntMAP helps us further demonstrate that mTORC1 activation controls diverse cellular pathways through a much broader spectrum of APA-driven selective proteolysis.

Similarly, finding several novel transcription factors whose expressions are driven by 3'-UTR APA with IntMAP also expands the identification of cellular pathways regulated by mTORC1-modulated transcription networks. mTORC1 is known to regulate diverse cellular mechanisms including lipid metabolism and stress responses through transcriptional activation or post-translational modification of transcription factors[46,47,62,63]. Thus, the molecular basis of C/EBPγ activation through mTOR-driven APA and subsequent regulation of transcription network provide a new mechanistic insight into cellular stress responses at the level of post-transcriptional regulation.

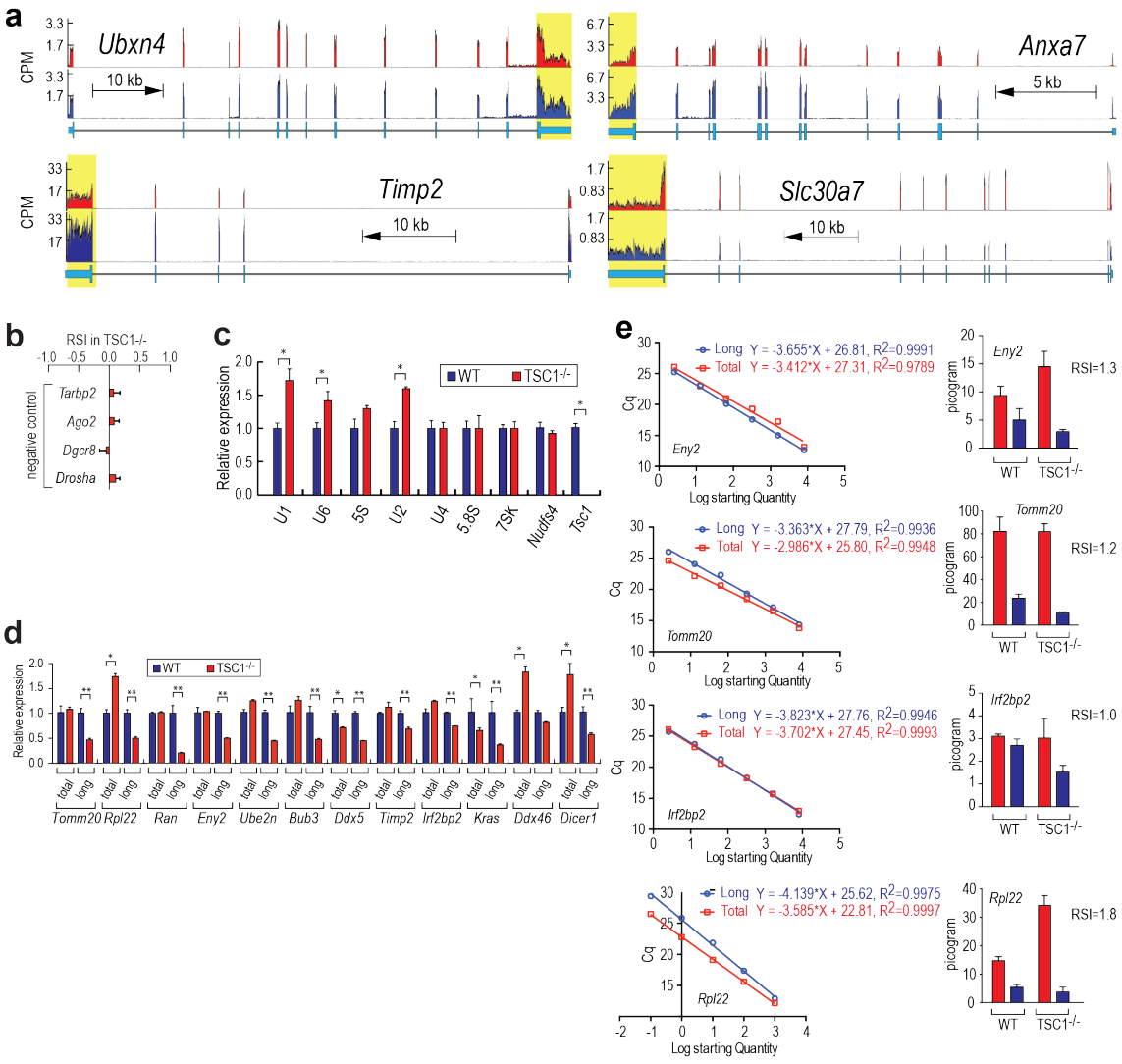
Overall, the data shown in this study showcase IntMAP as a powerful bioinformatics tool for a comprehensive profiling of transcriptome-wide 3'-UTR APA events. As exemplified by the findings of C/EBPγ, newly identified cellular pathways and transcription networks activated by mTORC1-driven APA events provide a broader list of downstream target pathways of mTORC1 signaling cascade. In this context, applications of IntMAP to other biological conditions or disease models will help elucidate new cellular pathways activated by the 3'-UTR APA events. Additionally, IntMAP is a new general approach for flexible integration of RNA-Seq short read

alignments with other mRNA quantification platforms to detect gene transcript variances. For example, other than integration of RNA-seq and PAS-seq, IntMAP can also be applied to the integration of RNA-Seq data and Exon-array expressions or NanoString color-barcode counts for isoform detection/quantifications. In these applications, similar qRT-PCR validation experiments can be applied to optimize the tuning parameter λ, which balance the contributions of RNA-Seq and other mRNA quantification platforms.

Acknowledgement

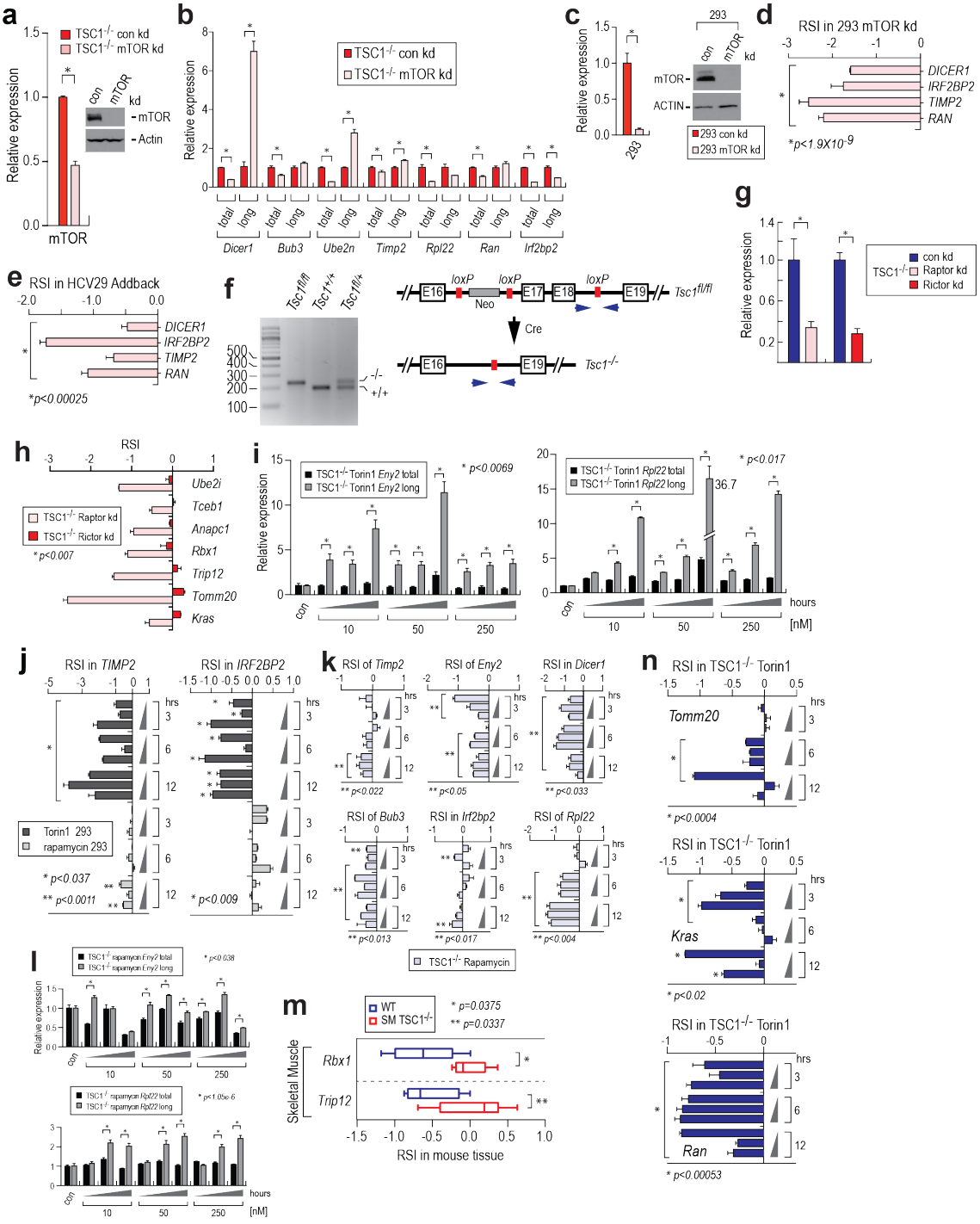
We thank Dr Kwiatkowski at Bringham and Women's Hospital and Harvard Medical School for providing us $TSC1^{+/+}$ and $TSC1^{-/-}$ MEF cell lines and HCV29 bladder cancer cell lines used in this study. We also thank Mr Todd Markowski and Dr LeeAnn Higgins at the Center for Mass Spectrometry and Proteomics, University of Minnesota, for their work in peptide fractionation and LC–MS/MS analysis. This work was supported in part by the following funding: Institutional Research Grant (118198-IRG-58-001-52-IRG76) from the American Cancer Society for J.Y., and the grant for the Bio & Medical Technology Development Program (2012M3A9B4028738) by the National Research Foundation of Korea (NRF) for K.-S.K and J.Y., NSF III 117153 for R.K. and GM097057, AG039758, ADA 7-12-BS-093 and W81XWH-13-1-0060 for D.-H.K.

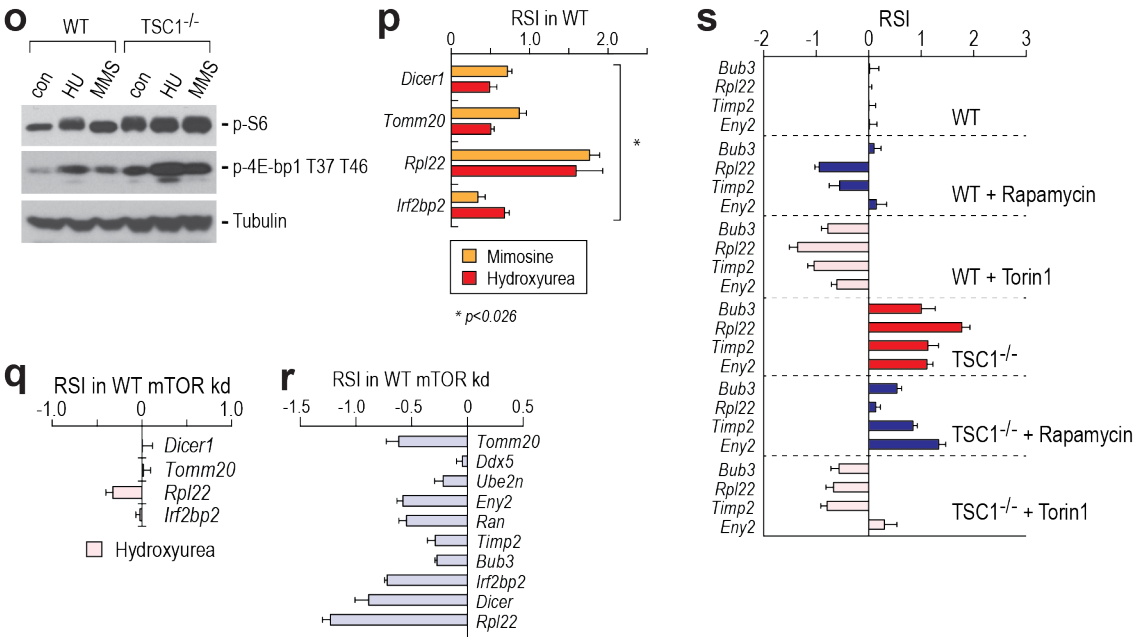
Supplementary Materials



Supplementary Fig. 1. Genome-wide 3'UTR shortening by mTORC1 activation. (A) Additional individual examples of transcripts with 3'UTR shortening are presented. The yellow boxes highlight the 3'UTR of annotated genes and sequence reads. (B) Four transcripts that did not show 3'UTR shortening in the mTORC1-activated transcriptome were used as negative controls for the RSI measurement. (C) Expression of several small

non-coding RNAs and mRNAs in WT and TSC1-/- MEFs measured by RT-qPCR. The relative expression level of each RNA was normalized to total RNAs. 7SK snRNA, and Nudfs4 expressions were used to confirm that each RT-qPCR experiments used the same amount of total RNAs. Rather than using a control mRNA for normalization control, we obtained comparative Ct values based on total amount of RNAs we used for RT-qPCR analysis. Then we performed another RT-qPCR analysis to measure the amount of control RNAs transcribed by different RNA polymerases including 7SK snRNA, and Nudfs4. By doing this, we could ensure that the same amount of total RNAs are used for each RT-qPCR and several normalization controls are used for quantitation. Student t-tests were performed to obtain statistical significance, which is presented by p-value. * $p<0.023$ (D) The data in Figure 1(D) were shown as the relative amount of total (long+short) or long transcripts for each tested genes. The level of total or long transcripts was normalized to total RNAs. * $p<0.05$, ** $p<0.015$ (E) Quantitation of RT-qPCR was done using standard curve method. Absolute quantity of PCR products using total or long-specific primers was measured and the standard curves were plotted using linear regression. The quantity of total and long transcripts in WT and $TSC1^{-/-}$ was calculated and plotted in picogram. The RSI for each gene was shown based on the absolute quantitation. Student t-test was done for each pair of total and long. * $p<0.019$

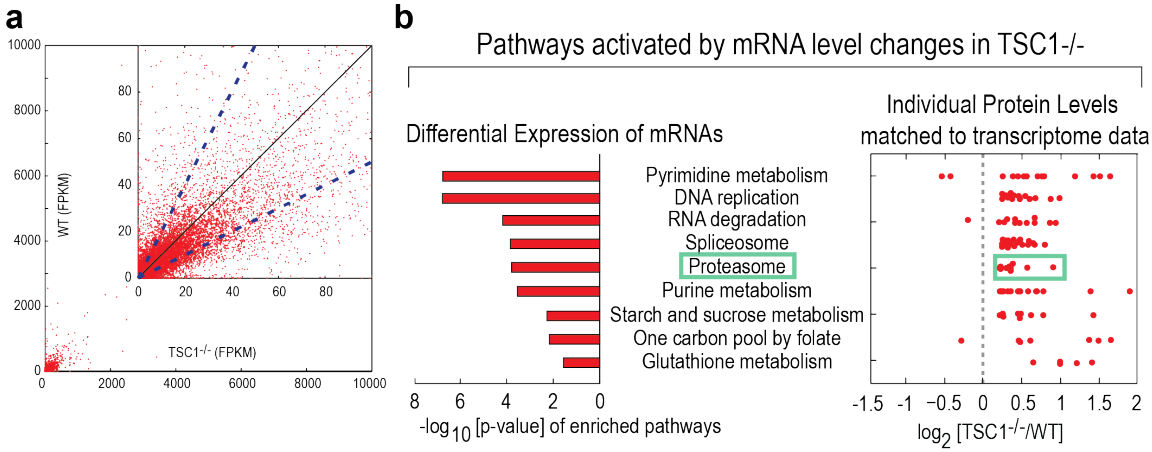

a
TSC1-/- con kd
TSC1-/- mTOR kd
Relative expression
mTOR
con
mTOR
kd
mTOR
Actin
b
TSC1-/- con kd
TSC1-/- mTOR kd
Relative expression
total
long
Dicer1
Bub3
Ube2n
Timp2
Rpl22
Ran
Irf2bp2
c
Relative expression
293
293
con
mTOR
kd
mTOR
ACTIN
293 con kd
293 mTOR kd
d
RSI in 293 mTOR kd
DICER1
IRF2BP2
TIMP2
RAN
*p<1.9X10-9
e
RSI in HCV29 Addback
DICER1
IRF2BP2
TIMP2
RAN
*p<0.00025
f
Tsc1fl/fl
Tsc1+/+
Tsc1fl/+
500
400
300
200
100
-/-
+/+
loxP
loxP
loxP
E16
Neo
E17
E18
E19
Tsc1fl/fl
Cre
E16
E19
Tsc1-/-
g
Relative expression
con kd
TSC1-/- Raptor kd
Rictor kd
h
RSI
Ube2i
Tceb1
Anapc1
Rbx1
Trip12
Tomm20
Kras
TSC1-/- Raptor kd
TSC1-/- Rictor kd
* p<0.007
i
TSC1-/- Torin1 Eny2 total
TSC1-/- Torin1 Eny2 long
* p<0.0069
Relative expression
con
10
50
250
TSC1-/- Torin1 Rpl22 total
TSC1-/- Torin1 Rpl22 long
* p<0.017
36.7
hours
[nM]
j
RSI in TIMP2
RSI in IRF2BP2
hrs
Torin1 293
rapamycin 293
* p<0.037
** p<0.0011
* p<0.009
k
RSI of Timp2
RSI of Eny2
RSI in Dicer1
RSI of Bub3
RSI in Irf2bp2
RSI of Rpl22
** p<0.022
** p<0.05
** p<0.033
** p<0.013
** p<0.017
** p<0.004
TSC1-/- Rapamycin
l
TSC1-/- rapamycin Eny2 total
TSC1-/- rapamycin Eny2 long
* p<0.038
Relative expression
TSC1-/- rapamycin Rpl22 total
TSC1-/- rapamycin Rpl22 long
* p<1.05e-6
hours
[nM]
m
WT
SM TSC1-/-
* p=0.0375
** p=0.0337
Skeletal Muscle
Rbx1
Trip12
RSI in mouse tissue
n
RSI in TSC1-/- Torin1
Tomm20
* p<0.0004
RSI in TSC1-/- Torin1
Kras
* p<0.02
RSI in TSC1-/- Torin1
Ran
* p<0.00053

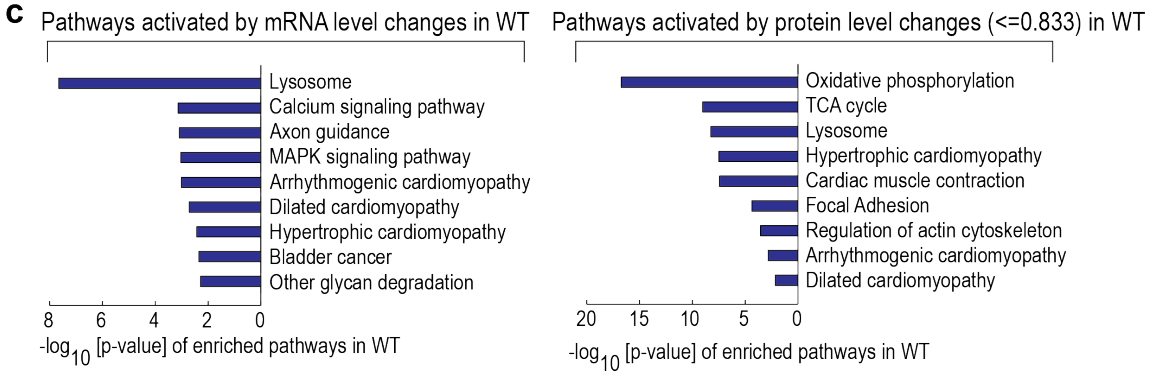


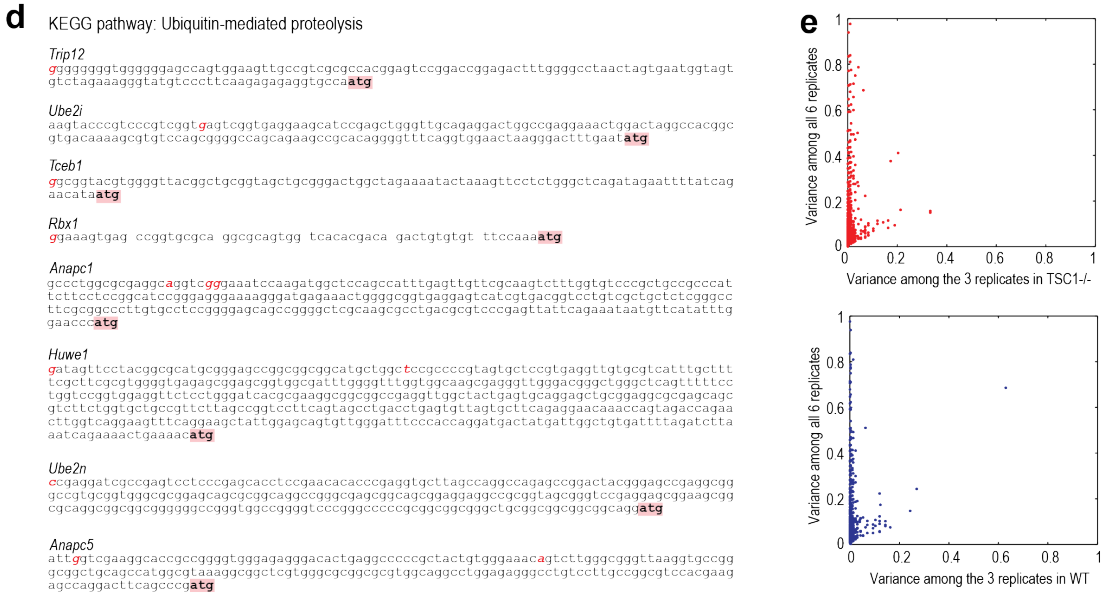
Supplementary Fig. 2. The mTORC1 pathway targets 3'UTR shortening. (A) A TSC1$^{-/-}$ stable cell line with shRNA to mTOR was established. mTOR knockdown (kd) was measured by RT-qPCR and western blotting. * p<0.00015 (B) The data in Figure 2B are presented as the relative amounts of total (long+short) or long transcripts for each tested gene. The levels of total or long transcripts were normalized to total RNAs. * p<0.012 (C) An mTOR knockdown (kd) stable cell line in HEK293 cells was established using shRNA to mTOR. RT-qPCR and western blotting were performed to confirm the kd of mTOR in the stable cell line. * p<3.7E-7 (D) The RSI for selected transcripts was measured by RT-qPCR using 293 mTOR kd cell line. (E) A bladder cancer cell line HCV29, which carry TSC1 deletion mutation and HCV29 Addback, a cell line with addback of TSC1 into HCV29, were used to measure the RSI. The RSI shows that 3'-

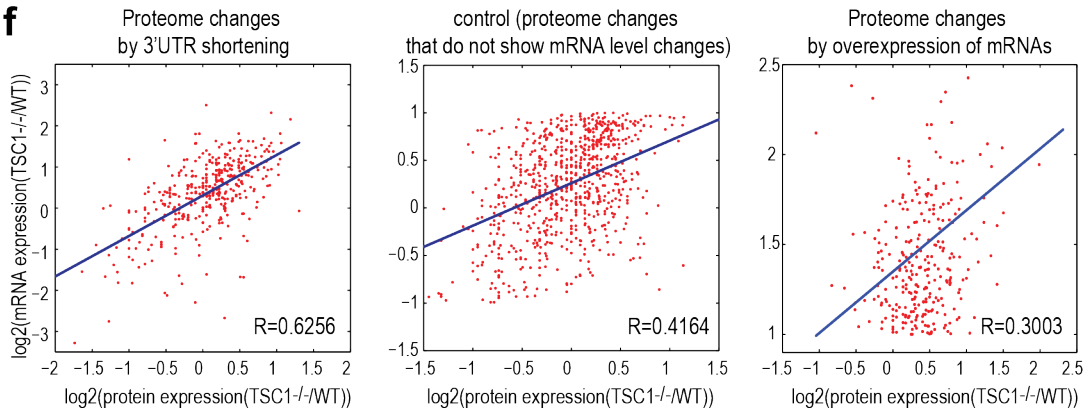
UTR lengthening occurs in HCV29 Addback when compared to HCV29. (F) The gene structure for Tsc1fl/fl mouse and the confirmation of Genotyping. (G) An establishment of stable cell lines for knockdown of Raptor or Rictor in $TSC1^{-/-}$ MEFs. RT-qPCR was performed to measure the relative decrease in Raptor and Rictor mRNAs in the stable cell lines compared to a control cell line. * $p<0.00041$ (H) The RSI for selected transcripts was measured by RT-qPCR using TSC1-/- MEFs with Rictor kd or Raptor kd. (I) The data in Figure 2D are presented as the relative amounts of total (long+short) or long transcripts for each tested transcript. The levels of total or long transcripts were quantitated using RT-qPCR by normalizing to total RNAs. Treatment of Torin1 accumulated long transcripts significantly over total transcripts for both Eny2 and Rpl22. (J) HEK293 cells were treated with selective mTOR inhibitors, rapamycin or Torin1 at various doses (10, 50 and 250 nM) and time courses (3, 6 and 12 hours). Changes in the 3'-UTR length of TIMP2 and IRF2BP2 were determined by measuring the RSI. The asterisks (* and **) indicate the conditions that accumulate the long 3'-UTR with statistical significance. (K) Rapamycin, which does not completely inhibit mTORC1, has a minimal effect on 3'-UTR lengthening. Rapamycin was treated in $TSC1^{-/-}$ MEFs at various doses (10, 50, and 250 nM) and time courses. The RSI was measured by RT-qPCR. Compared to Torin1 (Fig. 2D), rapamycin has a minimal or no effect on accumulating long transcripts for many tested genes. Rapamycin shows an improved effect on accumulating long transcripts for Rpl22 in certain conditions but it is still not as effective as Torin1. The asterisks (**) indicate the conditions that accumulate the long 3'-UTR with statistical significance. (L) The data in Supplementary Fig. 2K are presented as

the relative amounts of total (long+short) or long transcripts for each tested transcript. The asterisks (*) indicate the conditions that accumulate the long 3'-UTR with statistical significance. (M) Activation of mTORC1 in terminally differentiated skeletal muscle leads to 3'-UTR shortening. Total RNAs from skeletal muscles in wild type (WT) or skeletal muscle-specific knockout of Tsc1 (SM $TSC1^{-/-}$) mice were used for the RSI measurement. (N) $TSC1^{-/-}$ MEFs were treated with Torin1 at various doses (10, 50 and 250 nM) and time courses (3, 6 and 12 hours). Changes in the 3'-UTR length of Tomm20, Kras and Ran were determined by measuring the RSI. (O) Western blotting using WT and $TSC1^{-/-}$ MEFs treated with hydroxyurea (HU; 2 mM) or mimosine (MMS; 2 mM) for 12 hours. mTORC1 substrates (p-S6 and p-4E-bp1) were detected using specific antibodies. Western blotting for Tubulin was performed for a loading control. (P) The RSI was measured in WT MEFs treated with hudroxyurea or mimosine. (Q) The RSI was measured in WT MEFs with mTOR kd after the treatment with hydroxyurea. (R) WT MEFs contain 3'-UTR shortened transcripts. mTOR was knocked down using shRNA in WT MEFs and the effect of knockdown on the RSI was measured by RT-qPCR. The RSI indicates that there are 3'-UTR shortened transcripts in WT MEFs. (S) The RSIs in WT, WT treated with Rapamycin or Torin1, $TSC1^{-/-}$, $TSC1^{-/-}$ treated with Rapamycin or Torin1 were measured using RT-qPCR. The RSIs for each tested transcript in WT were used as a reference to present the RSIs in other cell lines. The RSIs in WT treated with Torin1 indicate that there are 3'-UTR shortened transcripts in WT MEFs. The RSIs in $TSC1^{-/-}$ treated with Torin1 indicate that some transcripts reach their 3'-UTR lengthening beyond the level of WT and some do not.


a
WT (FPKM)
TSC1-/- (FPKM)
b
Pathways activated by mRNA level changes in TSC1-/-
Differential Expression of mRNAs
Pyrimidine metabolism
DNA replication
RNA degradation
Spliceosome
Proteasome
Purine metabolism
Starch and sucrose metabolism
One carbon pool by folate
Glutathione metabolism
-log10 [p-value] of enriched pathways
Individual Protein Levels
matched to transcriptome data
log2 [TSC1-/-/WT]

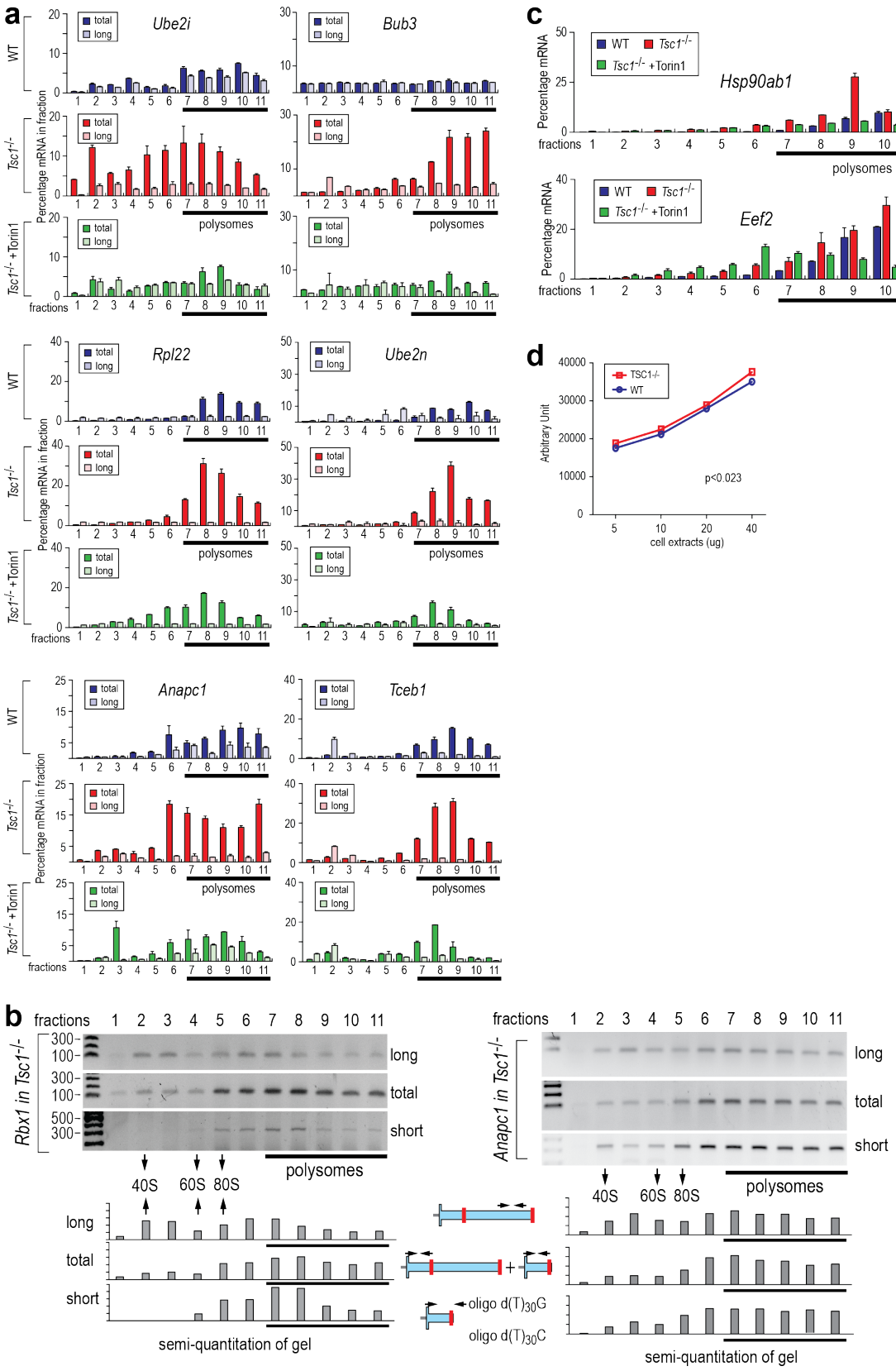

c
Pathways activated by mRNA level changes in WT
Lysosome
Calcium signaling pathway
Axon guidance
MAPK signaling pathway
Arrhythmogenic cardiomyopathy
Dilated cardiomyopathy
Hypertrophic cardiomyopathy
Bladder cancer
Other glycan degradation
-log10 [p-value] of enriched pathways in WT
Pathways activated by protein level changes (<=0.833) in WT
Oxidative phosphorylation
TCA cycle
Lysosome
Hypertrophic cardiomyopathy
Cardiac muscle contraction
Focal Adhesion
Regulation of actin cytoskeleton
Arrhythmogenic cardiomyopathy
Dilated cardiomyopathy
-log10 [p-value] of enriched pathways in WT

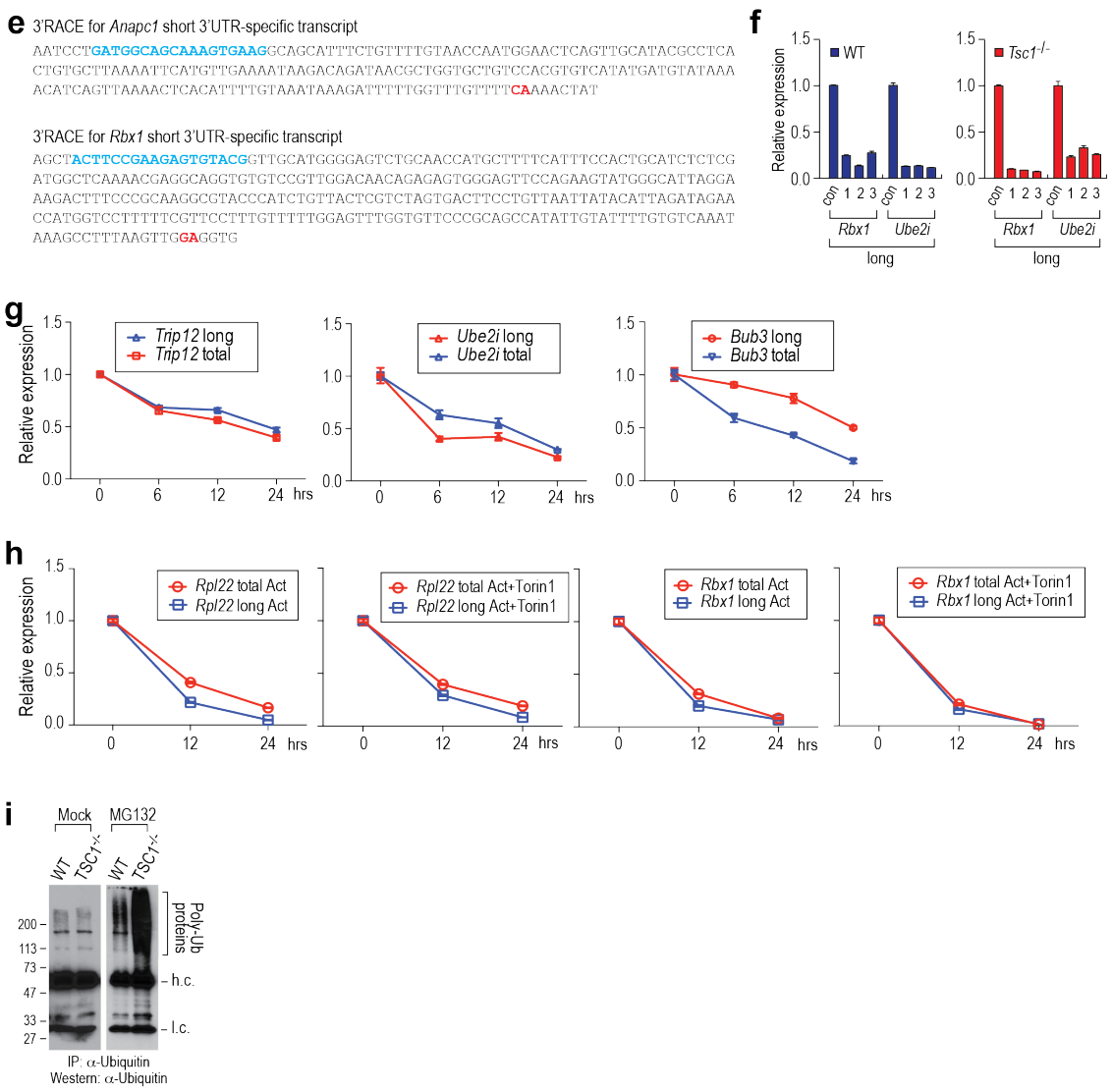

d
KEGG pathway: Ubiquitin-mediated proteolysis
Trip12
Ube2i
Tceb1
Rbx1
Anapc1
Huwe1
Ube2n
Anapc5
e
Variance among all 6 replicates
Variance among the 3 replicates in TSC1-/-
Variance among all 6 replicates
Variance among the 3 replicates in WT



Supplementary Fig. 3. Differential gene expression analysis of mTORC1-activated transcriptome. (A) Differential gene expression analysis of mTORC1-activated transcriptome. FPKM of each transcript was calculated and differential expression was quantitated by log2 [TSC1-/-/WT]. The dotted blue lines represent cut-off for 2-fold differences. (B) Analysis of enriched KEGG pathways by mRNA level changes in TSC1$^{-/-}$ MEF. The KEGG pathways enriched in TSC1$^{-/-}$ MEFs are shown in –log10[p-value]. Fold changes of individual proteins in each pathway due to differential gene expression are plotted in the right box. Proteasome pathway is highlighted in a light green box. (C) Analysis of enriched KEGG pathways by mRNA and protein level changes in WT. The KEGG pathways enriched by proteins showing less than 0.833 fold changes are shown. (D) 5'-UTR structure of the genes for E3 ubiquitin ligase complexes. 5'UTR sequences were examined for 5'-TOP, 5'-TOP-like and 5'-PRTE sequences. Exon 1 and 2 of each gene are presented based on RefSeq annotation. dbTSS (data base of transcription start sites) was used to predict transcription start sites (red bold in italic) for each gene. The initiation codons (atg) are highlighted in light red boxes. In some cases, there are multiple transcription start sites by different annotations in RefSeq and Ensembl. (E) Variance

analyses on quantitative proteomics. Y-axis represents all combined 6 replicate samples (3 WT and 3 $TSC1^{-/-}$). X-axis represents each group of 3 replicates. Each dot is a protein. The protein expression values across the three replicates within each group are more consistent than the protein expression values across all the six replicates. (F) Correlation coefficients between RNA-seq and proteomics data were determined; 3'-UTR shortening vs. no level changes in mRNAs vs. mRNAs with more than 2-fold changes.

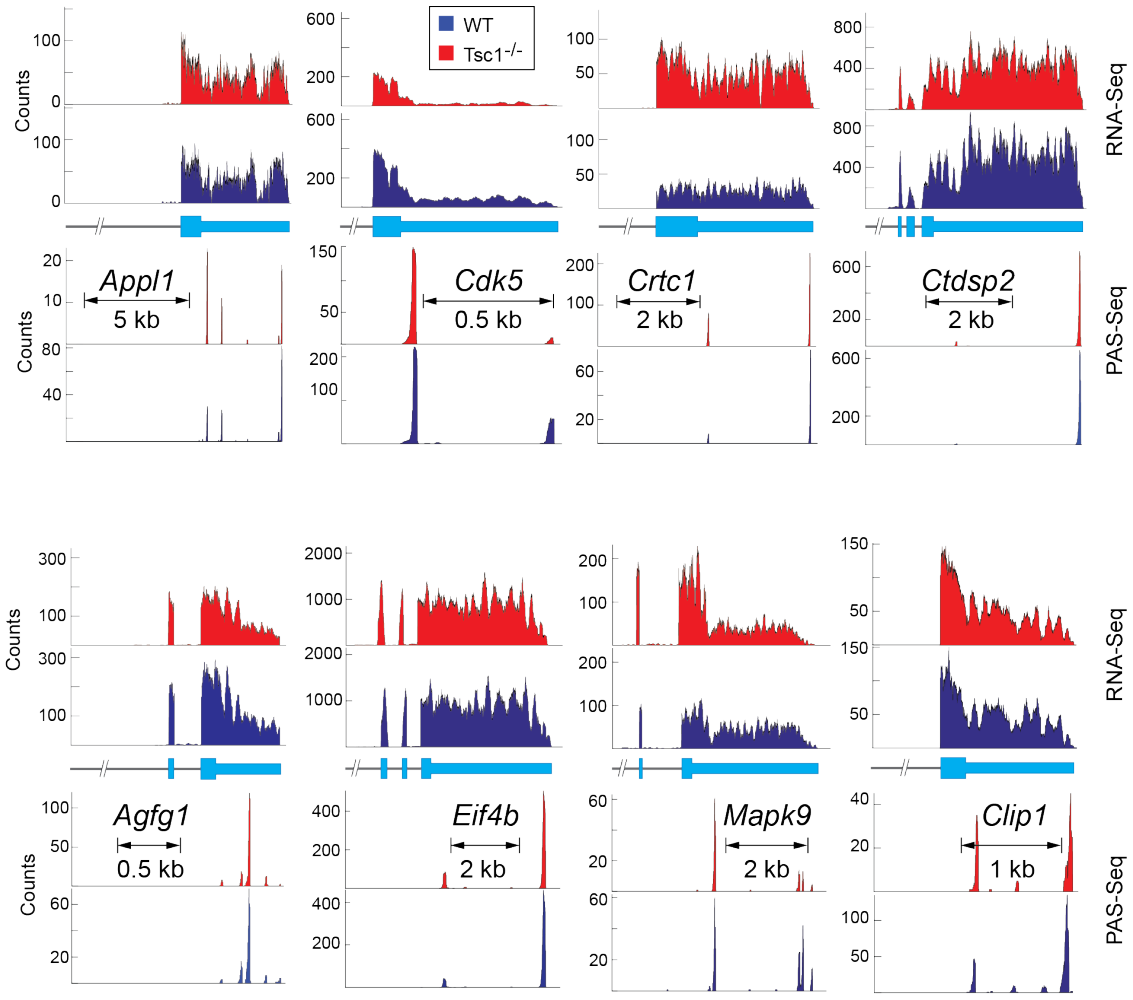

a
Ube2i
Bub3
Rpl22
Ube2n
Anapc1
Tceb1
WT
Tsc1-/-
Tsc1-/- +Torin1
total
long
Percentage mRNA in fraction
polysomes
fractions
c
Hsp90ab1
Eef2
WT
Tsc1-/-
Tsc1-/- +Torin1
Percentage mRNA
fractions
polysomes
d
TSC1-/-
WT
Arbitrary Unit
p<0.023
cell extracts (ug)
b
Rbx1 in Tsc1-/-
Anapc1 in Tsc1-/-
fractions
long
total
short
40S
60S
80S
polysomes
semi-quantitation of gel
oligo d(T)30G
oligo d(T)30C



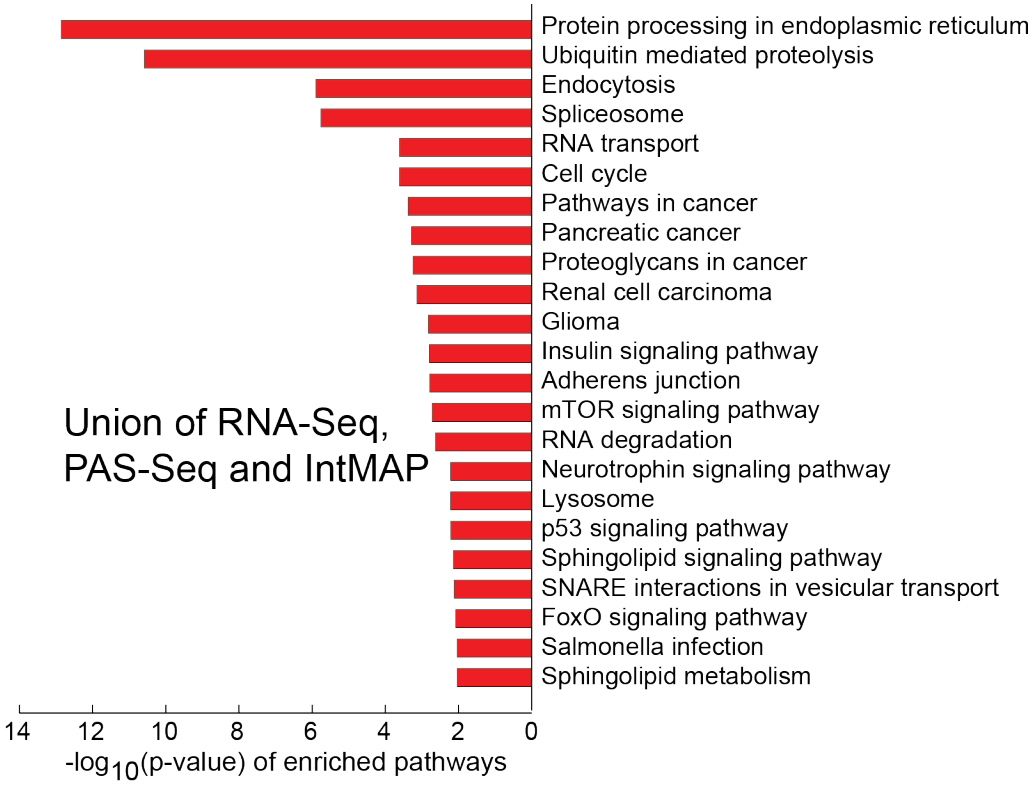
Supplementary Fig. 4. 3'-UTR shortening functions in the up-regulation of E3 ubiquitin ligase complexes. (A) Distribution of total (long+short) and long transcripts in fractions from sucrose gradient. The percentage of total or long transcripts for Ube2i, Rpl22, Bub3, Ube2n, Anapc1 and Tceb1 was measured using RT-qPCR. Polysome fractions corresponding to fractions 7 through 11 are marked by a black line. (B) RT-qPCR results of polysome profiling experiments were validated by RT-PCR and gel-

based assay. 3'RACE experiments were carried out to precisely locate the proximal poly(A) sites in Rbx1 and Anapc1 (Supplementary Fig 4E). Using degenerated oligo d(T)30G or oligo d(T)30C based on 3'RACE, short-specific transcripts were amplified from each fraction. Total and long transcripts were amplified using primer sets for RT-qPCR. The results are shown by agarose gel electrophoresis and semi-quantitation of the images using ImageStudio Light®. (C) Distribution of transcripts containing 5'TOP sequences in the fractions from sucrose gradient. The percentage of transcripts for Hsp90ab1 or Eef2 was measured using RT-qPCR. Polysome fractions corresponding to fractions 7 through 11 are marked by a black line. (D) Proteasome activity assay using cell extracts from WT and $TSC1^{-/-}$ MEFs. Proteasome activity was measured using fluorescence-based method. Three repeats of the experiment were performed. (E) 3'RACE to find proximal poly(A) sites in Rbx1 and Anapc1. The red fonts in the 3'-UTR sequences denote proximal poly(A) sites. The light blue fonts indicate the 5' forward primers for short-specific transcripts. (F) Knockdown of long 3'-UTR transcripts for Rbx1 and Ube2i after the transfection with siRNAs #1, #2 and #3. (G) The stability of total and long 3'-UTR transcripts for each tested gene was measured at various time points after the treatment of actinomycin D at the concentration of 1 ug/ml. (H) The stability of total and long 3'-UTR transcripts for each tested gene was measured at various time points after the treatment of actinomycin D at the concentration of 1 ug/ml and/or Torin1 at 50 nM. (I) Western blot analysis for polyubiquitination of cellular proteins. An anti-Ubiquitin antibody was used to immunoprecipitate ubiquitinated proteins from total cell lysates of WT and $TSC1^{-/-}$ MEFs in the absence or presence of

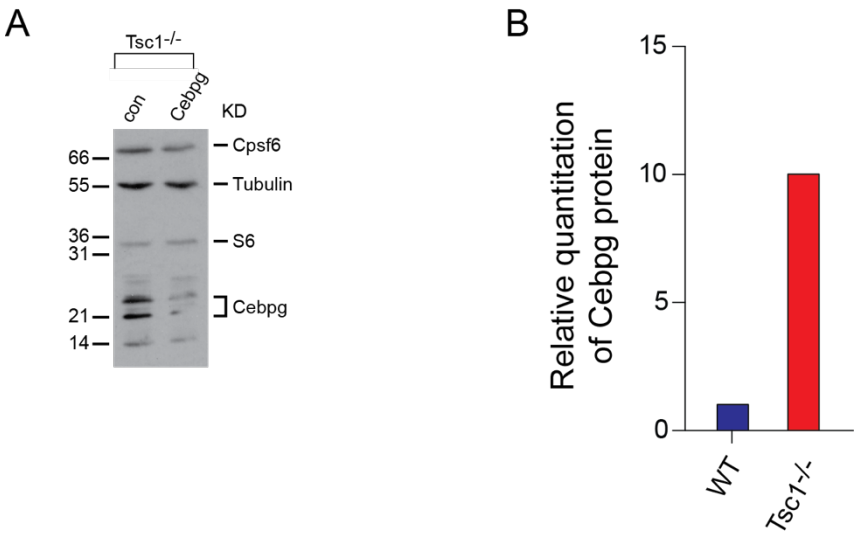
MG132. Polyubiquitinated proteins appeared as smearing bands in the western blotting are marked. H.c and l.c. represent heavy and light chains of the antibody, respectively.



Supplementary Fig. 5. Examples of RNA-Seq and PAS-Seq alignments for genes showing 3'-UTR APA events identified by IntMAP. The orientation of genes is presented from 5' (left) to 3' (right).



Supplementary Fig. 6. The full list of KEGG pathways enriched in the APA events by the union of RNA-Seq, PAS-Seq and IntMAP analyses.



Supplementary Fig. 7. Expression of Cepbg gene. (A) Western blot of Cebpg knockdown in TSC1$^{-/-}$ MEFs. (B) Quantitation of C/EBPγ in western blotting using WT and TSC1$^{-/-}$ extracts. The western results shown Figure 8B was quantitated using Image Studio Lite by Licor.

Supplementary Table 1. Tophat alignment statistics.

original reads			
Sample	Total#	Mapped#	Percentage
WT	127485580	110885148	86.98%
TSC1-/-	148503782	129711943	87.34%

Supplementary Table 2. 3'-UTR shortened transcripts in the mTORC1-activated transcriptome. https://static-content.springer.com/esm/art%3A10.1038%2Fncomms8218/MediaObjects/41467_2015_BFncomms8218_MOESM986_ESM.xls

Supplementary Table 3. The catalog of proteins identified in 2D-LC-MS/MS. https://static-content.springer.com/esm/art%3A10.1038%2Fncomms8218/MediaObjects/41467_2015_BFncomms8218_MOESM987_ESM.xls

Supplementary Table 4. The catalogue of differentially expressed genes in WT and $TSC1^{-/-}$ MEFs. https://static-content.springer.com/esm/art%3A10.1038%2Fncomms8218/MediaObjects/41467_2015_BFncomms8218_MOESM988_ESM.xls

Supplementary Table 5 (Supplementary table 1 in the original manuscript). The gene list of 3'-UTR APA by RNA-Seq and PAS-Seq.

Supplementary Table 6 (Supplementary table 2 in the original manuscript). The list of genes showing 3'-UTR APA by RNA-Seq, PAS-Seq and IntMAP.

Supplementary Table 7 (Supplementary Table 3 in the original manuscript). The enrichment analysis of 3'-UTR APA profiled by IntMAP or 3 unions of APA methods.

Supplementary Table 8 (Supplementary Table 4 in the original manuscript). The GO-term analysis using 280 genes belonging to IntMAP analysis only.

To access Supplementary Table 5–8:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6158760/bin/gky340_supplemental_files.zip

References

1. Hsin-Sung, Y. *et al.* Analyses of alternative polyadenylation: from old school biochemistry to high-throughput technologies. *BMB Reports* **50**, 201–207 (4).
2. Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, (2009).
3. Masamha, C. P. *et al.* CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **510**, 412 (2014).
4. Morris, A. R. *et al.* Alternative Cleavage and Polyadenylation during Colorectal Cancer Development. *Clin Cancer Res* **18**, 5256 (2012).
5. Jenal, M. *et al.* The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. *Cell* **149**, 538–553 (2012).
6. Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nature Communications* **5**, 5274 (2014).
7. Wang, W., Wei, Z. & Li, H. A change-point model for identifying 3′UTR switching by next-generation RNA sequencing. *Bioinformatics* **30**, 2162–2170 (2014).
8. Grassi, E., Mariella, E., Lembo, A., Molineris, I. & Provero, P. Roar: detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinformatics* **17**, 423 (2016).
9. Le Pera, L., Mazzapioda, M. & Tramontano, A. 3USS: a web server for detecting alternative 3′UTRs from RNA-seq experiments. *Bioinformatics* **31**, 1845–1847 (2015).
10. Kim, M., You, B.-H. & Nam, J.-W. Global estimation of the 3′ untranslated region landscape using RNA sequencing. *Methods* **83**, 111–117 (2015).

11. Birol, I. *et al.* Kleat: cleavage site analysis of transcriptomes. *Pac Symp Biocomput* 347–358 (2015).

12. Shenker, S., Miura, P., Sanfilippo, P. & Lai, E. C. IsoSCM: improved and alternative 3′ UTR annotation using multiple change-point inference. *RNA* **21**, 14–27 (2015).

13. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. *Nature* **469**, 97–101 (2011).

14. Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S. & Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* **27**, (2013).

15. Routh, A. *et al.* Poly(A)-ClickSeq: click-chemistry for next-generation 3′-end sequencing without RNA enrichment or fragmentation. *Nucleic Acids Research* **45**, e112–e112 (2017).

16. Zheng, D., Liu, X. & Tian, B. 3′READS+, a sensitive and accurate method for 3′ end sequencing of polyadenylated RNA. *RNA* **22**, 1631–1639 (2016).

17. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3’ untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).

18. Shepard, P. J. *et al.* Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**, 761–772 (2011).

19. Hoque, M. *et al.* Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nature Methods* **10**, 133 (2012).

20. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

21. Tian, B. & Manley, J. L. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**, 312–320 (2013).

22. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA* **106**, (2009).

23. Singh, P. Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res.* **69**, (2009).

24. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511–515 (2010).

25. Laplante, M. & Sabatini, D. M. mTOR Signaling in Growth Control and Disease. *Cell* **149**, 274–293 (2012).

26. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55 (2012).

27. Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109 (2012).

28. Zhang, Y. Coordinated regulation of protein synthesis and degradation by mTORC1. *Nature* **513**, (2014).

29. Parkin, S. E., Baer, M., Copeland, T. D., Schwartz, R. C. & Johnson, P. F. Regulation of CCAAT/Enhancer-binding Protein (C/EBP) Activator Proteins by

Heterodimerization with C/EBPγ (Ig/EBP). *Journal of Biological Chemistry* **277**, 23563–23572 (2002).

30. Cooper, C., Henderson, A., Artandi, S., Avitahl, N. & Calame, K. Ig/EBP (C/EBPγ) is a transdominant negative inhibitor of C/EBP family transcriptional activators. *Nucleic Acids Research* **23**, 4371–4377 (1995).

31. Huggins, C. J. *et al.* C/EBPγ Suppresses Senescence and Inflammatory Gene Expression by Heterodimerizing with C/EBPβ. *Mol. Cell. Biol.* **33**, 3242 (2013).

32. Huggins, C. J. *et al.* C/EBPγ Is a Critical Regulator of Cellular Stress Response Networks through Heterodimerization with ATF4. *Mol. Cell. Biol.* **36**, 693 (2016).

33. Kwiatkowski, D. J. *et al.* A mouse model of TSC1 reveals sex-dependent lethality from liver hemangiomas, and up-regulation of p70S6 kinase activity in Tsc1 null cells. *Human Molecular Genetics* **11**, 525–534 (2002).

34. Tee, A. R., Manning, B. D., Roux, P. P., Cantley, L. C. & Blenis, J. Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. *Curr. Biol.* **13**, (2003).

35. Knowles, M. A., Habuchi, T., Kennedy, W. & Cuthbert-Heavens, D. Mutation spectrum of the 9q34 tuberous sclerosis gene TSC1 in transitional cell carcinoma of the bladder. *Cancer Res.* **63**, (2003).

36. Guo, Y. TSC1 involvement in bladder cancer: diverse effects and therapeutic implications. *J. Pathol.* **230**, (2013).

37. Castets, P. Sustained activation of mTORC1 in skeletal muscle inhibits constitutive and starvation-induced autophagy and causes a severe, late-onset myopathy. *Cell Metab.* **17**, (2013).

38. Guertin, D. A. & Sabatini, D. M. The pharmacology of mTOR inhibition. *Sci. Signal.* **2**, (2009).

39. Thoreen, C. C. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J. Biol. Chem.* **284**, (2009).

40. Fingar, D. C., Salama, S., Tsou, C., Harlow, E. & Blenis, J. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes Dev.* **16**, (2002).

41. Ekim, B. mTOR kinase domain phosphorylation promotes mTORC1 signaling, cell growth, and cell cycle progression. *Mol. Cell Biol.* **31**, (2011).

42. Ma, X. M. & Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nature Reviews Molecular Cell Biology* **10**, 307 (2009).

43. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **9**, (2008).

44. Gruber, A. R., Martin, G., Keller, W. & Zavolan, M. Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. *Wiley Interdiscip. Rev. RNA* **5**, (2014).

45. Shimobayashi, M. & Hall, M. N. Making new contacts: the mTOR network in metabolism and signalling crosstalk. *Nat. Rev. Mol. Cell Biol.* **15**, (2014).

46. Ben-Sahra, I. & Manning, B. D. mTORC1 signaling and the metabolic control of cell growth. *Current Opinion in Cell Biology* **45**, 72–82 (2017).

47. Laplante, M. & Sabatini, D. M. Regulation of mTORC1 and its impact on gene expression at a glance. *Journal of Cell Science* **126**, 1713 (2013).

48. Howell, J. J., Ricoult, S. J., Ben-Sahra, I. & Manning, B. D. A growing role for mTOR in promoting anabolic metabolism. *Biochem. Soc. Trans.* **41**, (2013).

49. Chio, I. I. TRADD contributes to tumour suppression by regulating ULF-dependent p19Arf ubiquitylation. *Nat. Cell Biol.* **14**, (2012).

50. Teixeira, L. K. & Reed, S. I. Ubiquitin ligases and cell cycle control. *Annu. Rev. Biochem.* **82**, (2013).

51. Yanagiya, A. Translational homeostasis via the mRNA cap-binding protein, eIF4E. *Mol. Cell* **46**, (2012).

52. Chen, D., Shan, J., Zhu, W. G., Qin, J. & Gu, W. Transcription-independent ARF regulation in oncogenic stress-mediated p53 responses. *Nature* **464**, (2010).

53. Hagting, A. Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. *J. Cell Biol.* **157**, (2002).

54. Zhou, L. The role of RING box protein 1 in mouse oocyte meiotic maturation. *PLoS One* **8**, (2013).

55. Kraft, C. Mitotic regulation of the human anaphase-promoting complex by phosphorylation. *EMBO J.* **22**, (2003).

56. Jia, L. & Sun, Y. RBX1/ROC1-SCF E3 ubiquitin ligase is required for mouse embryogenesis and cancer cell survival. *Cell Div.* **4**, (2009).

57. Bartke, T., Pohl, C., Pyrowolakis, G. & Jentsch, S. Dual role of BRUCE as an antiapoptotic IAP and a chimeric E2/E3 ubiquitin ligase. *Mol. Cell* **14**, (2004).

58. Mangone, M. *et al.* The Landscape of C. elegans 3′UTRs. *Science* **329**, 432 (2010).

59. Mata, J. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biology* **10**, 1407–1414 (2013).

60. Martin, G., Gruber, A. R., Keller, W. & Zavolan, M. Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. *Cell Reports* **1**, 753–763 (2012).

61. Chang, J.-W. *et al.* mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. *Nat Commun* **6**, 7218 (2015).

62. Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960–976 (2017).

63. Aramburu, J., Ortells, M. C., Tejedor, S., Buxadé, M. & López-Rodríguez, C. Transcriptional regulation of the stress response by mTOR. *Sci. Signal.* **7**, re2 (2014).

CHAPTER THREE

mTORC1 Controls Alternative Splicing Events by Modulating the Splicing of *U2af1* Tandem Exons

This chapter contains an original research article previously published.

Jae-Woong Chang[#], Hsin-Sung Yeh[#], Meeyeon Park, Luke Erber, Jiao Sun, Sze Cheng, Alexander M. Bui, Naima Ahmed Fahmi, Ryan Nasti, Rui Kuang, Yue Chen, Wei Zhang, and Jeongsik Yong (2019) *Nucleic acids research* 47(19), 10373–10387.

[#]These authors contribute equally to this work

Hsin-Sung Yeh and Jae-Woong Chang performed the experiments and Wei Zhang performed the analyses in Figures 1, 3, and 5; Hsin-Sung Yeh performed the experiments and Wei Zhang performed the analyses in Figure 2; Hsin-Sung Yeh performed the experiments and Luke Erber performed the analyses in Figure 4.

Summary

U2 auxiliary factor 1 (U2AF1) functions in 3'-splice site selection during pre-mRNA processing. Alternative usage of duplicated tandem exons in *U2AF1* produces two isoforms, U2AF1a and U2AF1b, but their functional differences are unappreciated due to their homology. Through integrative approaches of genome editing, customized-transcriptome profiling and crosslinking-mediated interactome analyses, we discovered that the expression of U2AF1 isoforms is controlled by mTORC1 and they exhibit a distinctive molecular profile for the splice site and protein interactomes. Mechanistic dissection of mutually exclusive alternative splicing events revealed that U2AF1 isoforms' inherent differential preferences of nucleotide sequences and their stoichiometry determine the 3'-splice site. Importantly, U2AF1a-driven transcriptomes feature alternative splicing events in the 5'-untranslated region (5'-UTR) that are favorable for translation. These findings unveil distinct roles of duplicated tandem exon-derived U2AF1 isoforms in the regulation of the transcriptome and suggest U2AF1a-driven 5'-UTR alternative splicing as a molecular mechanism of mTORC1-regulated translational control.

Introduction

Eukaryotic pre-mRNA is spliced to mRNA by the spliceosome which is composed of small nuclear ribonucleoprotein complexes (snRNPs). Among those snRNPs in the spliceosome, U2 snRNP is critical for splicing by recognizing the branch point [1]. U2 auxiliary factors (U2AFs) are known to bind to polypyrimidine tracts near 3'-splice sites and recruit U2 snRNP to the branch point. U2AFs are a heterodimer consisting of U2AF1 (formerly known as U2AF35) and U2AF2 (formerly known as U2AF65). U2AF2 recognizes the polypyrimidine tract while U2AF1 is known to bind to the AG dinucleotide at 3'-splice site [2–6].

The U2AF1 gene contains duplicated tandem exons between exon 2 and 4. These two duplicated tandem exons (3a and 3b (formerly designated as exon Ab)) are mutually exclusive in splicing and yield two highly similar isoforms, U2AF1a and U2AF1b. They are evolutionary conserved and only differ by seven amino acids in the final protein products (97.1% identity) [7,8]. It has been shown that U2AF1a is more abundant than U2AF1b in various cell lines and tissues [7–9]. Because of inherent similarities and biased expression of U2AF1a, studies on the functional differences between U2AF1 isoforms are largely lacking. Other than a few examples, functional differences between tandem exon-derived isoforms are not well characterized due to similar reasons. However, examples of *PKM* and *FGFR2* genes provide evidence that tandem exon-derived isoforms function differently and distinctively affect cells [10–13].

It is known that three transcripts are transcribed from the *U2AF1* gene. Evolutionarily conserved, mutually exclusive tandem exons drive the transcription of the two isoform transcripts, *U2AF1a* and *U2AF1b*, and the inclusion of both exons produce *U2AF1c* transcript which is subjected to nonsense-mediated mRNA decay [8]. The differences between two U2AF1 isoforms encoded by alternative exon 3 usage occur at the atypical RNA recognition motif which is involved in the dimerization with U2AF2 [14]. Regardless of the seven amino acid differences, however, the two U2AF1 isoforms have biochemically been shown to be similar in forming U2AF heterodimers and functioning in pre-mRNA splicing [8].

The mammalian target of rapamycin (mTOR) pathway has pivotal roles in cell growth, protein translation, and survival[15]. Tuberous sclerosis complexes (TSC1 and TSC2) negatively regulate mTORC1 kinase and genetic knockdown or knockout of TSC (*Tsc1*$^{-/-}$ or *Tsc2*$^{-/-}$) hyperactivates mTORC1 [16,17]. We previously showed that mTORC1 is involved in alternative polyadenylation (APA) and promotes transcriptome-wide APA in 3'-untranslated regions (3'-UTRs), affecting diverse cellular pathways including ubiquitin-mediated proteolysis and ER stress responses [18,19]. These studies suggested that mTORC1 may function in the processing of pre-mRNA in addition to well-characterized roles in various cellular pathways.

In this study, we investigated the transcriptome changes upon mTORC1 activation and found that the stoichiometry of U2AF1 isoforms is drastically altered by cellular mTORC1 activity. We further delineated the functional differences of duplicated tandem exon-derived U2AF1 isoforms by taking integrative approaches of genome-editing and

high profiling methodologies. Unlike previous suggestions [9], our unbiased approaches revealed that U2AF1 isoforms contribute differentially to transcriptome changes by alternative splicing and affect protein synthesis by regulating 5'-UTR alternative splicing.

Materials and Methods

Cell lines and Cell culture

WT MEF and TSC1$^{-/-}$ MEF cells were obtained from Dr. Kwiakowski at Harvard University and they were previously described [16,18,19]. WT, TSC1$^{-/-}$ MEF, HEK293, HeLa and MDA-MB231 cells were cultured in Dulbecco's Modified Eagle Medium (Gibco, USA) with 10% (v/v) fetal bovine serum (FBS) and 100g/ml streptomycin and 100U/ml penicillin at 37C in 5% CO_2.

Construction of CRISPR/Cas9-sgRNA plasmids for U2AF1 tandem exon knockout

The target gRNA sequences were identified by crispor.tefor.net and chopchop.rc.fas.harvard.edu. The guide sequences were cloned into Addgene plasmid #48138 using the following primers. Targeting exon 3b, 5' end forward 5'-CACCgttgaatcaagatggtctgcg-3'reverse 5'-AAACcgcagaccatcttgattcaaC-3'; 3' end forward 5'-CACCgcacactgtaagtcccacagt-3' reverse 5'-AAACactgtgggacttacagtgtgc-3'. Targeting exon 3a #1, 5' end forward 5'-CACCgagaggtgtccccttagttgg-3' reverse 5'-AAACccaactaaggggacacctctg-3'; 3' end forward 5'-CACCgagttcagatctcgaggtgag reverse 5'-AAACctcacctcgagatctgaactc-3'. Targeting exon 3a #2, 5' end forward 5'-CACCgctgggctggcacttagcag-3' reverse 5'-AAACctgctaagtgccagcccagc-3'; 3' end forward 5'-CACCgggcaggagttcagatctcg-3' reverse 5'-AAACcgagatctgaactcctgccc-3'.

RNA sequencing and analyses

Poly(A+) RNAs were isolated from U2af1**a-** and U2af1**b-**only#1 cell lines treated with control or U2af1 targeting siRNA were sent out for paired end reads RNA-seq analysis. A total of 84,452,901 reads for U2af1a only#1 control cells, 91,886,993 reads for U2af1a only#1 siRNA cells, 84,722,415 reads for U2af1b only#1 control cells, and 86,708,880 reads for U2af1b only#1 siRNA cells were produced from Hi-Seq pipeline with length of 51bp of each end. The short reads were aligned to mm10 reference genome by TopHat, with up to two mismatches allowed. 93.4% of paired short reads from U2af1a only#1 control, 93.5% reads from U2af1a only#1 siRNA, 93.0% reads from U2af1b only#1 control, and 93.9% reads from U2af1b only#1 siRNA were mapped to the reference genome for alternative splicing analysis in the study.

AS-Quant for the analyses of alternative splicing events

AS-Quant first applies rMATS [20] to categorize potential alternative splicing events into four categories (cassette type, SE; mutually exclusive, MXE; alternative 5'-splice site, A5SS; alternative 3'-splice site, A3SS) based on the UCSC annotation. Then, for each categorized potential alternative splicing event, the mean short read coverages of the affected exon and the rest of exons in the transcript are measured, and we denote them as n and N based on the above context using RNA-seq alignment file. Next, a canonical 2 x 2 Chi-squared test is applied to report a p-value for each candidate event based on the n/N ratios in two cases. The candidate alternative splicing events with p-value ≤ 0.1 and the ratio difference larger than 0.1 between the two cases are considered for further analyses.

Western blotting

Antibodies used in this study include: anti-U2AF1 (ab86305, Abcam), anti-U2AF2 (sc-48804, Santa Cruz Biotechnology), anti-hnRNP A1 (4B10, Abcam), anti-RPS6 (#2317, Cell Signaling), and anti-pRPS6 (#2211, Cell Signaling), anti-TUBULIN (sc-53646, Santa Cruz Biotechnology), anti-SRSF3 (sc-13510, Santa Cruz Biotechnology), anti-EIF4EBP1 (#9452, Cell Signaling), anti-HNRNPC1/C2 (ab10294, Abcam), anti-Flag (F3165, SigmaAldrich).

Minigene reporter assay

Minigene U2af1 reporter gene fragment was amplified at genomic region from the start of exon 2 to the end of exon 4 with forward primer 5'-GCCATGGATCCAGTCAACTGTTCATTTTATTTC-3' and reverse primer 5'-.ATATTAGAGCGGCCGCCTCAAAGAACTCATCATAG-3'. The fragments were then digested with BamHI and NotI, then ligated into pcDNA3.1 (+) plasmid (Thermo Fisher Scientific).

CRISPR-induced homologous recombination

To insert a 3x-Flag tag at the C-terminal of *U2af1* gene via homologous recombination, the donor sequence was synthesized as a gBlocks gene fragment (IDT) and cloned into pAAV-SEPT-acceptor vector (Addgene). Type IIS restriction enzyme BspQI was used for the cloning. To induce efficient homologous recombination near the C-terminal locus of *U2af1*, a double-stranded break was created by CRISPR/Cas9 cloned into

PX458 (GFP+) using primer sequences as follows: forward 5'-CACCGACACACGGTAAAAAGGGCT-3' reverse 5'-AAACAGCCCTTTTTACCGTGTGTC. The two plasmids were co-transfected into U2af1**a**-only#1 and U2af1**b**-only#1 cell lines. Top 20% GFP+ cells were isolated for further screening by flow cytometry. Edited clones were confirmed by PCR of genomic DNA and Western blot analysis.

Immunoprecipitation and mass spectrometry

Cells were fixed and crosslinked with 0.2% formaldehyde for 10 minutes at room temperature and quenched with 0.15 M glycine pH 7.5, then washed with PBS twice. The pellet was resuspended with lysis buffer (25mM Tris pH7.4, 300mM NaCl, 2.5mM MgCl2, 0.5% Empigen; 0.5% NP-40) and sonicated with microtip for 4 times, 10 seconds each at 4W, and spun down at maximum speed for 10 min. The supernatants were incubated with anti-Flag M2 magnetic beads (Sigma-Aldrich, M8823) in an end-over-end rotator at 4°C for 2 hours. The beads were washed five times with lysis buffer. The protein complexes were eluted with 5 volumes of beads with 3x flag peptide (150 ng/ul). The elutions were precipitated with acetone and resuspended with 5x SDS sample buffer by boiling for 15 minutes. Samples were run on a 10% criterion gel. The gel was fixed with 40% ethanol and 10% AcOH and washed with ddH_2O. Lastly, the gel was stained with imperial stain and the stained areas were cut out for in-gel trypsin digestion. The gel pieces were washed with 50% ethanol twice for 2 hours and 16 hours with mixing. The gel pieces were washed twice with water for 10 minutes with mixing and then cut into mm^3 size pieces. After drying

with 100% acetonitrile, the gel pellets were reduced and alkylated with 10mM tris(2-carboxyethyl)phosphine and 40mM iodoacetamide respectively. The gel was washed with 50% acetonitrile and 50mM ammonium bicarbonate for 5 minutes and dried in speedvac. The gel pieces were digested with 0.1ug Trypsin overnight at 37C with rotation and extracted twice with 50% acetonitrile with 1% TFA and 100% acetonitrile. The peptide solution was dried with speedvac and desalted with C18.

LC-MS/MS Measurement

LC-MS/MS analysis was performed using a Proxeon Easy nLC 1000 Nano-UPLC system coupled with an Orbitrap Fusion mass spectrometer (ThermoFisher). Peptide samples were loaded onto custom packed C18 column (15 cm × 75 μm, ReproSil-Pur Basic C18, 2.5 μm, Dr. Maisch GmbH) and eluted for 2 hours using a 5-32% gradient of HPLC solvent B (0.1% formic acid in acetonitrile, v/v) and a flow rate of 200 nl/min. Fusion Orbitrap was operated in data-dependent mode. Survey scan MS were acquired with the orbitrap with a 380-1580 m/z range and a resolution of 60000. Ions were selected by using dynamic exclusion of 15 seconds, an intensity threshold of 1.0E4 and charge states of 2-7. The top 12 most intense ions per survey were selected for CID fragmentation and ion trap analysis

Raw Mass Spectrometry Data Processing

Raw mass spectrometry files were processed by MaxQuant (version 1.5.3.12) for database search and quantitative analysis. Cysteine carbamidometylation was set as a fixed modification and methionine oxidation and protein N-terminal acetylation were set as

variable modifications. The proteolytic enzyme was set as trypsin with a maximum of two missing cleavages. The data was searched against the UniProt mouse database (downloaded at 2013/09/27 with 43310 sequences), and we used a cutoff threshold setting at 1% False Discovery Rate (FDR) at protein and peptide levels. The precursor ion tolerance was set to 4.5 ppm and the fragment ion mass tolerance was set to 0.5 Da. The MaxLFQ algorithm provided by MaxQuant was selected for the label-free relative quantification of the samples. To perform relative quantification, the LFQ metrics were extracted from the MaxQuant-processed data and processed for statistical analysis using the Perseus software (version 1.5.5.1). Multiple hypothesis testing was performed using two-sided Student t-test and permutation-based FDR correction. The FDR was set at 5% and the S0 variance correction constant was set at 0.1 for all comparisons.

Co-immunoprecipitation

RSB-100 buffer (25 mM Tris, pH 7.4; 100 mM NaCl, 2.5 mM MgCl2; 0.02% Triton-X-100) was used as binding buffer for co-immunoprecipitation experiments. Nuclear extract of HEK293 cell line was prepared according to the REAP method [21]. Nuclear pellet was resuspended with RSB-100 and sonicated twice, 5 seconds each at 1 W. After 30 seconds spin-down at 8000g, the nuclear extract was incubated with protein G sepharose beads immobilized with anti-HNRNPA1, anti-HNRNPC1/C2 or anti-Flag antibodies for 1 hour at 4 degree on an end-over-end rotator. Beads were then washed three times. Elution was done by adding 4x SDS sample buffer to the beads followed by 10 minute-boiling. Elutions were run on 12.5% SDS-PAGE for Western blot analysis.

qPCR and RT-PCR

Total RNAs were isolated using Trizol reagents according to manufacturer's protocol. The extracted RNAs were reverse transcribed into cDNA using oligo-d(T) or random hexamer priming and superscript III (Thermo Fisher Scientific) from standard protocol supplied by the manufacturer. For qPCR, cDNA templates were amplified and the Ct values were quantified in real time using Eva Green or Taqman probes where indicated. Normalization of the Ct values were performed for relative quantitation. Absolute quantitation was made where indicated. Primers and Taqman probes used in qPCR assays include: *U2af1a* Taqman probe, 5′-(FAM)- TTTAGCCAGACCATTGCCCTCTTGA –(BHQ-1)–3′. *U2af1a* forward primer, 5′- ATGGCGGAATACTTGGCCTC -3′; reverse primer, 5′- GTCAGCAGACTGGGAAGAGT -3′; *U2af1b* Taqman probe, 5′-(FAM)-ACGGCTCACACTGTGCTGTGAGCGA-(BHQ-1)-3′. *U2af1b* forward primer, 5′-ATCGTAATCCCCAAAACAGTGC -3′; reverse primer, 5′-AGACTTCCTCAAAGAACTCATCAT -3; *Anapc10* forward 5'AAGCAGTTGGAGAGGACAGC-3' reverse 5'-ACCCTGGTTTGCAGGAAGAG-3'; *Hnrnph2* forward 5'-CACAGGGGAAGCTTTTGTGC-3' reverse 5'-GGACTTCAGCTCGGCTACTC-3'; Srsf1 forward 5'-ATCTCACGAGGGAGAAACTGC-3' reverse 5'- GTAACTGCGACTCCTGCTGT -3'; Srsf2 forward 5'- GCCCGAAGATCCAAGTCCAA-3' reverse 5'-TGGACTCTCGCTTCGACAC-3'; Srsf3 forward 5'- GCTGCCGTGTAAGAGTGGAA -3' reverse 5'- AGGACTCCTCCTGCGGTAAT -3'; Srsf4 forward 5'-AGCCGCAGTAAGAAGGAGAAA -3' reverse 5'- GTCCTCGGCGTGGTCTTTA -

3';Srsf5 forward 5'- AGGTCAAGAAGCAGGTCACG-3' reverse 5'-TCGGCTGTAAGACTTGCTCC-3';Srsf6 forward 5'-GTCTCGGAGCAAAGGTCGAT-3' reverse 5'- CTTGAGTGGGAATGGGAGCC-3';Srsf7 forward 5'- TGCAGAGGATGCAGTTCGAG-3' reverse 5'-GGGCAGGTGGCCTATCAAAA-3';Srsf9 forward 5'-TCACGAGGGTGAGACTTCCT-3' reverse 5'-GACCGCGACCGTGAGTAG-3'; Srsf10 forward 5'- TCTCGAAGCCGGAGTTATGA-3' reverse 5'-AGTCGGTCTACTGTTTCTAGGACT-3'; Srsf11 forward 5'-GATCTCGCTCGAGGAGGAGG -3' reverse 5'- TGGATTTGGAGTGTGACCGC -3'; Srsf12 forward 5'- GAAATCACAGTCACGCTCGC-3' reverse 5'-CTCTGGGAGACTTGCATGGG-3';Cpsf1 forward 5'-ACATACCGACGCTTGCTGAT-3' reverse 5'- TAGCGGTTTAGCAGTTCCCC-3'; Cpsf2 forward 5'- CGGAATTTGTAGGGGGCGTA-3' reverse 5'-ATCCGATGCGTCCAGTTTCT-3'; Cpsf3 forward 5'-GCACGTTTACAGCAAGAGGC-3' reverse 5'- TTCTACAGCCCGAGTCTCCA-3'; Cpsf4 forward 5'- GCACCCTCGATTTGAACTGC-3' reverse 5'-CTGCATGACCCCAATGACCT-3'; Cpsf5 forward 5'-AAGCCTTGTTTGCAGTCCCTA-3' reverse 5'- AATGATGGGTCCATACCCCG-3'; Cpsf6 forward 5'- TCACGGGAAAAGAGTCGTCG-3' reverse 5'-CGGTATTCTCGCTCTCGGTC-3'; Cpsf7 forward 5'-TGATTCTGCTGATGGACGGG-3' reverse 5'- GGCAGACCCATTAGGGGAAG-3'.

For RT-PCR, cDNAs were amplified using primer sets listed below and the PCR products were subjected to electrophoresis in 2% agarose gel with ethidium bromide for visualization of amplified DNA fragments. Primer sets used in RT-PCR assays are as follows: *Anapc10* forward 5'-GAACCGGAATTGTGGCGAATC-3' reverse 5'-GGAGGTGTCTTGTTCGGTGT-3'; *Anapc10* alternative tss forward 5'-GCTGTCCTCTCCAACTGCTT-3' reverse 5'-TGCTGTCTCCTCAGGCTTTG-3'; *Hnrnph2* forward 5'-GGTCGTCGTCTATCGTCTCG-3' reverse 5'-AGCTTGGCTCAATGCAAATTC-3'; *Serpinh1* forward 5'-CTGTCTGAGGAGCGATTGCC-3' reverse 5'-CAAGAGGCATAAGGTGCCCA-3'; *Gng12* forward 5'-GGGAAGGACTTTGGGGTGAG-3' reverse 5'-CTATGCTGTTGGTGCTTGCC-3'; *Pcbp2* forward 5'-TTGACCAAGCTGCACCAGTT-3' reverse 5'-TTGATTTTGGCGCCTTGACG-3'; *Ktn1* forward 5'-AGCTGACGAGTCTCAAAGGA-3' reverse 5'-CACGTAAGTCGATCGCTCCAT-3'; *Tial1* forward 5'-TCAGTCAGATCGGACCCTGT-3' reverse 5'-AGCAGCTGCATCTCTGTGTT-3'; *Puf60* forward 5'-TGCAATGGAGCAGAGCATCA-3' reverse 5'-ATGCTCTTGATGGGGCCAAA-3'; *Pex2* forward 5'-ATGTCCACAGGATCCATGCC-3' reverse 5'-TGGCTCAAAGCGAGCTAACA-3'.

Polysome fractionation

Isolation of polysome fractions from total cell lysates using sucrose gradient was carried out as previously described [18,19]. Briefly, cells were lysed in the polysome buffer

(20 mM Tris pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 100 mg/ml cycloheximide and 1% Triton X-100). Cell extracts were loaded onto sucrose gradient (5–45%). Fractionation was done by centrifugation at 190 000 × g for 2 hour at 4°C. Twelve fractions were collected for the analysis. Amounts of mRNAs in each fraction were calculated using absolute quantitation. Ten per cent (v/v) of total RNAs in each fraction were used for RT-qPCR.

Luciferase construct and assay

5'-UTRs with or without the alternative exons of *Hnrnph2*, *Anapc10*, *Pex2* and *Cwc22* were cloned into psiCheck1 (kindly provided by Dr. Aaron Goldstrohm at the University of Minnesota Twin Cities [22]). These recombinant luciferase reporters were transfected into $TSC1^{-/-}$ MEFs using Lipofectamine 3000 (Thermo Fisher Scientific). 18 hours after the transfection, the luciferase activity was measured using Dual-Glo reagent with the Glomax Discover luminometer (Promega). Expression of Renilla luciferase mRNA was measured to normalize the luciferase activity by RT-qPCR using the following primers: forward 5'- TCTCGTTAAGGGAGGCAAGC-3' reverse 5'-TGGAAAAGAACCCAGGGTCG-3'. Four replicates of the measurement were conducted for technical repeats.

Accession code

The accession number for the RNA-Seq data in this study is SRP215854.

Results

Mutually exclusive expression of *U2af1* isoforms is associated with cellular mTORC1 activity

To better understand how mTORC1 contributes to transcriptome changes, we analyzed our previous RNA-Seq data from WT and TSC1$^{-/-}$ MEFs [19] by focusing on the changes of gene expression in RNA-binding proteins (RBPs). Among the RBPs whose transcript levels changed upon mTORC1 activation, *U2af1* was particularly interesting: one of the two *U2af1* isoforms, *U2af1a* (uc008bvo.2, NM_024187.4), shows a ~2-fold difference in TSC1$^{-/-}$ MEFs while the *U2af1b* (uc012aov.1, NM_001163769.1) expression remained unchanged between WT and TSC1$^{-/-}$ MEFs (Fig. 1A-B), Supplementary Fig. 1A-D). This observation suggests that the biased expression of *U2af1a* isoform is associated with cellular mTORC1 signaling. To test this idea, we incubated TSC1$^{-/-}$ MEFs in Earle's balanced salt solution (EBSS) to reduce the cellular mTORC1 activity, followed by re-activation of the mTORC1 activity by incubating cells in serum-containing Dulbecco's modified eagle media (DMEM) (Fig. 1C, Supplementary Fig. 1E). In this experiment, while the level of *U2af1b* transcript remained unchanged, the expression of *U2af1a* was selectively increased (Fig. 1C). Together, these data show that the mTORC1 signaling pathway regulates the mutually exclusive alternative splicing of duplicated tandem exons in *U2AF1* expression.

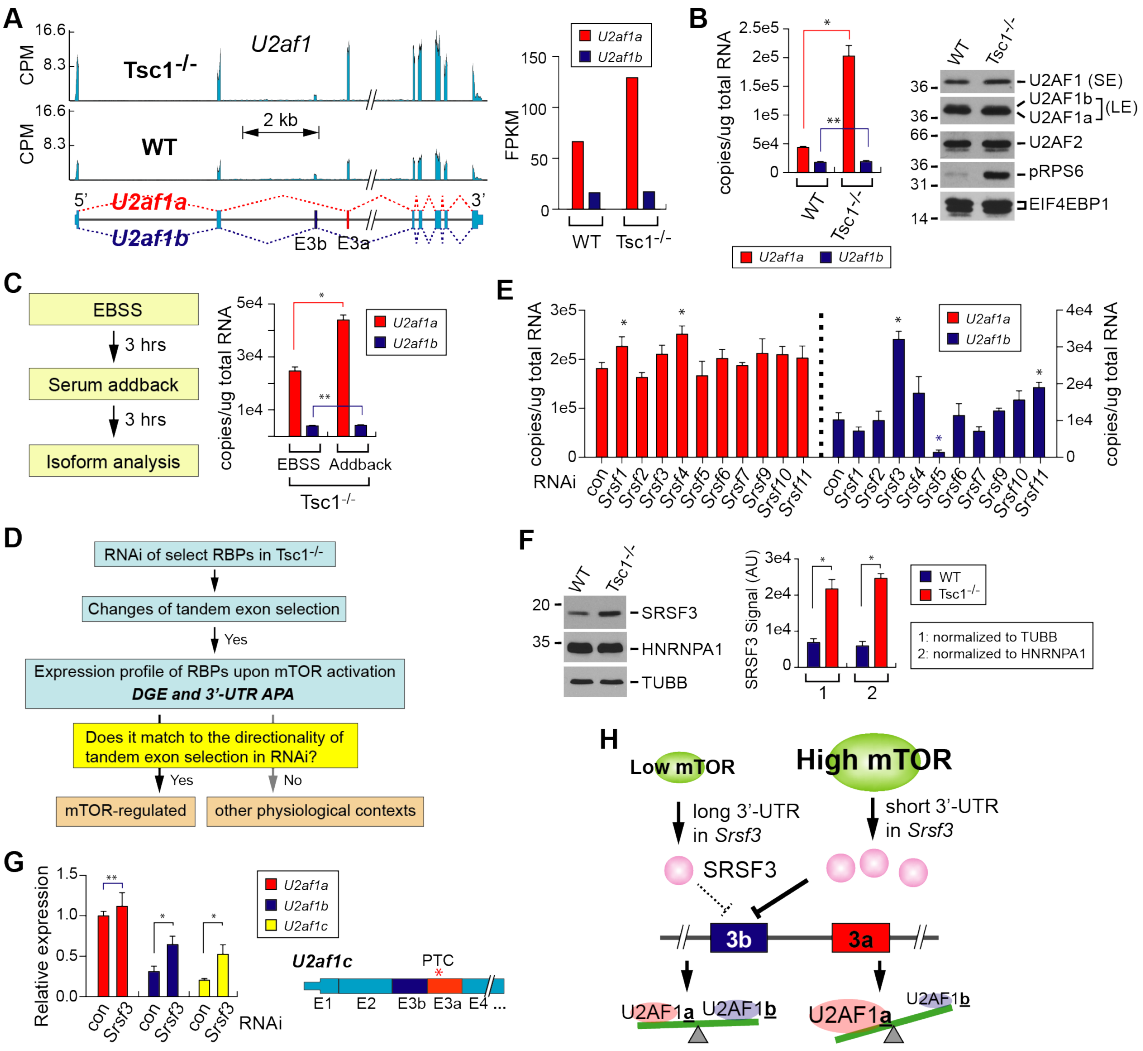


Figure 1. Cellular mTORC1 activity affects the expression profile of *U2af1* isoform.

(A) *U2af1a* is selectively up-regulated upon mTORC1 activation. (left) RNA-Seq reads alignments of *U2af1* isoforms in WT and *Tsc1*$^{-/-}$ transcriptomes. (right) Quantitation of *U2af1* isoforms in the RNA-Seq data from WT and TSC1$^{-/-}$ MEFs. **(B)** Expression of *U2af1* isoforms in WT and TSC1$^{-/-}$ MEFs was measured by Taqman qPCR with absolute quantitation. The data are presented as the mean (SD) (**p*=1.3e-6 , ***p*=0.60; two-tailed Student's *t* test, n=3 for technical repeats). Western blot analyses of U2AF1 isoforms were done using total cellular extracts from WT and TSC1$^{-/-}$ MEFs. Please note that U2AF1b

isoform is only visible in the longer exposure blot. Phospho-S6 (pRPS6) probing is for the validation of mTORC1 activation. SE and LE indicates short exposure and long exposure in western blot, respectively. **(C)** *U2af1a* is selectively up-regulated upon the activation of cellular mTORC1 signaling. (left) A workflow of the serum add-back experiment for the manipulation of cellular mTORC1 activity. (right) Absolute quantitation of *U2af1* isoforms by Taqman qPCR. The data are presented as the mean (SD) (**p*=1.6e-4, ***p*= 0.43; two-tailed Student's *t* test, n=3 for technical repeats). **(D)** A workflow of the screening strategy for mTORC1-regulated splicing factors that regulate the *U2af1* isoform expression. DGE, differential gene expression at the transcript level. **(E)** An RNAi screen to identify a regulator(s) of *U2af1* alternative splicing. SR splicing factors were knocked down by siRNAs in $TSC1^{-/-}$ MEFs and the expression of *U2af1* isoforms was measured by Taqman qPCR assay with absolute quantitation. Asterisks denote statistically significant changes of *U2af1* isoform expression upon the RNAi knockdown. The data are presented as the mean (SD) (**p*< 0.0086; two-tailed Student's *t* test, n=3 for technical repeats). **(F)** Western blot analysis of SRSF3 in WT and $TSC1^{-/-}$ MEFs. TUBULIN and HNRNPA1 were used as loading controls. Quantitation by ImageQuant software of the SRSF3 signals normalized to TUBULIN or HNRNPA1 is shown on the right. (**p*< 7.5e-4; two-tailed Student's *t* test, n=3 for biological repeats; see Supplementary Fig. 1N for the other two repeats). **(G)** Relative expression of *U2af1a*, *U2af1b* and *U2af1c* transcripts (structure shown on right; PTC, premature termination codon) upon RNAi knockdown of *Srsf3*. Puromycin was added for 8 hours at the concentration of 5 μg/ml. The data are presented as the mean (SD)

(*p<0.010, **p=0.31; two-tailed Student's t test, n=3 for technical repeats). **(H)** A proposed model for regulation of *U2af1* tandem exon splicing by mTORC1 and SRSF3.

U2AF1a-polarized expression has been found in tissues and cell lines previously [8] and a weak branch point consensus sequence upstream of exon 3b has been proposed to be the reason for this observation [23]. If this is the case, however, the increase of *U2af1* expression by mTORC1 as shown in Fig. 1C, would also accompany an increase in *U2af1b* expression, although with a lower degree. Thus, our results suggest that upon mTORC1 activation, additional active suppression mechanism(s) for exon 3b inclusion and/or promotion mechanism(s) for exon 3a inclusion exists. To test this idea, we performed an siRNA-mediated screen targeting various splicing regulators to identify potential mTORC1-regulated factors that regulate *U2af1* isoform expression. After following the screening strategy presented in Fig. 1D, which takes into account the effects of these RBPs on the splicing of *U2af1* tandem exons upon knockdown and their expression profile changes in response to mTORC1 activation, *Srsf3*, *Srsf5*, and *Cpsf5* emerged as candidates of mTORC1-regulated factors for *U2af1* isoform expression regulation (Fig. 1E and Supplementary Fig. 1A-B, F-J). *Srsf3* was of our particular interest because, as opposed to *Srsf5* and *Cpsf5*, it not only passed our screening criteria (Fig. 1D) but also displayed 3'-UTR shortening by APA, a recently characterized post-transcriptional signature in the mTORC1-activated transcriptome (Supplementary Fig. 1K-M) [19]. The *Srsf3* knockdown significantly increased the expression of *U2af1b* about ~3 fold while *U2af1a* transcript level was relatively unaffected, suggesting that SRSF3 has a suppressive role on the inclusion of exon 3b (Fig. 1E and Supplementary Fig. 1F). Consistent with our previous findings on the role of 3'-UTR shortening in the promotion of protein synthesis [18,19], polysome profiling of *Srsf3* transcripts and western blot analysis showed that the SRSF3

protein level significantly increased due to the 3'-UTR APA in TSC1$^{-/-}$ compared to WT MEFs. (Fig. 1F and Supplementary Fig. 1N, O). A further analysis of *U2af1* transcript variants showed that the knockdown of *Srsf3* coupled with the inhibition of nonsense-mediated mRNA decay by puromycin treatment increased the expression of *U2af1c* transcript (not annotated in mouse mm10, the same structure of the transcript is annotated as NM_001025204 in human hg38) which contains both exon 3 a and 3b (Fig. 1G) [8,24,25]. These results suggest that the exon 3b inclusion is actively suppressed, while the exon 3a selection is constitutively active. We next made a reporter construct containing a genomic DNA fragment of *U2af1* gene ranging from exon 2 to exon 4 (Supplementary Fig. 1P). We then manipulated the cellular level of SRSF3 in TSC1$^{-/-}$ MEFs harboring the reporter construct by transient overexpression and used qPCR to measure the selection of *U2af1* tandem exons from the reporter construct. Consistent with the measurements for endogenous *U2af1* isoform expression, the overexpression of SRSF3 in the TSC1$^{-/-}$ MEFs significantly reduced the inclusion of exon 3b from the reporter construct (Supplementary Fig. 1Q). Together, these results identify SRSF3 as one of the factors that contribute to *U2af1* tandem exon splicing by mTORC1 signaling and establish a regulatory pathway of U2AF1 isoform expression: a transcriptional activation of *U2af1* gene upon the mTORC1 activation constitutively selects exon 3a for its splicing and expression while suppressing the inclusion of exon 3b and drives the biased expression of U2AF1a (Fig. 1H).

U2AF1 isoform-specific transcriptomes display an overlapping but distinctive alternative splicing profile

Although two isoforms are produced from *U2AF1*, most studies on U2AF1 and its mutations do not distinguish functional differences that might be conferred by these isoforms [6,14,26–36]. However, our above findings indicate that the stoichiometry of U2AF1 isoforms may change depending on cellular contexts. In fact, western blot analyses of U2AF1 showed that the stoichiometry of U2AF1 isoforms is dynamic across the tested cell lines, confirming that the cellular contents of U2AF1 isoforms are diverse and further suggesting that U2AF1a isoform cannot be presumed to be predominantly expressed in every biological or cellular model (Supplementary Fig. 2A). To comprehensively understand how U2AF1 isoforms differentially contribute to the transcriptome, we adopted the CRISPR/Cas9 genome editing tool to create cell lines that only express one of the two U2AF1 isoforms in TSC1$^{-/-}$ MEF background. To separately knockout each of the *U2af1* isoforms, we designed pairs of CRISPR/Cas9 constructs that could create double-stranded breaks flanking one of the tandem exons to induce non-homologous end joining (NHEJ), leading to the removal of the targeted exon. The resulting *U2af1* locus would only have one usable exon 3, exon 3a or exon 3b, achieving the creation of U2af1**a**- or U2af1**b**-only cell lines (Fig. 2A (left panel) and Supplementary Fig. 2B). The exclusive expression of one isoform is confirmed by RNA-Seq and western blot analyses (Fig. 2A (right panel) and B). We selected two clones of U2af1**a**- and U2af1**b**-only TSC1$^{-/-}$ MEF cell lines for future experiments to avoid artifacts from clonal variations. Of note, CRISPR/Cas9 genome editing did not drastically alter the overall expression of U2AF proteins, nor affect the

cellular mTORC1 activity, as evidenced by western blot analyses on U2AF1, U2AF2 and phospho-S6 (Fig. 2A and B).

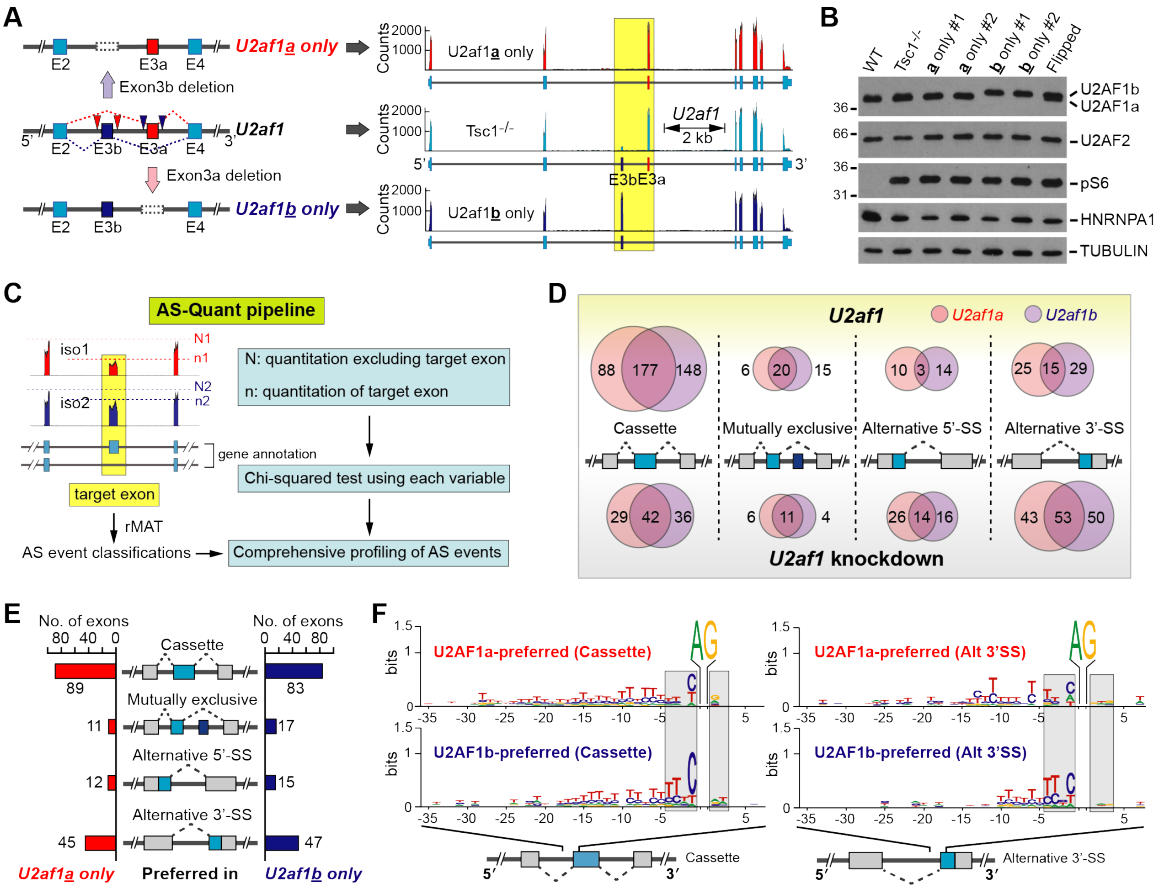


Figure 2. U2AF1 isoforms display distinctive alternative splicing profiles. (A) (left) Schematic for the generation of U2AF1 isoform-specific cell lines in TSC1$^{-/-}$ MEFs. Location of guide RNA (gRNA) pairs to produce U2af1**a**-only and U2af1**b**-only cells are indicated by red and blue triangles, respectively. (right) RNA-Seq read alignments of *U2af1* gene locus in U2af1**a**-only, U2af1**b**-only and control TSC1$^{-/-}$ MEFs. The yellow box highlights tandem exon regions in *U2af1*. **(B)** Western blot analyses of U2af1**a**-only, U2af1**b**-only, control TSC1$^{-/-}$, and WT MEFs. Exon 3a targeting experiment created several heterozygous clones, which were named as "flipped" since the U2AF1a/U2AF1b

ratio is flipped compared to control TSC1$^{-/-}$MEFs. A flipped clone is also loaded to aid visualizing the migration shift of U2AF1 isoforms. Two **a**-only and **b**-only cell lines were analyzed. **(C)** Schematic of custom-developed AS-Quant (***A***lternative ***S***plicing ***Quant***itation) pipeline for a quantitative analysis of alternative splicing. **(D)** Types of alternative splicing events dependent on the cellular level of U2AF1a or U2AF1b isoform. *U2af1* knockdown-dependent alternative splicing events are categorized and the numbers of events identified in U2af1**a**-only cell line (orange circles), U2af1**b**-only cell line (purple circles), and in both cell lines (overlapped regions) are presented. (upper) Number of alternative exons that are more included in the presence of *U2af1*. (lower) Number of alternative exons that are more included in the absence of *U2af1* (upon knockdown). **(E)** Types of alternative splicing events preferred by U2AF1a or U2AF1b. Alternative splicing events identified by a direct comparison between U2af1**a**- and U2af1**b**-only cell lines are presented. Alternative splicing events are categorized and the number of exons that are preferentially included in U2af1**a**-only (left) and U2af1**b**-only cell line (right) are shown. **(F)** The frequency of upstream nucleotides of the 3'-splice site of cassette type (left) or alternative 3'-splice site type (right) preferred by U2AF1 isoforms. The certainty (bit = log(frequency/2.4), ranging from 0 to 1.5) of nucleotides in each position of the upstream intron and the downstream exon regions from the AG dinucleotide of the 3'-splice site is illustrated. X-axis denotes the position of upstream and downstream nucleotides from the AG dinucleotide and Y-axis represents the certainty of the nucleotides.

To examine the transcriptome-wide changes of gene expression by U2AF1 isoforms, we performed RNA-Seq experiments using the #1 clone of U2af1**a**- and U2af1**b**-only cell lines in the presence or absence of *U2af1* knockdown. The knockdown of *U2af1* in these cell lines did not affect the level of U2AF2 (Supplementary Fig. 2C). The analyses of RNA-Seq data from corresponding cell lines were focused on alternative splicing events using our custom-developed AS-Quant (*A*lternative *S*plicing *Quant*itation; Fig. 2C) pipeline. AS-Quant first applies rMATS [20] to categorize potential alternative splicing events into four categories (cassette type or skipped exon, mutually exclusive exons, alternative 5'-splice site, and alternative 3'-splice site) based on the mm10 UCSC mouse genome annotation. Then the quantitation of affected exons in a transcript compared to the rest of exons in the transcript is further tested by the Chi-squared method to determine the alternative inclusion/exclusion of the tested exon between the two cases (Fig. 2C).

Since a previous report showed that U2AF1a has a much broader impact on alternative splicing than U2AF1b [9], we asked whether U2AF1 isoforms have different capacities in alternative splicing. To this end, alternative splicing events in the RNA-Seq data of the control vs. *U2af1* knockdown in U2af1**a**- or U2af1**b**-only cell line were analyzed using AS-Quant. Overall, we identified 568 exons in 451 genes to be U2AF1a-dependent alternative splicing events in U2af1**a**-only cell line, and 647 exons in 501 genes as U2AF1b-dependent alternative splicing events in U2af1**b**-only cell line; out of a total of 880 identified alternative splicing events, 335 of these exons are common in the two datasets (Fig. 2D). Notably, cassette and mutually exclusive type of alternative splicing showed a much higher overlap than alternative 5'-splice site and 3'-splice site events

between the two isoforms (Fig. 2D). Moreover, in both knockdown experiments, the presence of U2AF1 is crucial for the exon inclusion in cassette type alternative splicing (upper far left of Fig. 2D) and alternative uses of 3'-splice sites (lower far right of Fig. 2D), supporting the suggested role of U2AF1 in exon inclusion/definition and 3'-splice site definition upon splicing (Supplementary Table 1-2) [28,37,38]. Collectively, these results show that the two U2AF1 isoforms have similar capacities to function as alternative splicing regulators but they appear to have different specificities.

Therefore, to further examine the functional differences of U2AF1 isoforms in alternative splicing regulation, we then sought to identify alternative splicing events that are differentially regulated by U2AF1 isoforms by directly comparing the RNA-Seq data from U2af1**a**- and U2af1**b**-only cell lines with AS-Quant. In this case, the data from cells expressing comparable levels of each U2AF1 isoform are directly compared, without considering the data from knockdown experiments. This approach excludes alternative exons that are redundantly regulated by U2AF1 isoforms or other RBPs and only reveals exons that are differentially regulated by the splicing machineries only harboring U2AF1a or U2AF1b. In this analysis, we identified 157 exons in 139 genes and 162 exons in 142 genes that are preferentially included in U2af1**a**- or U2af1**b**-only cells, respectively (Fig. 2E and Supplementary Table 3). That is, 157 exons are more preferentially included and 162 exons are more preferentially skipped in an U2af1**a**-only environment compared to an U2af1**b**-only environment. To validate the alternative splicing events identified by AS-Quant, we randomly selected alternative splicing events and quantified these splicing events by RT-PCR using total RNAs purified from the two clones of U2af1**a**- and U2af1**b**-

only cells (Supplementary Fig. 2D). All tested alternative splicing events showed splicing patterns consistent with the RNA-Seq data analyses using AS-Quant. In addition, all tested genes showed similar splicing pattern changes within the U2af1**a**- or U2af1**b**-only clones, demonstrating that these alternative splicing events indeed show U2AF1 isoform preferences and did not arise from clonal variations (Supplementary Fig. 2D).

Although U2AF1 has been indicated to bind to the consensus AG dinucleotide motif in the 3'-splice site, our analyses of U2AF1 isoform-dependent alternative splicing events strongly suggest that each U2AF1 isoform prefers additional distinct sequence contexts around the 3'-splice site and renders the specificity in exon choice for the splicing reaction. Therefore, we analyzed the nucleotide frequency surrounding the 3'-splice site of the cassette and alternative 3'-splice site type splicing events based upon U2AF1 expression as well as U2AF1 isoform preferences (-35 ~ +5 bp relative to the AG dinucleotide of 3'-splice site). In line with the well-established role of U2AF complex in splicing, exons whose inclusions are commonly promoted by U2AF1a and U2AF1b have a strong poly-pyrimidine tract frequency downstream of the -20 position compared to exons not promoted by U2AF1 (Supplementary Fig. 2E and F). And interestingly, while sharing the feature of prominent poly-pyrimidine tract, nucleotide frequencies upstream of the 3'-splice site of U2AF1a-preferred exons have different sequence signatures compared to that of U2AF1b-preferred exons. Specifically, for both cassette and alternative 3'-splice site types, there is a higher frequency of C at the -1 position for U2AF1b-preferred exons. U2AF1b-preferred exons also have a stronger T preference at -3 and -4 positions (Fig. 2F and Supplementary Fig. 2E). In contrast, U2AF1a is less selective in the sequence

preference at these positions and shows a weaker preference to polypyrimidine tracts compared to U2AF1b (Fig. 2F and Supplementary Fig. 2E). These analyses indicate that the two types of U2AF dimers consisting of two different U2AF1 isoforms have distinct nucleotide-binding preferences at the splice site. Together, these data provide evidence that the two isoforms have comparable involvements in the general splicing mechanism, yet a subset of alternative exons are differentially regulated by the two isoforms, demonstrating the functional differences between U2AF1a and U2AF1b in alternative splicing.

U2AF1 isoform stoichiometry is a mechanistic factor for mutually exclusive alternative splicing

As shown above and in other studies [13,39–42], genes with tandem duplicated exons can produce highly similar isoforms with distinct functions. Therefore, it is important to understand how the mutually exclusive tandem duplicated exons are processed, which can involve more dynamic reorganization of RBPs and *cis*-acting sequence elements in introns and exons compared to other types of alternative splicing. Intriguingly, AS-Quant identified a number of tandem duplicated mutually exclusive alternative splicing events to be differentially regulated by U2AF1 isoforms. Although previous reports suggest that U2AF1 function in alternative splicing of several mutually exclusive duplicated tandem exons, mechanistic insights of these alternative splicing events are largely lacking [23,36]. Moreover, these studies did not consider the relevance of the functional differences of U2AF1 isoforms in the regulation of mutually exclusive alternative splicing events.

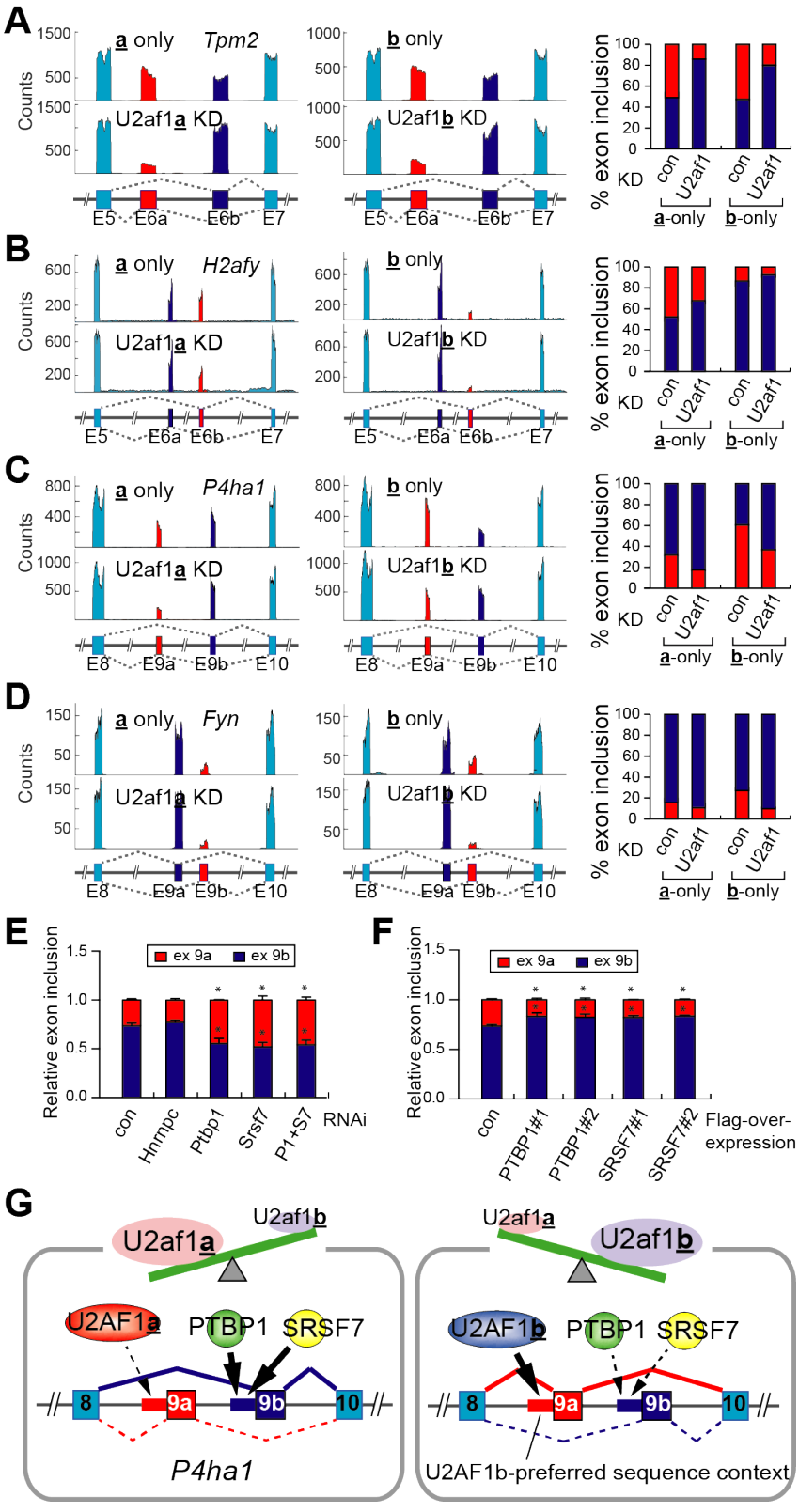

A
B
C
D
E
F
G
a only
b only
Tpm2
H2afy
P4ha1
Fyn
U2af1a KD
U2af1b KD
Counts
% exon inclusion
con
U2af1
a-only
b-only
E5
E6a
E6b
E7
E8
E9a
E9b
E10
Relative exon inclusion
ex 9a
ex 9b
RNAi
Hnrnpc
Ptbp1
Srsf7
P1+S7
PTBP1#1
PTBP1#2
SRSF7#1
SRSF7#2
Flag-over-expression
U2af1a
U2af1b
U2AF1a
U2AF1b
PTBP1
SRSF7
8
9a
9b
10
P4ha1
U2AF1b-preferred sequence context

Figure 3. Stoichiometry of U2AF1 isoforms determines alternative 3'-splice site. (A-D) RNA-Seq read alignments of *Tpm2* (A), *H2afy* (B), *P4ha1* (C) and *Fyn* (D) gene loci in U2af1**a**-only, U2af1**b**-only and *U2af1* knockdown in corresponding cells. Please note that the designation of mutually exclusive exon a and b of these genes is in the order of exons from 5' to 3'-end direction for convenience. Inclusion of exon a or b is shown based on the quantitation of RNA-Seq data with the matching color code. **(E)** An RNAi screening to identify a factor(s) for *P4ha1* exon 9b alternative splicing. Relative inclusion of exon 9a or exon 9b in *P4ha1* expression was presented after the knockdown of indicated RNA-binding proteins (RBPs) in $TSC^{-/-}$ MEFs. The asterisks indicate the statistically significant decrease in the inclusion of exon 9b compared to the control. The data are presented as the mean (SD) (*p<0.0062, two-tailed Student's t test, n=3 for technical repeats). **(F)** The effect of PTBP1 and SRSF7 overexpression on the inclusion of *P4ha1* exon 9b. PTBP1 or SRSF7 was overexpressed in $TSC^{-/-}$ MEFs and the relative inclusion of exon 9b was measured. Two independent repeats of the experiments are shown. The asterisks indicate the statistically significant increase in the inclusion of exon 9b compared to control. The data are presented as the mean (SD) (*p<0.01, two-tailed Student's t test, n=3 for technical repeats). **(G)** A proposed model for the mutually exclusive alternative splicing of *P4ha1* upon the changes of U2AF1 isoform stoichiometry. The stoichiometry of U2AF1 isoforms in cells determines the usage of one of the tandem exons' splice site based on the nucleotide composition and the other splice site is selected by other splicing factors. In this case, PTBP1 and SRSF7 are one of the splicing factors involved in the mutually exclusive alternative splicing of *P4ha1*.

In our datasets, RNA-Seq read alignments of several duplicated tandem exons displayed mutually exclusive alternative splicing when comparing U2af1**a**- and U2af1**b**-only cells; moreover, RNAi knockdown of *U2af1* in those cells also showed characteristic U2AF1 isoform-dependent changes of mutually exclusive alternative splicing. For example, the inclusion of exon 6a in the mutually exclusive alternative splicing of *Tpm2* is dependent upon the overall level of U2AF1, but independent of which U2AF1 isoforms is present, as the shift of exon inclusion to 6b occurred similarly in the RNAi knockdown of *U2af1* in both U2af1**a**- and U2af1**b**-only cells (Fig. 3A). In contrast, U2AF1-isoform dependent tandem exon splicing became apparent in the *H2afy* expression. In this case, the inclusion of exon 6b decreased as *U2af1* was knocked down in both isoform-specific cell lines. Notably, however, the inclusion of exon 6b was more favored in U2af1**a**-only cells compared to U2af1**b**-only cells (Fig. 3B). These observations suggest that U2AF1a, as compared to U2AF1b, is more specific to exon 6b inclusion and is more capable of competing against exon 6a inclusion by unknown splicing factor(s). In the cases of *P4ha1* and *Fyn* expression, U2AF1b was more critical for the inclusion of a specific exon between the tandem exons. In *P4ha1* mutually exclusive alternative splicing, U2af1**b**-only cells exhibit more favorable exon 9a inclusion (59.9%) than U2af1**a**-only cells (31.0%) (Fig. 3C). Furthermore, the knockdown of *U2af1* shifted the exon inclusion to 9b in both cell lines (40.1% to 63.5% in U2af1**b**-only cells and 69.0% to 83.5% in U2af1**a**-only cells) (Fig. 3C). Similar observations were made in the expression of *Fyn*. In this case, exon 9b is more preferentially selected in U2af1**b**-only compared to U2af1**a**-only cells (25.7% vs. 13.9% exon 9b inclusion) and the knockdown of *U2af1* in both cell lines decreased exon

9b inclusion (8.2% in **b**-only vs. 9.5% in **a**-only). Thus, with a varying degree, it seems that U2AF1b is more specific than U2AF1a is to the splicing of *P4ha1* exon 9a and *Fyn* exon 9b; the other exon in those tandem exons is likely to be spliced by an unknown splicing factor(s) as both U2AF1 isoform knockdown increases the inclusion of the other exon (Fig. 3C and D). Collectively, these results show that the stoichiometry of U2AF1 isoforms determines the selection of one of tandem exons for mutually exclusive alternative splicing and the other exon selection is completed by an unknown splicing factor(s). Indeed, a series of knockdown experiments on selected RBPs identified SRSF7 and PTBP1 as two of the splicing factors for *P4ha1* tandem exon splicing because the knockdown of *Srsf7* or *Ptbp1* in $TSC^{-/-}$ MEFs decreased the inclusion of exon 9b (Fig. 3E and Supplementary Fig. 3A-C). Consistent with these observations, overexpression of *Ptbp1* or *Srsf7* in $TSC^{-/-}$ MEFs increased the inclusion of exon 9b (Fig. 3F and Supplementary Fig. 3D). Together, these results support a model in *P4ha1* tandem exon splicing where U2AF1b has a higher preference to exon 9a inclusion than U2AF1a and furthermore, that PTBP1/SRSF7 have a role in the inclusion of exon 9b (Fig. 3G). Therefore, the stoichiometry of U2AF1 isoforms and the level of competing splicing factors in a given cellular context likely determine the selection of tandem exons in mutually exclusive splicing. Of note, the concept that the stoichiometry of U2AF1 isoforms affects alternative splicing also applies to a simpler splicing type, e.g. cassette type (Supplementary Fig. 3E-H). These demonstrate the importance of considering the functional differences of U2AF1 isoforms in mechanistic studies of the regulation of alternative splicing.

Isoform-specific interactomes of U2AF1 feature common but refined cellular functions

Differences in nucleotide preference and splicing regulation by U2AF1 isoforms raise the question of whether they form different functional complexes in cells. To identify proteins interacting with U2AF1 isoforms, we performed CRISPR-induced homologous recombination to insert a C-terminal Flag-tag to *U2af1* gene in U2af1**a**-only and U2af1**b**-only cell lines (Fig. 4A and Supplementary Fig. 4A). Characterization of the resulting Flag-tagged U2AF1 isoform-specific cell lines by western blots indicated that the Flag-tag was added to one allele of the *U2af1* gene in both U2af1**a**-only and U2af1**b**-only cell lines (Fig. 4B left). Immunoprecipitation (IP) with an α-Flag antibody followed by western blots confirmed that endogenous Flag-tagged U2AF1 isoforms pull down U2AF2, providing evidence that Flag-U2AF1 isoforms form endogenous U2AF complexes (Fig. 4B right). Since protein-protein interactions in the spliceosome are highly dynamic and transient, to capture the interactomes of U2AF1 isoforms, we performed the proteomics part of ribo-proteomics approach using formaldehyde-mediated crosslinking and Flag-IP in the presence of RNase A followed by mass spectrometry analysis [43,44]. Mass spectrometry analysis of the immunoprecipitated samples showed high enrichment of U2AF heterodimer, suggesting the enrichment of U2AF1-interacting proteins in the co-IP (Fig. 4C). We used the intensity based label-free quantification (LFQ) algorithm to assess the relative abundance of interactors normalized to each U2AF1 isoform (Supplementary Fig. 4B). From this approach, we identified 127 U2AF1a interactors and 192 U2AF1b interactors significantly enriched over control Flag-IP (Fig. 4C, Supplementary Fig. 4C

and Supplementary Table 4). Of these identified interactors, 23 and 88 proteins were specific to U2AF1a and U2AF1b, respectively (Fig. 4C, Supplementary Fig. 4C and Supplementary Table 4). Gene Ontology (GO) term analyses of the U2AF1 interactomes show that overall U2AF1a and U2AF1b interactomes are similar to each other and are highly enriched for the GO terms 'mRNA processing' and 'splicing processes' (Fig. 4D and Supplementary Fig. 4D and E). However, U2AF1a displays a greater association with interactors belonging to 'mRNA processing' and 'splicing processes' while U2AF1b interactors are highly enriched in 'translation' (Supplementary Fig. 4D). Thus, our data suggest that the isoforms of U2AF1 form overlapping yet distinct protein complexes. Importantly, all enriched GO terms contain a subset of proteins exclusive to either U2AF1 isoform (Fig. 4D and Supplementary Fig. 4E). For instance, SF3A1 is specific to the U2AF1a interactome while MBNL2 is exclusive to the U2AF1b interactome (Fig. 4D). To validate these results, we first conducted co-IP and western blot analyses using antibodies specific to identified interactors (HNRNPC1/C2 and HNRNPA1) of both U2AF1 isoforms in the presence of RNase A. As shown in Fig 4E, both HNRNPC1/C2 and HNRNPA1 directly interact with U2AF complexes endogenously. To further confirm the isoform-specific interactomes of U2AF1, we co-expressed MBNL2 (identified to bind to U2AF1b but not U2AF1a) and U2AF1a-Flag or U2AF1b-Flag in HEK293 cells and performed Flag-IP in the presence or absence of RNase A, followed by western blot analyses. As shown in Fig 4F, MBNL2 prefers to bind to U2AF1b over U2AF1a; this bias is not RNase dependent. Together, these validate the results of U2AF1 interactome analyses, which strongly suggest

that U2AF1 isoforms have overlapping cellular functions yet provide refined or different regulatory roles by forming distinctive protein complexes.

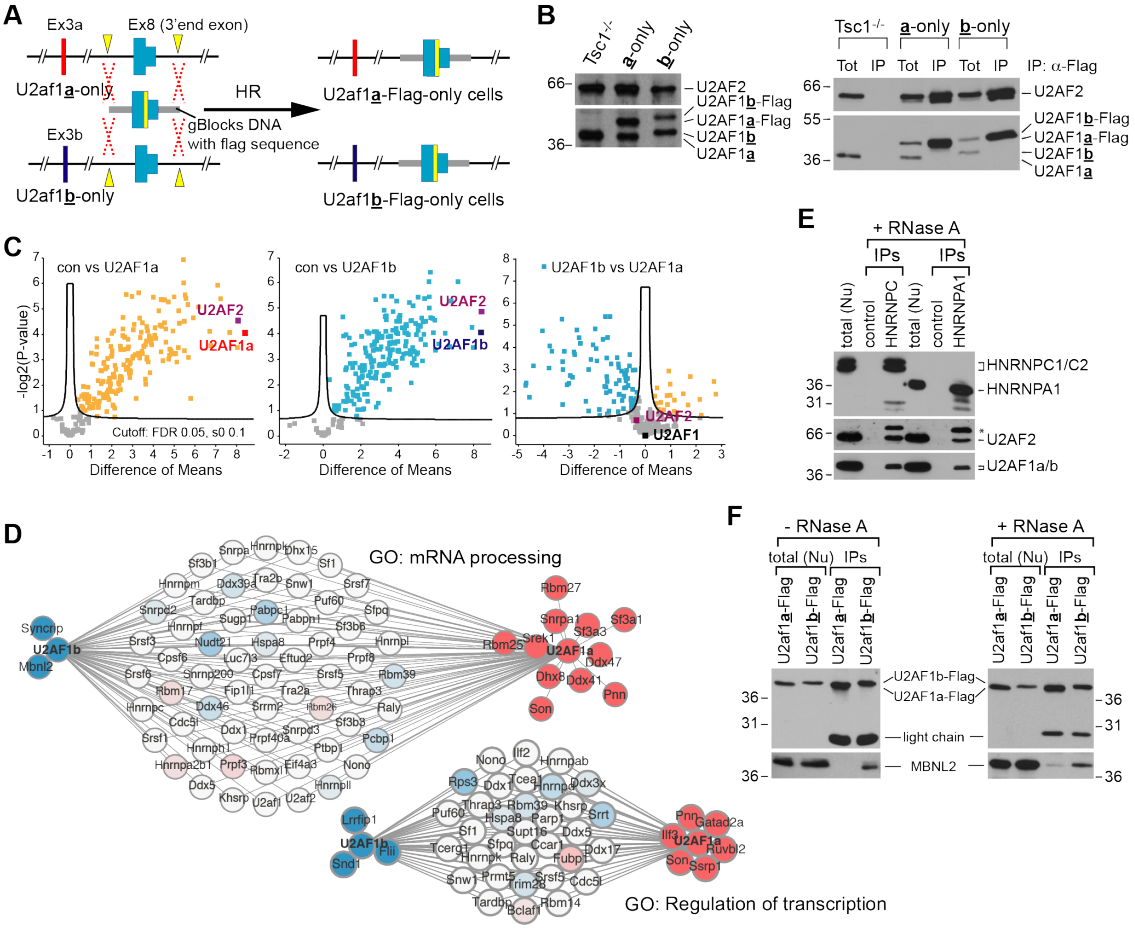


Figure 4. Overlapping but distinct interactome profiles of U2AF1 isoforms represents refined functional differences. (A) Schematic for CRISPR-induced homologous recombination (HR) to generate C-terminal Flag-tagged U2AF1 isoform-specific cell lines. Yellow rectangular box represents the Flag-tag. **(B)** Western blot analyses confirming the addition of a Flag-tag to U2AF1 (left). Anti-Flag immunoprecipitation (IP) and western blot analyses using total cellular extracts from Flag-tagged U2AF1 isoform-specific cell lines. Only Flag-tagged U2AF1 along with U2AF2 was immunoprecipitated (right). Tot:

Total cell lysate, 1% of input was loaded. **(C)** Volcano plots illustrating enrichment of both U2AF1 isoforms and corresponding interactors. The plot compares the log2 mean protein LFQ intensity difference between the control, U2AF1a and U2AF1b baits against the negative logarithmized p-values. **(D)** Interactome analyses of U2AF1a and U2AF1b. Interactomes of U2AF1a and U2AF1b in GO term mRNA processing (GO:0006397) and regulation of transcription (GO:0006355) are illustrated. Proteins colored in solid blue and red represent unique interactors of U2AF1b and U2AF1a, respectively. **(E, F)** Co-IP and western blotting validation of U2AF1 isoform interactome analysis. (E) Anti-HNRNPC1/C2 or HNRNPA1 antibodies were used for co-IPs in the presence of RNase A. Nuclear fraction of HEK293 cells was used for co-IPs. 2.5% of input was loaded as total. The asterisk denotes a non-specific band which may come from undissociated antibody chains. (F) MBNL2 and U2AF1a-Flag or U2AF1b-Flag were co-expressed in HEK293 cells. Flag-IP was performed with nuclear fractions in the absence or presence of RNase A. 10% of input was loaded as total. Anti-Flag and Anti-MBNL2 antibodies were used for immunoblotting.

U2AF1a-mediated 5'-UTR alternative splicing promotes translation

To understand the physiological consequences of U2AF1 isoform-mediated alternative splicing events (Fig. 2E), we surveyed regions of these alternative splicing events and found that 70% of the alternative splicing events affect the coding capacity of genes while 24% and 6% of the alternative splicing events occur in the 5'-UTR and 3'-UTR, respectively (Fig. 5A). This distribution is very similar to that of the known alternative splicing events in mouse genome (mm10, Supplementary Fig. 5 O). To look into the functional proteomes regulated by U2AF1 isoforms, we first searched the alternative splicing events affecting coding DNA sequence (CDS) regions against Pfam domain database. Out of the 224 CDS alternative splicing events, 64 events affected 74 functional domains annotated by Pfam. Among those, 28 Pfam domains were associated with the GO term while 46 Pfam domains were not (Fig. 5B). Almost a half of the GO term-associated Pfam domains including 'Pkinase_Tyr' and 'Homeobox' clustered together (green box in Fig. 5B). Collectively, these indicate that the alternative splicing events differentially regulated by U2AF1 isoforms could widely affect various cellular pathways.

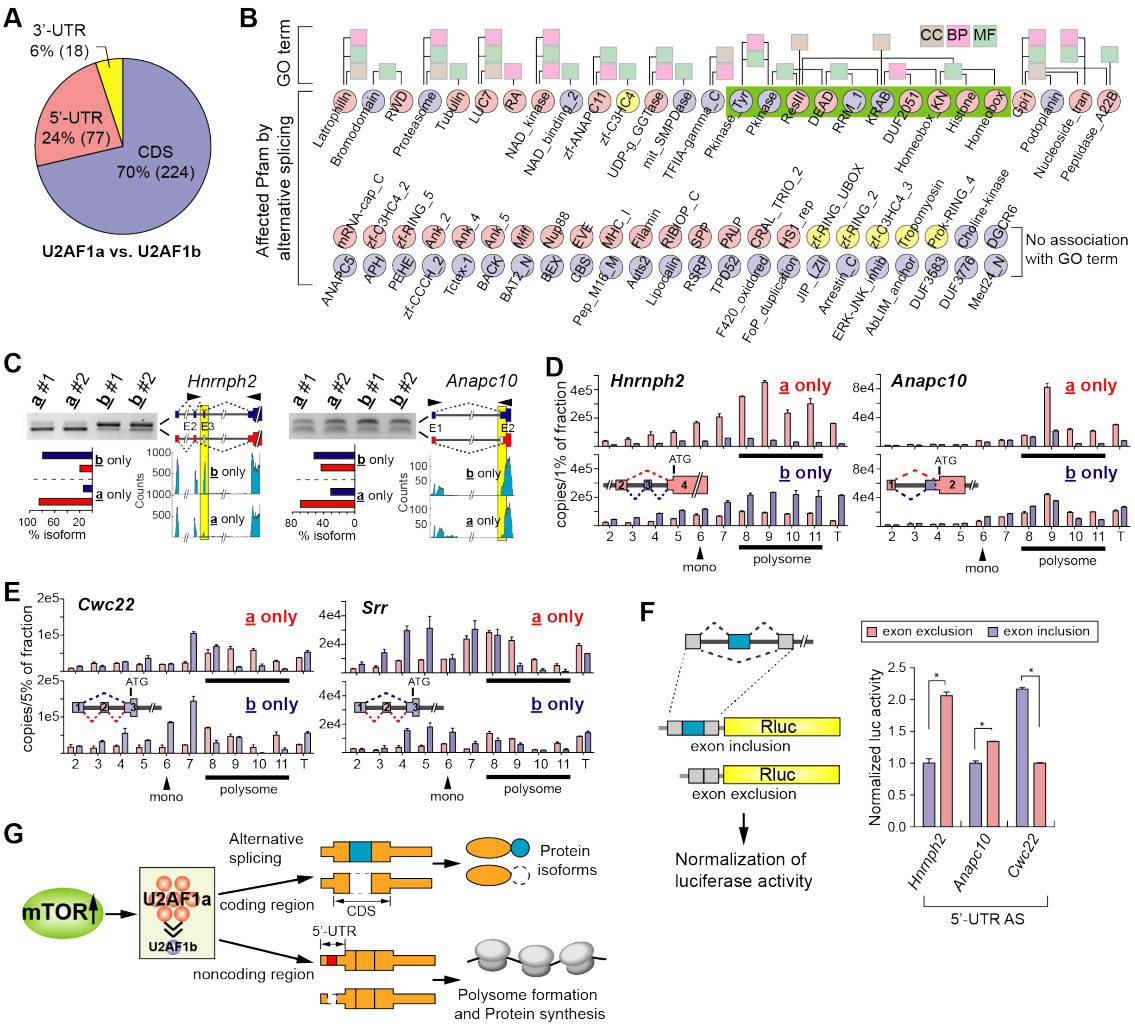


Figure 5. U2AF1 isoform-regulated alternative splicing in 5'-UTR modulates translation. (A) Distribution of alternative splicing in the regions of mRNA. Alternative splicing events displaying differences between U2af1**a**- and U2af1**b**-only cells were shown. **(B)** Affected Pfam domains by U2AF1 isoform-coordinated alternative splicing events were analyzed and their linkage to GO term is presented. Pfam domains affected by U2AF1a and U2AF1b-mediated alternative splicing are highlighted in light red and light blue, respectively. The Pfam domains highlighted in yellow are affected by both U2AF1a and U2AF1b-mediated alternative splicing. CC: Cellular Components; BP: Biological Processes; MF: Molecular Functions **(C)** Examples of 5'-UTR alternative splicing events

in U2af1**a**- and **b**-only cells. RT-PCR and agarose gel electrophoresis were conducted to validate alternative splicing events. RNA-Seq read alignments and quantitation of alternative splicing events are shown. Arrows indicate the position of primer binding sites for RT-PCR analyses. Splicing isoforms and their quantitation are color-coded as illustrated; yellow boxes highlight the alternative exons. Asterisk denotes a non-specific PCR product. **(D)** Polysome profiling analyses on the cytosolic fraction of U2af1**a**-only and U2af1**b**-only cells. Distribution of 5'-UTR alternative splicing transcripts (left, *Hnrnph2*; right, *Anapc10*) in polysome fractionation were analyzed by absolute quantitation using qPCR. T: 10% of input. Splicing isoforms are color-coded as illustrated. Monosome and polysome fractions are indicated. **(E)** The same analyses described in (D) were conducted on *Cwc22* and *Srr* genes. **(F)** Luciferase assays on the effects of 5'-UTR alternative splicing events on translation efficiency. The 5'-UTRs including or excluding the alternative exons of *Hnrnph2, Anapc10,* and *Cwc22* were placed into the 5'-UTR of luciferase reporter. The fold-changes of luciferase signals between the exon-included and exon-excluded 5'-UTR reporter construct pairs of the three genes are shown in bar graphs. The data are presented as the mean (SD) (*p<8.6e-5, two-tailed Student's t test, n=4 for technical repeats). **(G)** A proposed model for U2AF1 isoform-coordinated translational regulation by 5'-UTR alternative splicing and the connection to mTORC1 signaling. In this model, mTORC1-regulated changes of U2AF1 expression profile contributes to the proteome regulation by multiple ways. Alternative splicing in coding regions produce protein isoforms while alternative splicing in the 5'-UTR regulates differential translation.

Interestingly, the average length of 5'-UTR of transcripts with annotated alternative splicing events is 527.7 nucleotides, which is much longer than that of 5'-UTR of transcripts without reported alternative splicing events (226.6 nucleotides) in the mouse transcriptome (Supplementary Fig. 5A). In addition, the relative proportion of alternative exon length to the entire 5'-UTR is about 29.9% (Supplementary Fig. 5B). Since 5'-UTR is known to contain diverse elements for translational regulation [45,46], it is presumed that the alternative splicing events in the 5'-UTR reconfigure these regulatory *cis*-elements. Indeed, a search for potential regulatory elements in the affected 5'-UTRs in our data (U2af1**a**-only vs U2af1**b**-only) using UTRScan [47] identified several known 5'-UTR motifs and upstream open reading frames (uORFs) that were reconfigured by U2AF1 isoform-regulated alternative splicing events (Supplementary Fig. 5C and D). Out of the 77 genes showing U2AF1 isoform-mediated 5'-UTR alternative splicing events, 35 genes are predicted to reconfigure one or more uORFs (Supplementary Fig. 5D). In this case, not only the frequency but also the average length of uORFs were significantly changed by alternative splicing in the 5'-UTR.

RNA-Seq read alignments and semi-quantitative analyses of several genes on 5'-UTR alternative splicing validated U2AF1 isoform-specific events in all tested U2af1**a**- and U2af1**b**-only cells (Fig. 5C and Supplementary Fig. 5E). To examine whether these 5'-UTR alternative splicing events are associated with translational regulation, we conducted polysome fractionation using the cytoplasmic extracts from U2af1**a**- and U2af1**b**-only cells (Supplementary Fig. 5F) and analyzed the distribution of two alternative 5'-UTR isoforms by qPCR with absolute quantitation. *Hnrnph2* is mostly expressed as the exon 3-skipped

isoform in U2af1**a**-only cells whereas the exon 3-included isoform is highly expressed in U2af1**b**-only cells (Fig. 5C). Our quantitative analyses showed that, given the input amounts, the *Hnrnph2* exon 3-skipped isoform (shown in light red) formed polysomes more efficiently in both U2af1**a**- and U2af1**b**-only cells, while the transcript with exon inclusion (shown in light blue) was less efficient in forming polysomes (Fig. 5D left and Supplementary Fig. 5G, M). Similar differential polysomal distributions due to alternative splicing were observed in *Anapc10* (Fig. 5D right and Supplementary Fig. 5H, M). Interestingly, exon skipping is not always favored for polysome formations. In the case of *Cwc22* and *Srr* where exon inclusion in the 5'-UTR occurs more often in U2af1**a**-only cells (Supplementary Fig. 5I-J, M), the exon 2-included isoforms of both *Cwc22* and *Srr* (shown in light red) formed polysomes more efficiently compared to the exon 2-skipped isoforms (Fig. 5E and Supplementary Fig. 5K-M). Interestingly, in these select genes, alternative splicing events promoted by U2AF1a (skipping in *Hnrnph2* and *Anapc10*, inclusion in *cwc22* and *Srr*) leads to increase in polysome formation of the transcripts. To validate the findings of polysome fractionation analyses, we cloned the 5'-UTRs of *Hnrnph2, Anapc10, Cwc22*, and *Pex2*, including or excluding the alternative exon, into the 5'-UTR of luciferase reporter constructs (Fig. 5F). We then transfected these constructs into $TSC1^{-/-}$ MEFs and compared their luciferase activities to measure the effects of these 5'-UTRs on translation efficiency. As shown in Fig. 5F and Supplementary Fig. 5 N, consistent with the results of polysome fractionation analyses, the exclusion of the 5'-UTR alternative exons of *Hnrnph2*, *Anapc10*, and *Pex2* and the inclusion of the 5'-UTR alternative exon of *Cwc22* lead to higher translation of luciferase compared to their counterparts. Together,

these data show that the alternative splicing events in the 5'-UTR modulated by U2AF1 isoforms coordinate translation and suggest that the stoichiometry of U2AF1 isoforms plays a key role in the regulation of translation, potentially uncoupling the correlation between mRNA and protein abundance in cells.

Discussion

U2AF1 has been extensively studied for its crucial role in pre-mRNA splicing and the pathogenesis of myelodysplasia syndrome (MDS) [1,6,9,14,26–29,31–33,35,36,48–50]. Albeit two isoforms are expressed from *U2AF1*, early studies on U2AF1 were not able to functionally differentiate two isoforms [8,9,49]. Furthermore, commonly used mammalian cell lines often express more U2AF1a than U2AF1b [7–9]. Accordingly, most, if not all studies on U2AF1 and its pathogenic mutations do not distinguish U2AF1 isoforms [6,9,14,26–32,34–36,48,51–53]. A high sequence similarity along with the same molecular weight between the two isoforms make it challenging to study one isoform over the other. In addition, RNAi knockdown approaches for functional studies on U2AF1 isoforms are poised to generate U2AF1a-biased outcomes as many cell systems used for this kind of approach underrepresent U2AF1b expression [7–9]. In contrast, our genome editing approach to produce both isoform-specific cell lines provides an unbiased biological system to understand the function of U2AF1 isoforms. In fact, unlike previous reports [7,9], our study could bring up underrepresented U2AF1b functions as our data show that the number of exons and genes exclusively regulated by each U2AF1 isoform is similar (Fig. 2D and E). In conjunction with our interactome analyses, these indicate that two U2AF1 isoforms distinctively contribute to the transcriptome and have nuanced functional differences in cells. These conclusions are along the same lines with other well characterized tandem exon-derived isoforms including *PKM* and *FGFR2* [10–13,39]. In this regard, our findings of distinctive functions of U2AF1 isoforms raise important questions regarding pathogenic U2AF1 mutations in MDS. Since most, if not all, studies on U2AF1 pathogenic mutations do not

differentiate the two U2AF1 isoforms nor profile their expressions, they are not comprehensive in understanding the pathogenic mechanisms of MDS by not recognizing the potential functional impacts of the dynamic expressions of U2AF1 isoforms on the phenotypes that are attributed to U2AF1 mutations [31–36,50].

Obtaining the endogenously Flag-tagged U2AF1 isoform-specific cell lines with CRISPR-induced homologous recombination allowed us to enrich for U2AF1 isoform-specific interactomes by simply performing Flag-IP. More importantly, it eliminated the need to exogenously over-express bait proteins that may skew the stoichiometry of interactomes, and allowed us to IP endogenous U2AF1 isoform-specific complexes. The results of this comprehensive interactome analysis of U2AF1 isoforms suggest that the specificity of the isoform-specific interactomes could be a key characteristic of distinct functionality of the isoforms. Thus, these isoform-specific interactomes could not only help explain the different sequence preferences of the two isoforms, they could also represent previously unknown or sophisticated functions of U2AF1. Indeed, consistent with our findings of a possible link between U2AF1b and translation, a recent study suggested a role of U2AF1 and its mutations in translation regulation in the cytoplasm, although this study still lacked the information on U2AF1 isoforms [54]. Collectively, these demonstrate that our approach of interactome analyses is beneficial in revealing unknown and sophisticated functions of U2AF1 isoforms.

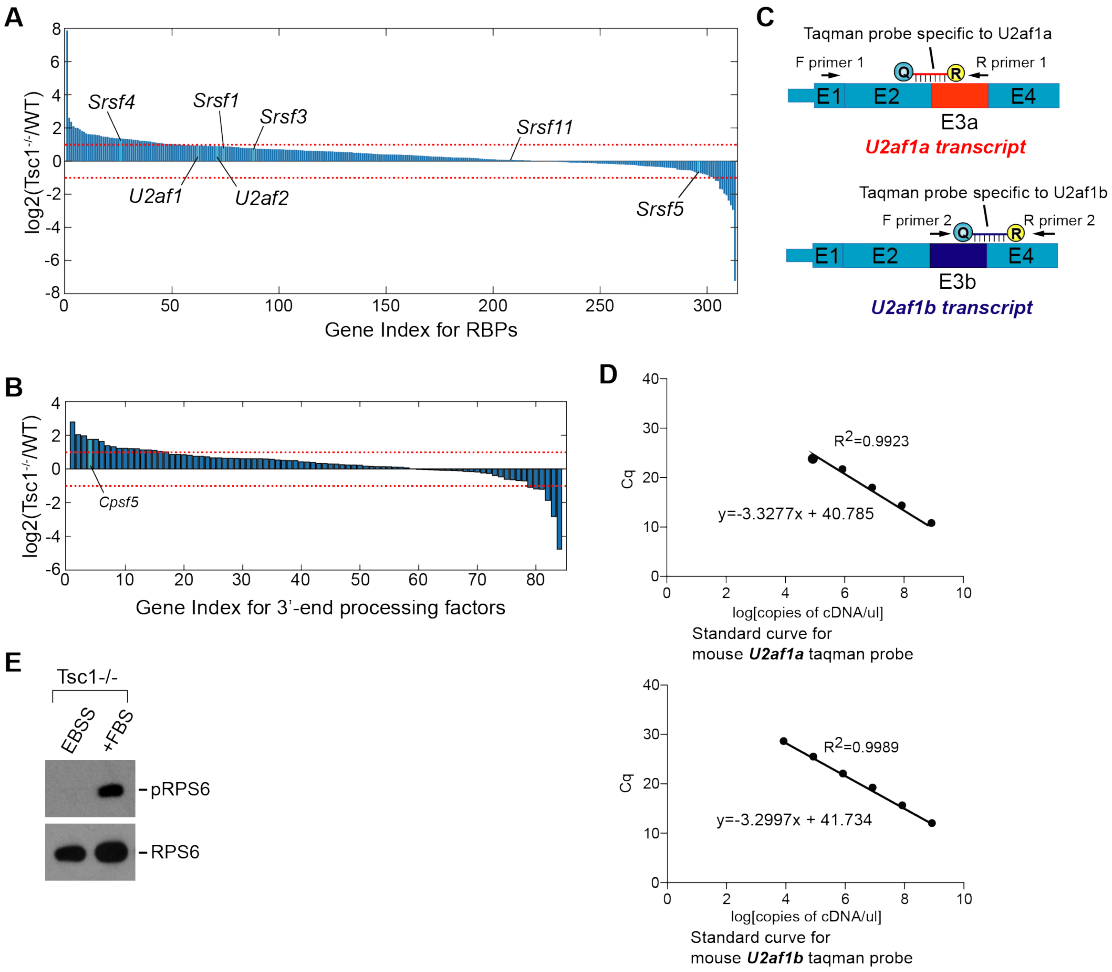
One of the surprising outcomes of U2AF1 isoform-specific alternative splicing is the translational regulation through 5'-UTR alternative splicing. As shown by various computational analyses, alternative splicing in the 5'-UTR driven by U2AF1 isoform

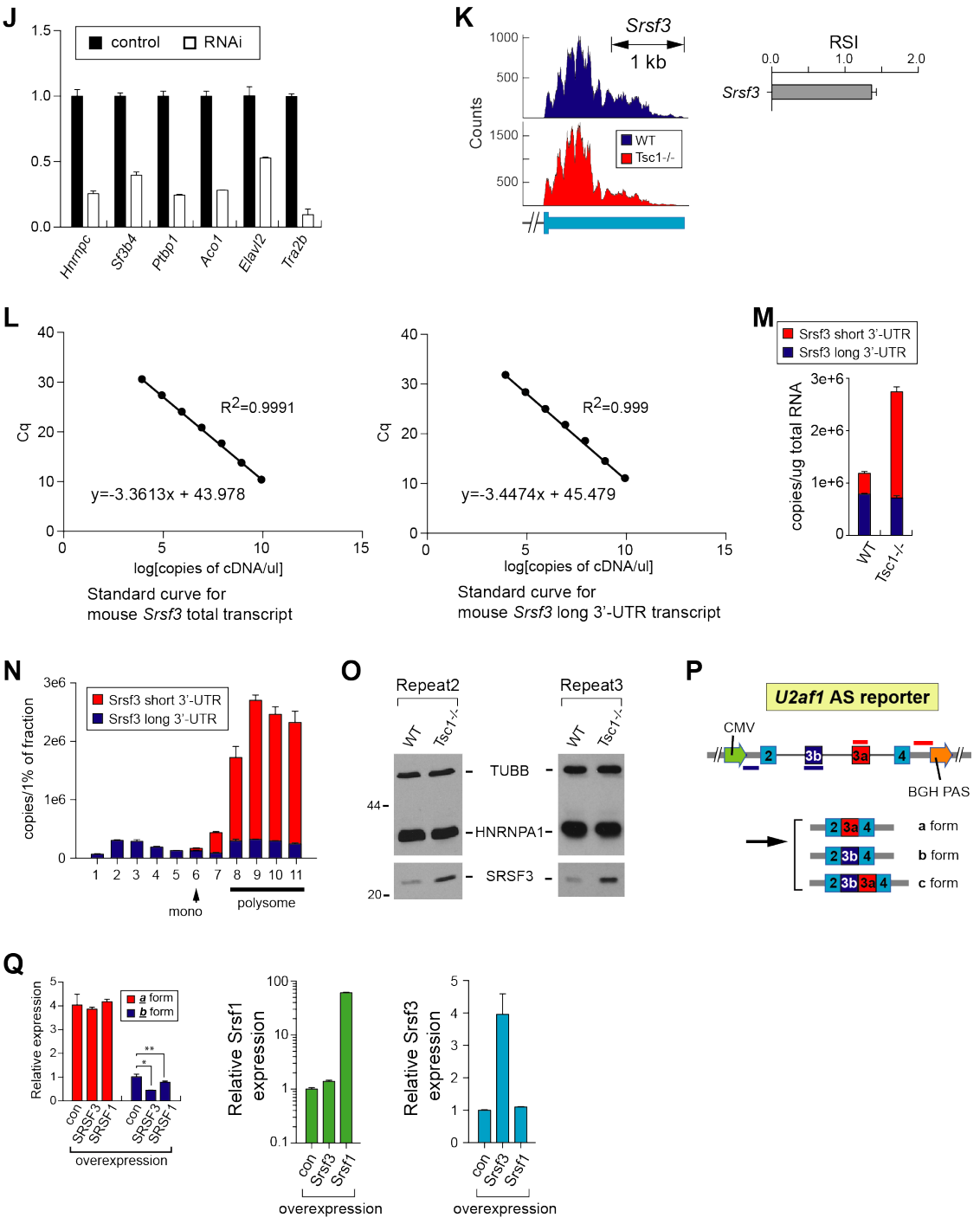
preferences as well as U2AF1 isoform knockdowns dynamically rearranges known *cis*-regulatory elements and uORFs (Supplementary Fig. 5C-D, Q-S). Thus, 5'-UTR alternative splicing reprograms multiple features in the 5'-UTR and can regulate translation. One of the most well characterized molecular signatures of 5'-UTR in translational regulation is the translational activation of 5'-TOP (terminal oligopyrimidine) containing mRNAs by mTOR [55]. Intriguingly, most 5'-UTR alternative splicing events specific to U2af1**a**- and U2af1**b**-only cells do not contain a 5'-TOP signature (only 2 out of 77 events in Fig. 5A contain 5'-TOP feature, Supplementary Fig. 5T). This pattern was consistent with the dataset of U2af1**a**-only control vs knockdown and U2af1**b**-only control vs knockdown (7 out of 130 events and 5 out of 132 events contain 5'-TOP feature, Supplementary Fig. 5P, U). A recent study using transcription start site profiling reported that distinct classes of non-5'-TOP mRNAs were subjected to mTOR-regulated translational control. Interestingly, the study found that mTOR-dependent translation of these non-5'-TOP mRNAs have short or long 5'-UTRs and the length of 5'-UTR is associated with cellular pathways targeted by non-5'-TOP mRNAs [56]. Our findings in this study present U2AF1 isoform-regulated alternative splicing in the 5'-UTR as a previously unrecognized translational regulatory mechanism and provide mTORC1-regulated U2AF1 isoform profile as a molecular link between mTORC1 and non-5'-TOP mRNA translation (Fig. 5G).

Acknowledgements

This work was supported by National Science Foundation [NSF-III1755761] to WZ and National Institutes of Health [1R01GM113952-01A1] and Department of Defense – Congressionally Directed Medical Research Programs [W81XWH-16-1-0135] to JY.

Supplementary Materials

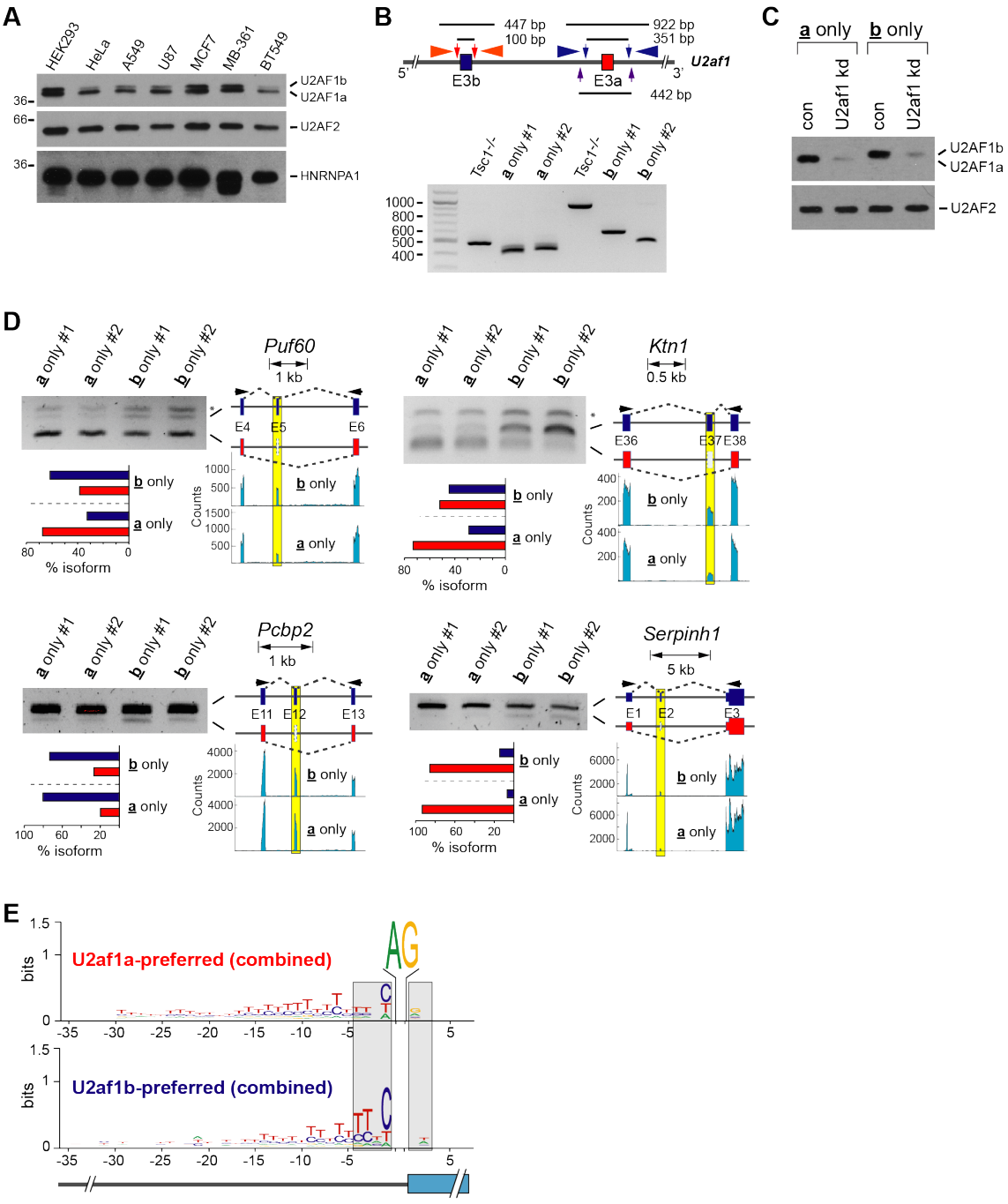


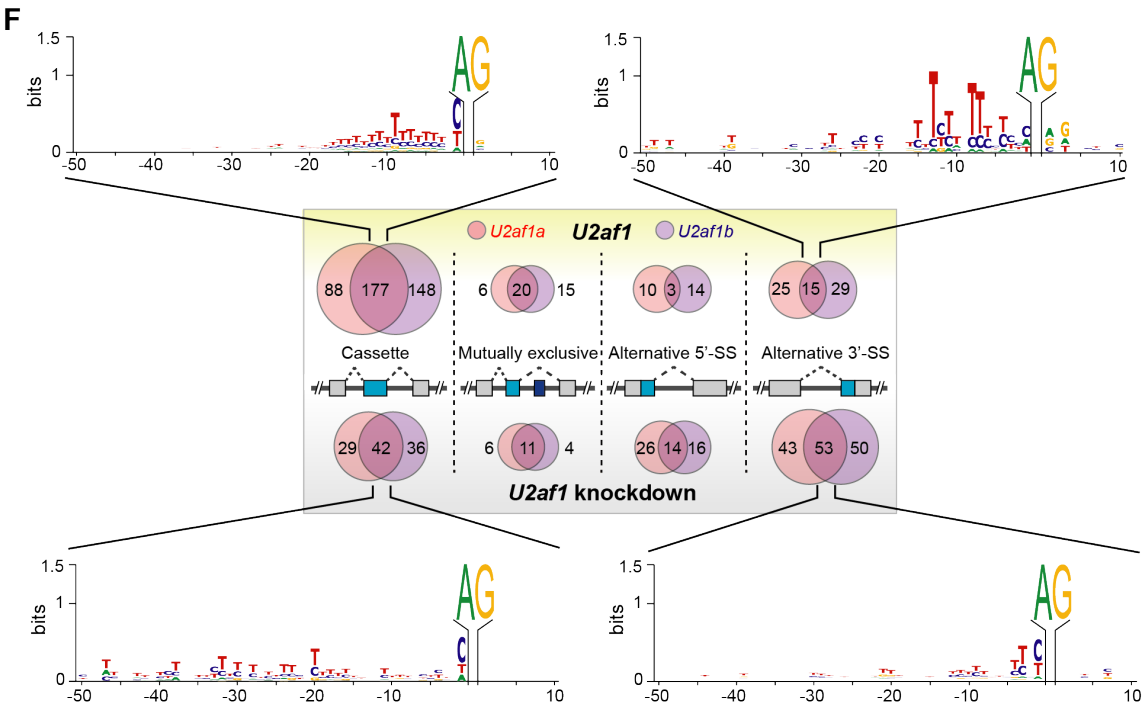


Supplementary Figure 1. (**A, B**) Histogram plots of transcript expression level change in TSC1$^{-/-}$ compared to WT MEFs of select RBPs (A) and 3'-end processing factors (B)

based on RNA-Seq quantification. Genes mentioned in the main text are indicated. (**C**) A diagram showing the locations of Taqman probes and PCR primers used in Taqman qPCR assays on *U2af1* transcript isoforms. (**D**) Standard curves for absolute quantitation of mouse *U2af1* transcript isoforms using Taqman qPCR assay. (**E**) Western blot analysis on $TSC1^{-/-}$ MEF cells under EBSS or FBS add-back treatment. The changes of mTORC1 activity by EBSS and FBS treatment are evidenced by pS6 (pRPS6) blotting, while S6 blotting serves as a loading control. (**F**) Knockdown of various SR splicing factors by RNAi. Relative expression of SR proteins in control and knockdown samples was measured using qPCR. (**G**) Knockdown of various polyadenylation factors by RNAi. Relative expression of CPSF proteins in control and knockdown samples was measured using qPCR. (**H**) Changes of *U2af1* isoform expression upon the knockdown of various polyadenylation factors. The amounts of *U2af1a* and *U2af1b* isoforms were measure by Taqman qPCR assay. (**I, J**) Small scale RNAi screen for regulators of alternative splicing. Each indicated splicing factor was knocked down by RNAi and the expression of *U2af1* isoforms was measured by Taqman qPCR assay. (**K**) *Srsf3* transcript undergoes 3'-UTR alternative polyadenylation (APA) upon the mTOR activation. (left) RNA-seq read alignments of the *Srsf3* 3'-UTR in WT and $TSC1^{-/-}$ MEF cell lines. (right) RSI (relative shortening index) measurement of *Srsf3* 3'-UTR APA in $TSC1^{-/-}$ relative to WT MEF cell lines. (**L**) Standard curves for absolute quantitation using qPCR assay on mouse *Srsf3 long 3'-UTR and total* transcripts. (**M**) Absolute quantitation of *Srsf3* long and short 3'-UTR transcripts using qPCR. The standard curves in (L) were used for the calculation. (**N**) Polysome profiling for *Srsf3* long and short 3'-UTR transcripts in $TSC1^{-/-}$ MEFs.

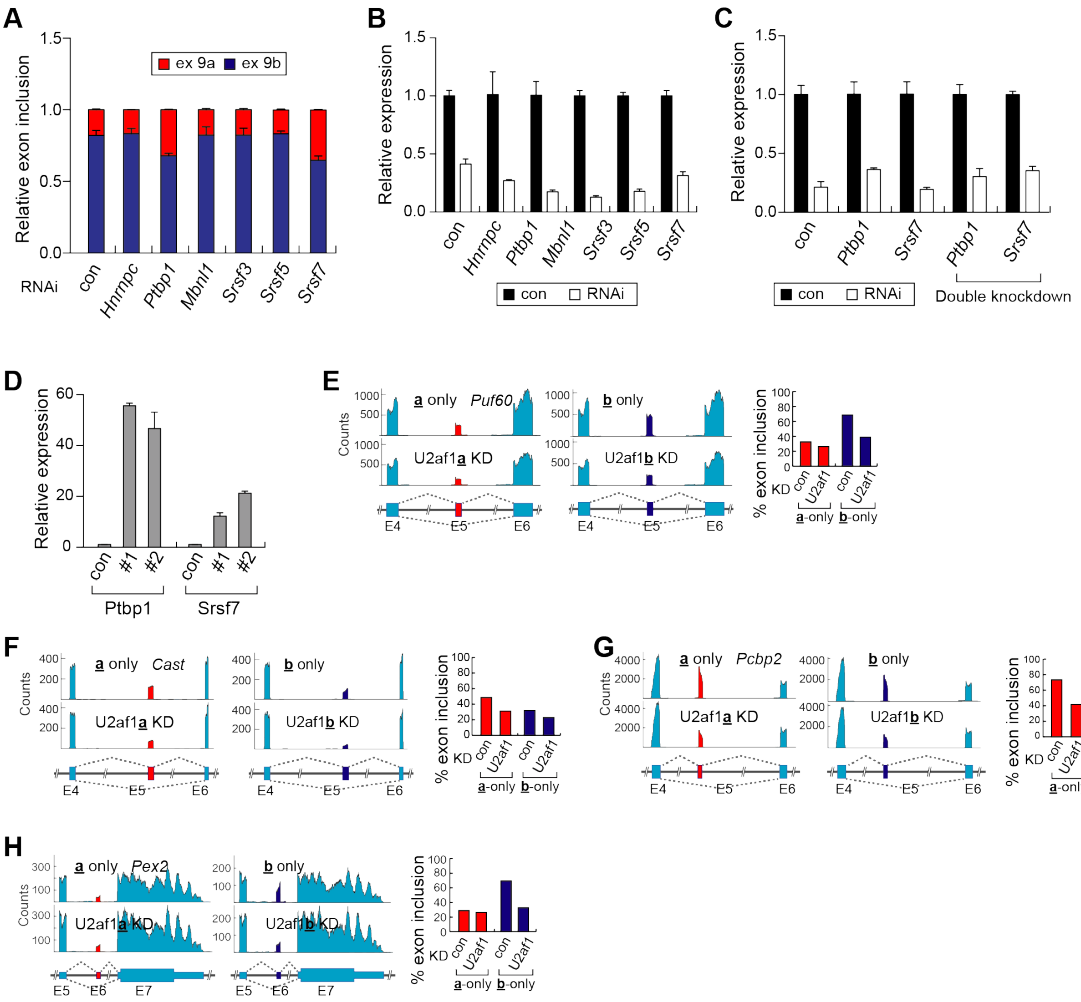
Absolute quantitation of long 3'-UTR and total *Srsf3* transcripts was conducted. The amount of short 3'-UTR *Srsf3* transcript was calculated by the subtraction of long from total amounts. **(O)** The biological repeats of the Western blot analysis of SRSF3 in WT and $TSC1^{-/-}$ MEFs shown in Fig. 1F. TUBULIN and HNRNPA1 were used as loading controls. (**P**) Schematic of *U2af1* tandem exon splicing reporter construct. (**Q**) Overexpression of SRSF3 suppresses exon 3b inclusion in *U2af1* splicing reporter assay. Two splicing isoforms from the reporter was measured by Taqman qPCR with relative quantitation upon the overexpression of SRSF3, SRSF1 and vector alone. Overexpression of *Srsf1* and *Srsf3* in the experiments was validated by qPCR with relative quantitation. The data are the mean (SD) (**p*=0.010, ***p*=0.13 ; two-tailed Student's *t* test, n=3 for technical repeats).


A
HEK293
HeLa
A549
U87
MCF7
MB-361
BT549
U2AF1b
U2AF1a
U2AF2
HNRNPA1
36
66
36
B
447 bp
100 bp
922 bp
351 bp
U2af1
5'
3'
E3b
E3a
442 bp
Tsc1-/-
a only #1
a only #2
Tsc1-/-
b only #1
b only #2
1000
800
600
500
400
C
a only
b only
con
U2af1 kd
con
U2af1 kd
U2AF1b
U2AF1a
U2AF2
D
a only #1
a only #2
b only #1
b only #2
Puf60
1 kb
E4
E5
E6
b only
a only
% isoform
Counts
Ktn1
0.5 kb
E36
E37
E38
Pcbp2
E11
E12
E13
Serpinh1
5 kb
E1
E2
E3
E
U2af1a-preferred (combined)
U2af1b-preferred (combined)
bits
AG



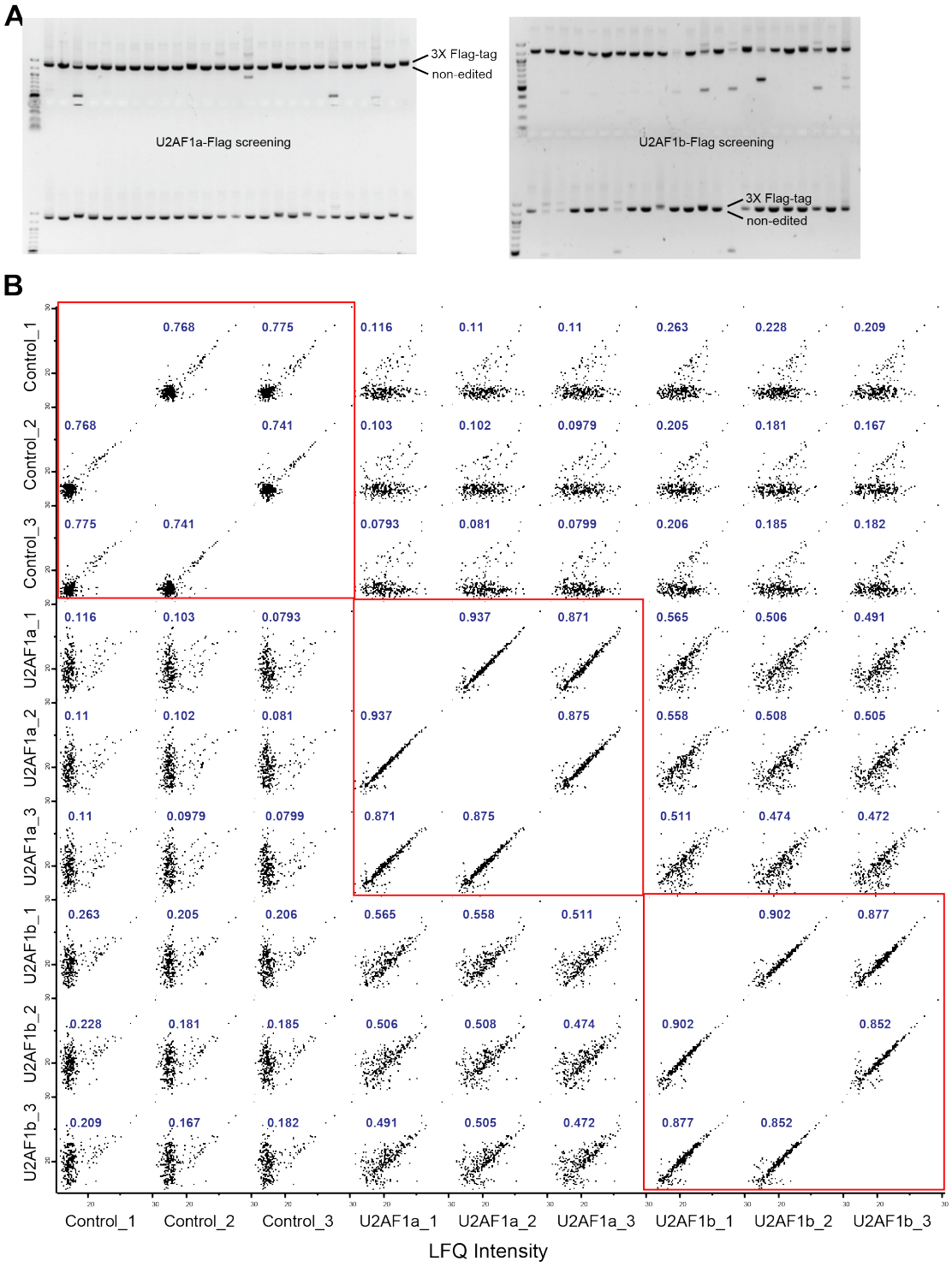
Supplementary Figure 2. (**A**) Western blot analyses of U2AF1 isoforms using various cancer cell lines. (**B**) (upper panel) A diagram illustrating the design of CRISPR/Cas9-mediated generation of U2af1**a**- and U2af1**b**-only cell lines. The locations of exon-flanking gRNAs are denoted by arrows. Primer sets used for gDNA screening are denoted by triangles. The distances between primer sets and the flanking gRNA pairs are indicated. (lower panel) PCR using the denoted gDNA screening primer sets followed by agarose gel electrophoresis. The downshift of the bands in U2af1**a**- and U2af1**b**-only cells demonstrates the successful homozygous editing by the exon-flanking gRNA pairs via NHEJ. (**C**) Western blot analysis on TSC1$^{-/-}$ MEF, U2af1**a**-only#1, and U2af1**b**-only#1 cells treated with control (con) or U2af1-targeting siRNA (U2af1 kd). (**D**) Splicing of four representative alternative splicing events showing U2AF1 isoform preferences. For each panel, RNA-Seq read alignments and quantitation of alternative

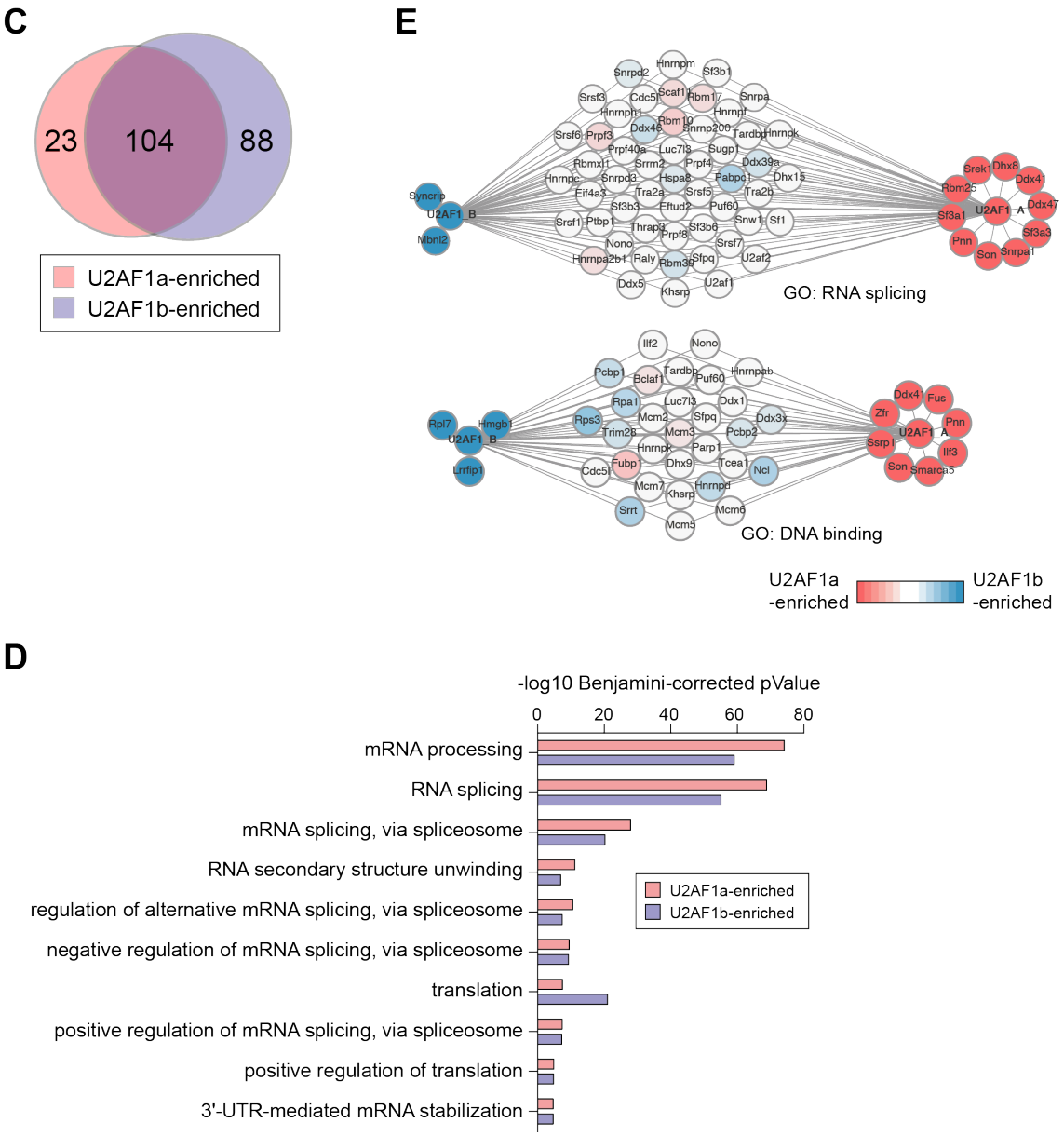
exon of U2af1**a**-only#1 and U2af1**b**-only#1 cell lines are shown. RT-PCR validation in U2af1**a**-only#1, U2af1**a**-only#2, U2af1**b**-only#1, and U2af1**b**-only#2 cell lines are shown. Splicing isoforms are color-coded as illustrated; yellow boxes highlight the alternative exons. Asterisks denote non-specific bands. (**E**) The frequency of upstream nucleotides of 3'-splice site preferred by U2AF1 isoforms. The certainty of nucleotides in each position was calculated by considering both cassette and alternative 3'-splice site type splicing. (**F**) The frequency of upstream nucleotides of the 3'-splice site of exons commonly dependent on both U2AF1 isoforms. (upper) Exons that are more included in the presence of *U2af1*. (lower) Exons that are more included in the absence of *U2af1* (upon knockdown). (left) 3'-splice site of cassette type alternative splicing. (right) 3'-splice site of alternative 3'-splice site type alternative splicing.



Supplementary Figure 3. (**A**) An initial RNAi screen for RBPs that promote the inclusion of exon 9b in *P4ha1*. Relative exon inclusion was calculated by relative quantitative PCR. (**B**) The measurement of knockdown efficiency of RBPs shown in (A). Relative qPCR was used for quantitation. (**C**) The measurement of knockdown efficiency of RBPs shown in Fig. 3E. Relative qPCR was used for quantitation. (**D**) The measurement of *Ptbp1* and *Srsf7* overexpression shown in Figure 3F. Relative qPCR was used for quantitation. **(E-H)** Examples of cassette type alternative splicing events that depend upon U2AF1 isoform in various manners. RNA-Seq read alignments of *Puf60*

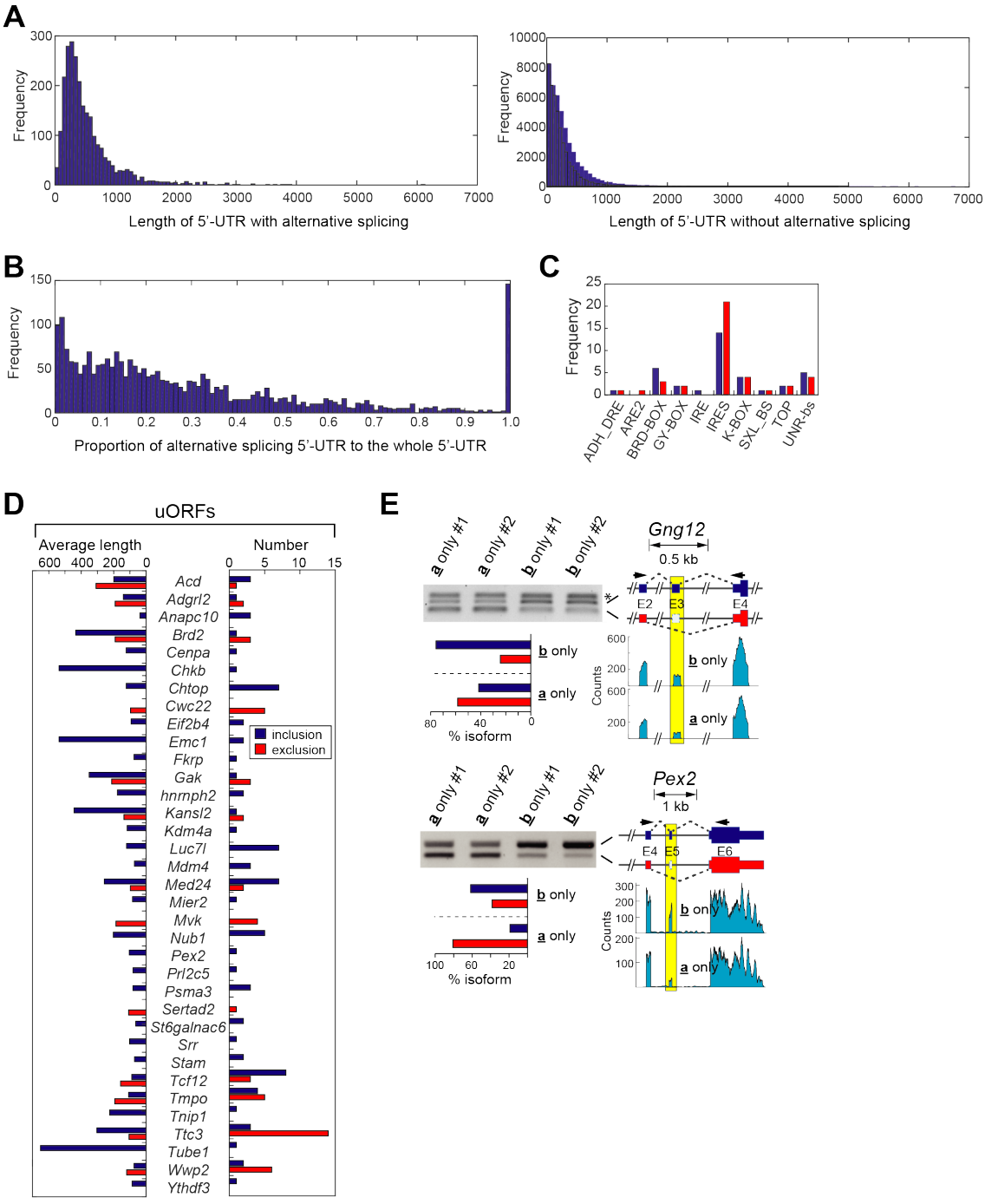
(preferring U2AF1b) (E), *Cast* (preferring U2AF1a) (F), *Pcbp2* (no U2AF1 isoform preference) (G) and *Pex2* (only dependent upon U2AF1b) (H) gene loci in U2af1**a**-only, U2af1**b**-only and *U2af1* knockdown in corresponding cells are shown. Inclusion of the alternative exon is plotted based on the quantitation of RNA-Seq data with the matching color code.

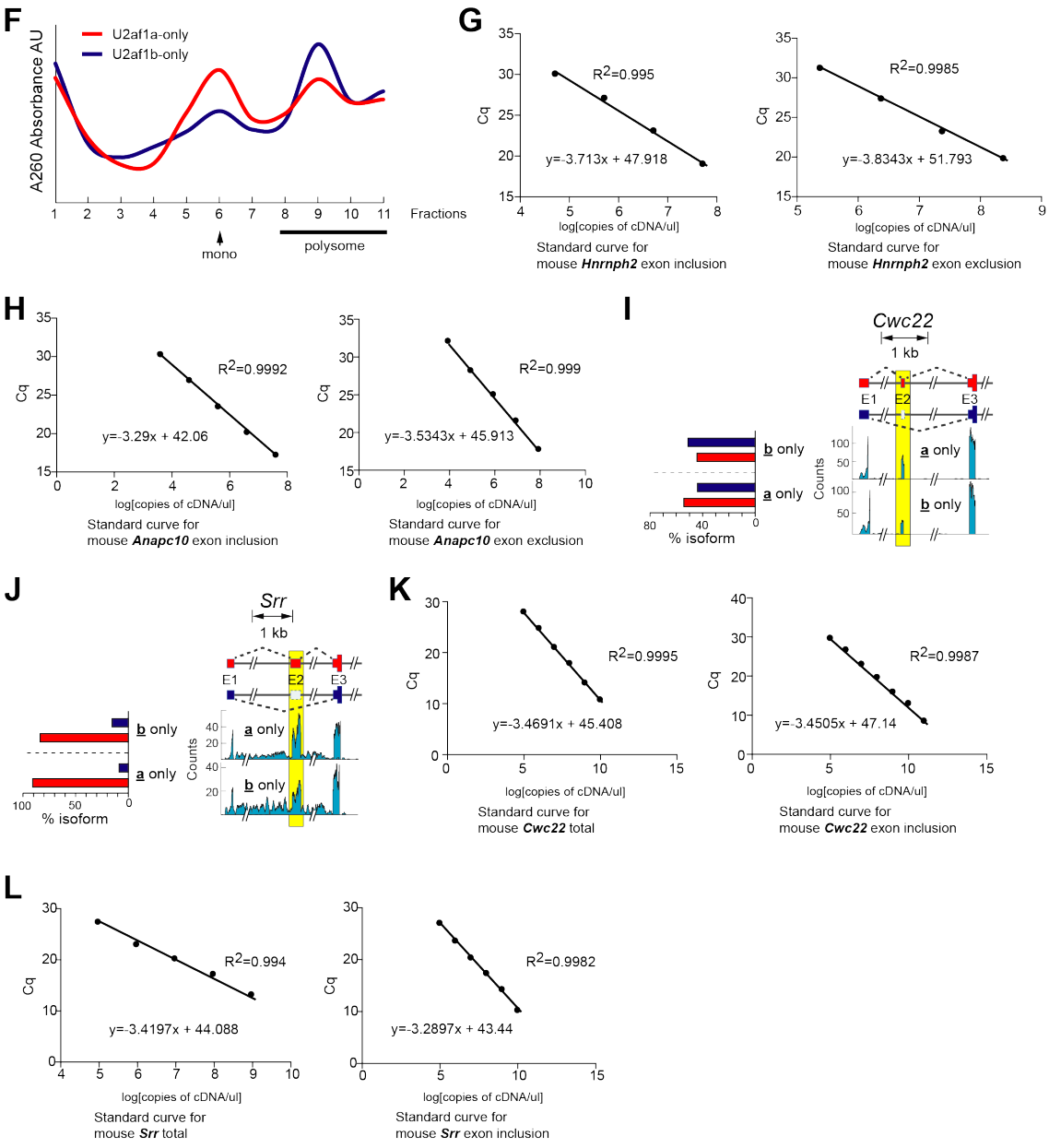

A
3X Flag-tag
non-edited
U2AF1a-Flag screening
U2AF1b-Flag screening
3X Flag-tag
non-edited
B
Control_1
Control_2
Control_3
U2AF1a_1
U2AF1a_2
U2AF1a_3
U2AF1b_1
U2AF1b_2
U2AF1b_3
0.768 0.775 0.116 0.11 0.11 0.263 0.228 0.209
0.768 0.741 0.103 0.102 0.0979 0.205 0.181 0.167
0.775 0.741 0.0793 0.081 0.0799 0.206 0.185 0.182
0.116 0.103 0.0793 0.937 0.871 0.565 0.506 0.491
0.11 0.102 0.081 0.937 0.875 0.558 0.508 0.505
0.11 0.0979 0.0799 0.871 0.875 0.511 0.474 0.472
0.263 0.205 0.206 0.565 0.558 0.511 0.902 0.877
0.228 0.181 0.185 0.506 0.508 0.474 0.902 0.852
0.209 0.167 0.182 0.491 0.505 0.472 0.877 0.852
Control_1
Control_2
Control_3
U2AF1a_1
U2AF1a_2
U2AF1a_3
U2AF1b_1
U2AF1b_2
U2AF1b_3
LFQ Intensity

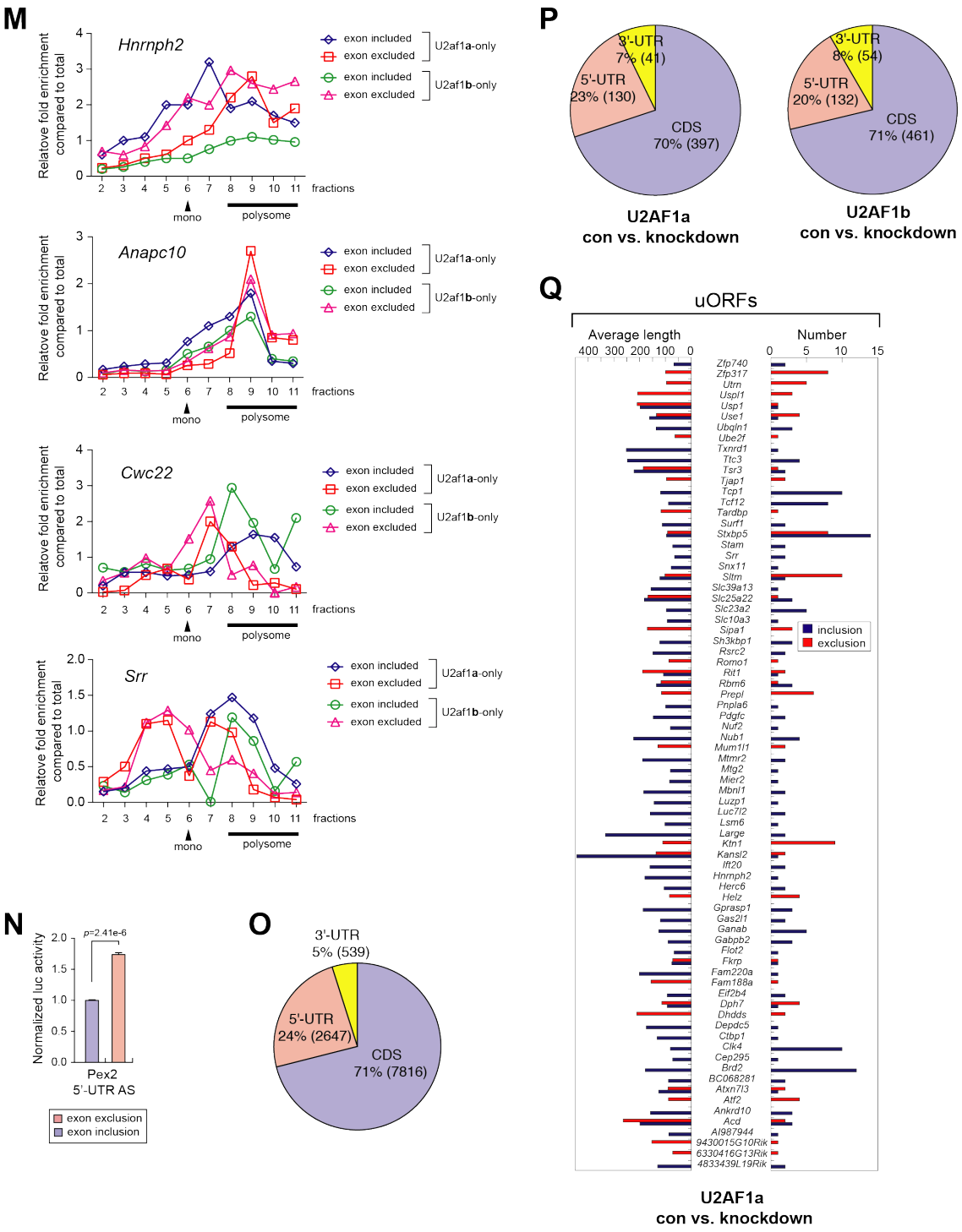


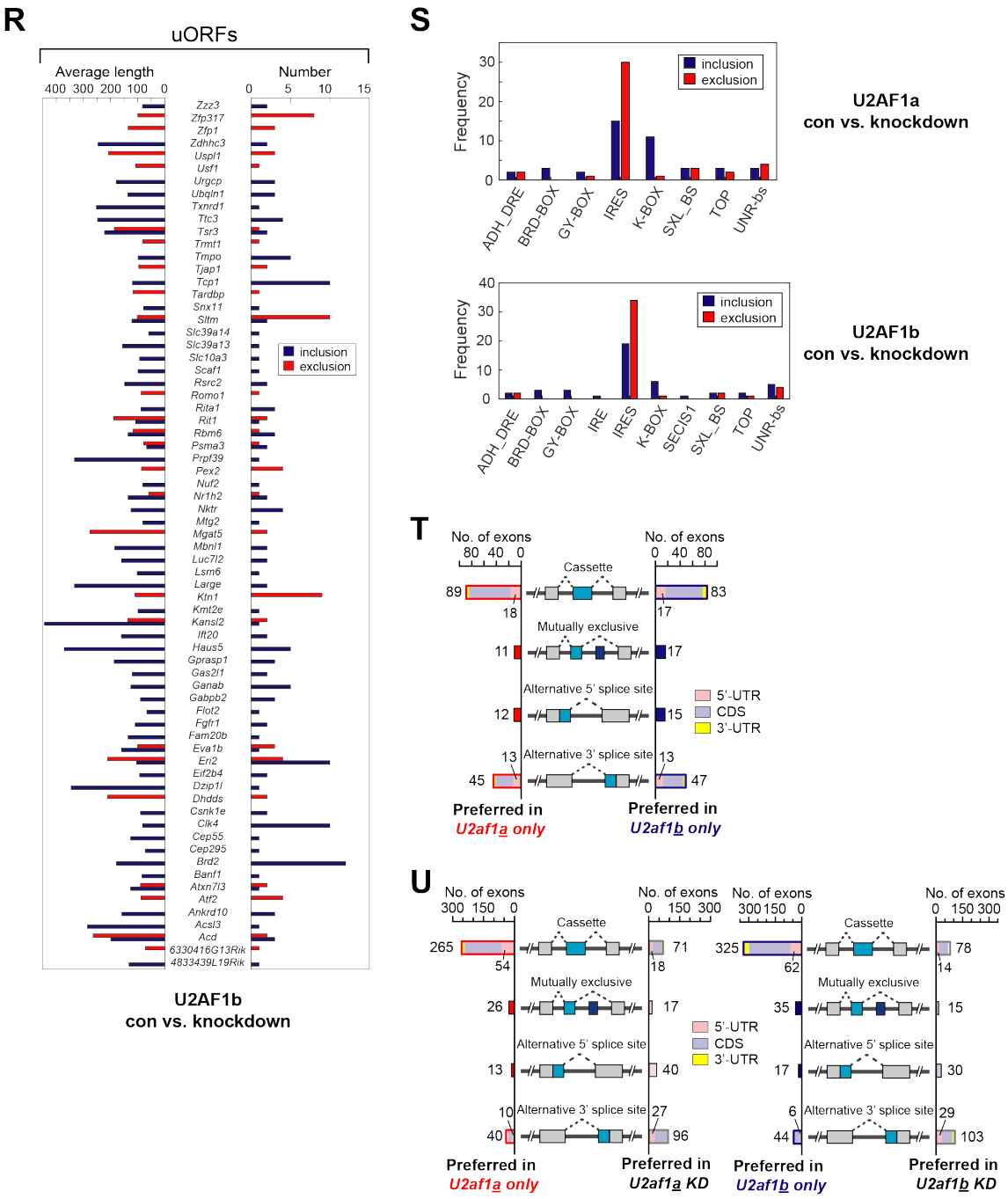
Supplementary Figure 4. (**A**) Screening of U2AF1a-Flag and U2AF1b-Flag cell lines after CRISPR/Cas9-initiated HR using synthesized gBlocks DNA. A slight shift caused by 3X-Flag tag insertion into the C-terminus of *U2af1* locus in U2af1**a**- and U2af1**b**-only cell lines was marked. (**B**) Multi-scatter plot illustrating the protein LFQ intensity and R^2 correlation among three sample repeats. Within technical replicates of each U2AF1 pull-down, the average R-squared pair-wise correlation coefficients of the protein LFQ

intensity profiles was 0.885, indicating high similarity. In contrast, the pair-wise comparison of the U2AF1 isoform-specific pulldowns had a lower average R-squared correlation coefficients (0.510) than the background protein correlation profile represented by the control Flag-IP samples using untagged $TSC1^{-/-}$ MEF cell line (0.761), indicating significant U2AF1 isoform-dependent difference in protein interactor abundance. (**C**) A summary of U2af1 isoform-specific interactomes. Overlapped and unique proteins identified by interactome studies are shown in a Venn Diagram. (**D**) GO-term analyses of U2AF1a- and U2AF1b-specific interactomes. (**E**) Interactomes of U2AF1a and U2AF1b in GO term RNA splicing and DNA binding are illustrated. Proteins colored in solid blue and red represent unique interactors of U2AF1b and U2AF1a, respectively.


A
Frequency
Length of 5'-UTR with alternative splicing
Length of 5'-UTR without alternative splicing
B
Proportion of alternative splicing 5'-UTR to the whole 5'-UTR
C
ADH_DRE
ARE2
BRD-BOX
GY-BOX
IRE
IRES
K-BOX
SXL_BS
TOP
UNR-bs
D
uORFs
Average length
Number
inclusion
exclusion
Acd
Adgrl2
Anapc10
Brd2
Cenpa
Chkb
Chtop
Cwc22
Eif2b4
Emc1
Fkrp
Gak
hnrnph2
Kansl2
Kdm4a
Luc7l
Mdm4
Med24
Mier2
Mvk
Nub1
Pex2
Prl2c5
Psma3
Sertad2
St6galnac6
Srr
Stam
Tcf12
Tmpo
Tnip1
Ttc3
Tube1
Wwp2
Ythdf3
E
a only #1
a only #2
b only #1
b only #2
Gng12
0.5 kb
E2
E3
E4
b only
a only
% isoform
Counts
Pex2
1 kb
E4
E5
E6


F
U2af1a-only
U2af1b-only
A260 Absorbance AU
Fractions
mono
polysome
G
R2=0.995
y=-3.713x + 47.918
Cq
log[copies of cDNA/ul]
Standard curve for mouse Hnrnph2 exon inclusion
R2=0.9985
y=-3.8343x + 51.793
Standard curve for mouse Hnrnph2 exon exclusion
H
R2=0.9992
y=-3.29x + 42.06
Standard curve for mouse Anapc10 exon inclusion
R2=0.999
y=-3.5343x + 45.913
Standard curve for mouse Anapc10 exon exclusion
I
Cwc22
1 kb
E1
E2
E3
b only
a only
% isoform
Counts
J
Srr
1 kb
E1
E2
E3
b only
a only
% isoform
Counts
K
R2=0.9995
y=-3.4691x + 45.408
Standard curve for mouse Cwc22 total
R2=0.9987
y=-3.4505x + 47.14
Standard curve for mouse Cwc22 exon inclusion
L
R2=0.994
y=-3.4197x + 44.088
Standard curve for mouse Srr total
R2=0.9982
y=-3.2897x + 43.44
Standard curve for mouse Srr exon inclusion


M
Hnrnph2
Anapc10
Cwc22
Srr
Relatove fold enrichment compared to total
fractions
mono
polysome
exon included
exon excluded
U2af1a-only
U2af1b-only
N
Normalized luc activity
p=2.41e-6
Pex2
5'-UTR AS
exon exclusion
exon inclusion
O
3'-UTR
5% (539)
5'-UTR
24% (2647)
CDS
71% (7816)
P
3'-UTR
7% (41)
5'-UTR
23% (130)
CDS
70% (397)
U2AF1a
con vs. knockdown
3'-UTR
8% (54)
5'-UTR
20% (132)
CDS
71% (461)
U2AF1b
con vs. knockdown
Q
uORFs
Average length
Number
inclusion
exclusion
U2AF1a
con vs. knockdown



Supplementary Figure 5. (**A**) Length distribution of 5'-UTR with or without alternative splicing annotated in mm10 UCSC mouse genome annotation. (**B**) Proportion of alternative exon in 5'-UTRs with annotated alternative splicing events. The frequency of each portion was displayed in the y-axis. (**C**) Rearrangement of known 5'-UTR

regulatory elements by U2AF1a- and U2AF1b-mediated alternative splicing. Red boxes and blue boxes indicate the elements and frequencies changed by U2AF1a and U2AF1b, respectively. (**D**) Average length and number of uORFs changed by U2AF1a- and U2AF1b-mediated alternative splicing. Red boxes and blue boxes indicate the dynamics of uORFs by exon exclusion and inclusion, respectively. (**E**) Examples of 5'-UTR alternative splicing events in U2af1**a**- and **b**-only cells. RT-PCR and agarose gel electrophoresis were conducted to validate alternative splicing events. RNA-Seq read alignments and quantitation of alternative splicing events are shown. Arrows indicate the position of primer binding sites for RT-PCR analyses. Splicing isoforms and their quantitation are color-coded as illustrated; yellow boxes highlight the alternative exons. Asterisk denotes a non-specific PCR product. (**F**) Absorbance measurement of polysome fractionations. Fractions corresponding to monosme and polysomes are indicated. (**G, H**) Standard curves of qPCR assay for *Hnrnph2* (g) and *Anapc10* (h) 5'UTR alternative transcript isoforms. (**I, J**) Examples of 5'-UTR alternative splicing events in U2af1**a**- and **b**-only cells. RNA-Seq read alignments and quantitation of alternative splicing events are shown. Splicing isoforms and their quantitation are color-coded as illustrated; yellow boxes highlight the alternative exons. Asterisk denotes a non-specific PCR product. (**K, L**) Standard curves of qPCR assay for *Cwc22* (k) and *Srr* (l) 5'UTR alternative transcript isoforms. Due to technical difficulties, total- and long-specific qPCR primer sets were designed and used for absolute quantitation. (**M**) Data from polysome profiling analyses of Figure 5 (D-E) normalized to total input. Splicing isoforms are color-coded as denoted. **(N)** Luciferase assays on the effects of 5'-UTR alternative splicing events on translation

efficiency. The 5'-UTRs including or excluding the alternative exon of *Pex2* were placed into the 5'-UTR of luciferase reporter. The fold-change of luciferase signals between the exon-included and exon-excluded 5'-UTR reporter construct pair are shown in bar graphs. **(O, P)** Overall distribution of alternative splicing in the regions of mRNA was shown for (O) the mouse genome (mm10) and (P) U2AF1a control vs knockdown and U2AF1b control vs knockdown. **(Q, R)** Average length and number of uORFs changed by alternative splicing events in (Q) U2AF1a control vs knockdown and (R) U2AF1b control vs knockdown datasets. Red boxes and blue boxes indicate the dynamics of uORFs by exon exclusion and inclusion, respectively. **(S)** Rearrangement of known 5'-UTR regulatory elements by alternative splicing events in U2AF1a control vs knockdown and U2AF1b control vs knockdown datasets. Red boxes and blue boxes indicate the elements and frequencies changed by exon exclusion and inclusion, respectively. **(T)** Distribution of alternative splicing regions of mRNA in U2af1**a**- and U2af1**b**-only cell datasets. In each alternative splicing category, the distribution of regions is represented by color code. The number of alternative splicing events in cassette and alternative 3'-splice site categories was displayed. **(U)** Distribution of alternative splicing regions of mRNA in U2AF1a control vs knockdown and U2AF1b control vs knockdown datasets. In each alternative splicing category, the distribution of regions is represented by color code. The number of alternative splicing events in cassette and alternative 3'-splice site categories was displayed.

Supplementary Table 1. Alternative splicing events upon *U2af1* knockdown in U2Af1**a**-only cell.

Supplementary Table 2. Alternative splicing events upon *U2af1* knockdown in U2Af1**b**-only cell.

Supplementary Table 3. Differential alternative splicing events between U2Af1**a**-only and U2AF1**b**-only cells.

Supplementary Table 4. List of identified U2AF1 isoform-specific interactors.

To access Supplementary Tables 1–4, go to https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkz761#supplementary-data**.**

References

1. Shi,Y. (2017) Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.*, **18**, 655–670.

2. Zamore,P.D., Patton,J.G. and Green,M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355**, 609–614.

3. Zhang,M., Zamore,P.D., Carmo-Fonseca,M., Lamond,A.I. and Green,M.R. (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 8769–73.

4. Singh,R., Valcárcel,J. and Green,M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, **268**, 1173–6.

5. Valcárcel,J., Gaur,R.K., Singh,R. and Green,M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science*, **273**, 1706–9.

6. Guth,S., Tange,T.Ø., Kellenberger,E. and Valcárcel,J. (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol. Cell. Biol.*, **21**, 7673–81.

7. Pacheco,T.R., Moita,L.F., Gomes,A.Q., Hacohen,N. and Carmo-Fonseca,M. (2006) RNA Interference Knockdown of hU2AF ^{35} Impairs Cell Cycle Progression and Modulates Alternative Splicing of Cdc25 Transcripts. *Mol. Biol. Cell*, **17**, 4187–4199.

8. Pacheco,T.R., Gomes,A.Q., Barbosa-Morais,N.L., Benes,V., Ansorge,W., Wollerton,M., Smith,C.W., Valcárcel,J. and Carmo-Fonseca,M. (2004) Diversity of

vertebrate splicing factor U2AF35: identification of alternatively spliced U2AF1 mRNAS. *J. Biol. Chem.*, **279**, 27039–49.

9. Kralovicova,J., Knut,M., Cross,N.C.P. and Vorechovsky,I. (2015) Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3′ splice-site organization and activity of U2AF-related proteins. *Nucleic Acids Res.*, **43**, 3747–3763.
10. Israelsen,W.J. and Vander Heiden,M.G. (2015) Pyruvate kinase: Function, regulation and role in cancer. *Semin. Cell Dev. Biol.*, **43**, 43–51.
11. Dayton,T.L., Jacks,T. and Vander Heiden,M.G. (2016) PKM2, cancer metabolism, and the road ahead. *EMBO Rep.*, **17**, 1721–1730.
12. Shirakihara,T., Horiguchi,K., Miyazawa,K., Ehata,S., Shibata,T., Morita,I., Miyazono,K. and Saitoh,M. (2011) TGF-β regulates isoform switching of FGF receptors and epithelial-mesenchymal transition. *EMBO J.*, **30**, 783–795.
13. Warzecha,C.C., Sato,T.K., Nabet,B., Hogenesch,J.B. and Carstens,R.P. (2009) ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. *Mol. Cell*, **33**, 591–601.
14. Kielkopf,C.L., Rodionova,N.A., Green,M.R. and Burley,S.K. (2001) A Novel Peptide Recognition Mode Revealed by the X-Ray Structure of a Core U2AF35/U2AF65 Heterodimer. *Cell*, **106**, 595–605.
15. Kim,J. and Guan,K.-L. (2019) mTOR as a central hub of nutrient signalling and cell growth. *Nat. Cell Biol.*, **21**, 63–71.
16. Kwiatkowski,D.J., Zhang,H., Bandura,J.L., Heiberger,K.M., Glogauer,M., el-

Hashemite,N. and Onda,H. (2002) A mouse model of TSC1 reveals sex-dependent lethality from liver hemangiomas, and up-regulation of p70S6 kinase activity in Tsc1 null cells. *Hum. Mol. Genet.*, **11**, 525–34.

17. Tee,A.R., Fingar,D.C., Manning,B.D., Kwiatkowski,D.J., Cantley,L.C. and Blenis,J. (2002) Tuberous sclerosis complex-1 and -2 gene products function together to inhibit mammalian target of rapamycin (mTOR)-mediated downstream signaling. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 13571–6.
18. Chang,J.-W., Zhang,W., Yeh,H.-S., Park,M., Yao,C., Shi,Y., Kuang,R. and Yong,J. (2018) An integrative model for alternative polyadenylation, IntMAP, delineates mTOR-modulated endoplasmic reticulum stress response. *Nucleic Acids Res.*, **46**, 5996–6008.
19. Chang,J.-W., Zhang,W., Yeh,H.-S., de Jong,E.P., Jun,S., Kim,K.-H., Bae,S.S., Beckman,K., Hwang,T.H., Kim,K.-S., *et al.* (2015) mRNA 3′-UTR shortening is a molecular signature of mTORC1 activation. *Nat. Commun.*, **6**, 7218.
20. Shen,S., Park,J.W., Lu,Z., Lin,L., Henry,M.D., Wu,Y.N., Zhou,Q. and Xing,Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, E5593-601.
21. Suzuki,K., Bose,P., Leong-Quong,R.Y.Y., Fujita,D.J. and Riabowol,K. (2010) REAP: A two minute cell fractionation method. *BMC Res. Notes*, **3**, 294.
22. Van Etten,J., Schagat,T.L. and Goldstrohm,A.C. (2013) A guide to design and optimization of reporter assays for 3′ untranslated region mediated regulation of mammalian messenger RNAs. *Methods*, **63**, 110–118.

23. Kralovicova,J. and Vorechovsky,I. (2017) Alternative splicing of U2AF1 reveals a shared repression mechanism for duplicated exons. *Nucleic Acids Res.*, **45**, 417–434.

24. Carter,M.S., Doskow,J., Morris,P., Li,S., Nhim,R.P., Sandstedt,S. and Wilkinson,M.F. (1995) A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *J. Biol. Chem.*, **270**, 28995–9003.

25. Durand,S. and Lykke-Andersen,J. (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat. Struct. Mol. Biol.*, **20**, 702–709.

26. Park,S.M., Ou,J., Chamberlain,L., Simone,T.M., Yang,H., Virbasius,C.-M., Ali,A.M., Zhu,L.J., Mukherjee,S., Raza,A., *et al.* (2016) U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. *Mol. Cell*, **62**, 479–490.

27. Yoshida,K., Sanada,M., Shiraishi,Y., Nowak,D., Nagata,Y., Yamamoto,R., Sato,Y., Sato-Otsubo,A., Kon,A., Nagasaki,M., *et al.* (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, **478**, 64–69.

28. Yip,B.H., Steeples,V., Repapi,E., Armstrong,R.N., Llorian,M., Roy,S., Shaw,J., Dolatshad,H., Taylor,S., Verma,A., *et al.* (2017) The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J. Clin. Invest.*, **127**, 2206–2221.

29. Wu,S., Romfo,C.M., Nilsen,T.W. and Green,M.R. (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. *Nature*, **402**, 832–835.

30. Yoshida,H., Park,S.-Y., Oda,T., Akiyoshi,T., Sato,M., Shirouzu,M., Tsuda,K., Kuwasako,K., Unzai,S., Muto,Y., *et al.* (2015) A novel 3' splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes Dev.*, **29**, 1649–60.

31. Shao,C., Yang,B., Wu,T., Huang,J., Tang,P., Zhou,Y., Zhou,J., Qiu,J., Jiang,L., Li,H., *et al.* (2014) Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.*, **21**, 997–1005.

32. Voith von Voithenberg,L., Sánchez-Rico,C., Kang,H.-S., Madl,T., Zanier,K., Barth,A., Warner,L.R., Sattler,M. and Lamb,D.C. (2016) Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc. Natl. Acad. Sci. U. S. A.*, **113**, E7169–E7175.

33. Chusainow,J., Ajuh,P.M., Trinkle-Mulcahy,L., Sleeman,J.E., Ellenberg,J. and Lamond,A.I. (2005) FRET analyses of the U2AF complex localize the U2AF35/U2AF65 interaction in vivo and reveal a novel self-interaction of U2AF35. *RNA*, **11**, 1201–1214.

34. Shirai,C.L., White,B.S., Tripathi,M., Tapia,R., Ley,J.N., Ndonwi,M., Kim,S., Shao,J., Carver,A., Saez,B., *et al.* (2017) Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nat. Commun.*, **8**, 14060.

35. Shirai,C.L., Ley,J.N., White,B.S., Kim,S., Tibbitts,J., Shao,J., Ndonwi,M., Wadugu,B., Duncavage,E.J., Okeyo-Owuor,T., *et al.* (2015) Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*, **27**, 631–643.

36. Jenkins,J.L. and Kielkopf,C.L. (2017) Splicing Factor Mutations in Myelodysplasias:

Insights from Spliceosome Structures. *Trends Genet.*, **33**, 336–348.

37. Merendino,L., Guth,S., Bilbao,D., Martínez,C. and Valcárcel,J. (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3′ splice site AG. *Nature*, **402**, 838–841.

38. Graveley,B.R., Hertel,K.J. and Maniatis,T. (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA*, **7**, 806–18.

39. Simsek,D., Tiu,G.C., Flynn,R.A., Byeon,G.W., Leppek,K., Xu,A.F., Chang,H.Y. and Barna,M. (2017) The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*, **169**, 1051-1065.e18.

40. Denton,R.M., Pullen,T.J., Armstrong,C.T., Heesom,K.J., Rutter,G.A., Koike,M., Koike,K., McCormack,J.G., Denton,R.M., McCormack,J.G., *et al.* (2016) Calcium-insensitive splice variants of mammalian E1 subunit of 2-oxoglutarate dehydrogenase complex with tissue-specific patterns of expression. *Biochem. J.*, **473**, 1165–78.

41. Pohl,M., Bortfeldt,R.H., Grützmann,K. and Schuster,S. (2013) Alternative splicing of mutually exclusive exons-A review. *BioSystems*, **114**, 31–38.

42. Zheng,X., Boyer,L., Jin,M., Mertens,J., Kim,Y., Ma,L., Ma,L., Hamm,M., Gage,F.H. and Hunter,T. (2016) Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. *Elife*, **5**, 1–25.

43. Yeh,H.-S., Chang,J.-W. and Yong,J. (2016) Ribo-Proteomics Approach to Profile RNA–Protein and Protein–Protein Interaction Networks. In. Humana Press, New York, NY, pp. 165–174.

44. Yong,J., Kasim,M., Bachorik,J.L., Wan,L. and Dreyfuss,G. (2010) Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol. Cell*, **38**, 551–62.

45. Leppek,K., Das,R. and Barna,M. (2017) Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.*, **19**, 158–174.

46. Hinnebusch,A.G., Ivanov,I.P. and Sonenberg,N. Translational control by 5′-untranslated regions of eukaryotic mRNAs.

47. Grillo,G., Turi,A., Licciulli,F., Mignone,F., Liuni,S., Banfi,S., Gennarino,V.A., Horner,D.S., Pavesi,G., Picardi,E., *et al.* (2010) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.*, **38**, D75–D80.

48. Guth,S., Martínez,C., Gaur,R.K. and Valcárcel,J. (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol. Cell. Biol.*, **19**, 8263–71.

49. Pacheco,T.R., Coelho,M.B., Desterro,J.M.P., Mollet,I. and Carmo-Fonseca,M. (2006) In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. *Mol. Cell. Biol.*, **26**, 8183–90.

50. Fei,D.L., Zhen,T., Durham,B., Ferrarone,J., Zhang,T., Garrett,L., Yoshimi,A., Abdel-Wahab,O., Bradley,R.K., Liu,P., *et al.* (2018) Impaired hematopoiesis and leukemia development in mice with a conditional knock-in allele of a mutant splicing factor gene *U2af1*. *Proc. Natl. Acad. Sci.*, **115**, E10437–E10446.

51. Tavanez,J.P., Madl,T., Kooshapur,H., Sattler,M. and Valcárcel,J. (2012) hnRNP A1 proofreads 3' splice site recognition by U2AF. *Mol. Cell*, **45**, 314–29.

52. Nguyen,H.D., Yadav,T., Giri,S., Saez,B., Graubert,T.A. and Zou,L. (2017) Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol. Cell*, **65**, 832-847.e4.

53. Nguyen,H.D., Leong,W.Y., Li,W., Reddy,P.N.G., Sullivan,J.D., Walter,M.J., Zou,L. and Graubert,T.A. (2018) Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. *Cancer Res.*, **78**, 5363–5374.

54. Palangat,M., Anastasakis,D.G., Fei,D.L., Lindblad,K.E., Bradley,R., Hourigan,C.S., Hafner,M. and Larson,D.R. (2019) The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. *Genes Dev.*, **33**, 482–497.

55. Masvidal,L., Hulea,L. and Furic,L. (2017) Ivan Topisirovic & Ola Larsson (2017) mTOR-sensitive translation: Cleared fog reveals more trees. *RNA Biol.*, **14**, 1299–1305.

56. Gandin,V., Masvidal,L., Hulea,L., Gravel,S.-P., Cargnello,M., McLaughlan,S., Cai,Y., Balanathan,P., Morita,M., Rajakumar,A., *et al.* (2016) nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. *Genome Res.*, **26**, 636–48.

CHAPTER FOUR

Conclusions and Perspectives

This chapter contains a mini review article previously published.

Hsin-Sung Yeh and Jeongsik Yong (2020) *Journal of Lipid and Atherosclerosis* 9(1):8-22.

Hsin-Sung Yeh wrote this chapter in its entirety

Functional Difference between U2AF1 Isoforms and Its Potential Roles in Cancer Biology

Splicing factor U2AF1 defines a 3'-splice site around a polypyrimidine tract and regulates alternative mRNA processing[1–4]. As briefly discussed in Chapter 3, mutations and changes in the expression of U2AF1 affect transcriptome-wide AS events, thus altering the functional proteome[5]. Specifically, U2AF1 has recurrent somatic mutations in ~11% of myelodysplastic syndromes (MDS) patients as well as in various other cancer types[6–8]. The mutated U2AF1 proteins show altered sequence binding preference and thus AS events[6,9]. Interestingly, although it is still unclear how U2AF1 mutants contribute to tumorigenesis, mutant-expressing cells tend to be more sensitive to splicing inhibitors[10]. Moreover, it has been shown that myc-driven cancer cells are vulnerable to perturbation of splicing components including U2AF1[11]. Together, these studies highlight the splicing machinery, specifically U2AF1, as an important target for cancer research and therapeutics.

As we demonstrated in Chapter 3, while most studies on U2AF1 do not differentiate the two isoforms due to their high similarity and the U2AF1**a**-dominant expression in most cellular model systems, the two U2AF1 isoforms indeed function differently in AS regulations. For example, also demonstrated in the previous chapter, the isoform-specific AS regulations affect the expression of select genes by controlling their translation efficiency. Therefore, given U2AF1's association with cancer biology, the

next step for this project could be to determine whether this functional difference between the two isoforms is significant in U2AF1's role in cancer. Of note, there are roughly 80 protein coding genes that contain duplicate tandem exons in humans. Since in most cases the tandem exons are highly similar, and the switching of the tandem exons cannot be detected by conventional gene expression analysis using RNA-Seq or western blot, the physiological relevance of the seesawing of most tandem exons has not been elucidated. Nevertheless, as demonstrated in well-known studies on *PKM* (pyruvate kinase muscle type), *FGFR2* (fibroblast growth factor receptor 2), and *FGFR3* genes, mutually exclusive, duplicate tandem exons can function differently and drastically alter cellular pathways in cancers[12–17].

Intriguingly, our interrogations on The Cancer Genome Atlas (TCGA) data collection show that the biased increase of U2AF1**a** expression compared to the **b** isoform expression is widespread across many types of cancer (Fig. 1A). Particularly, the difference in the expression of the two isoforms is most significant in Breast Invasive Carcinoma (BRCA) data collection (Fig. 1A-B). Furthermore, triple-negative breast cancer (TNBC) patients show more significant differences of U2AF1 isoform expression, suggesting that a higher stoichiometry of U2AF1**a** over **b** isoform is critical in TNBC transcriptome homeostasis (Fig. 1). The cellular mTORC1 activity, often deregulated in TNBC and other breast cancer subtypes, selectively drives U2AF1**a** upregulation, which will alter U2AF1**a**-dependent AS events in general. Thus, these findings establish the regulatory axis of mTORC1-modulating of U2AF1 isoforms—transcriptome-wide AS events in breast cancer. AS is a hallmark of cancer but the current understandings of AS

in the cancer transcriptome and its cancer relevance are far from complete. Moreover, U2AF1 isoform functional differences in the context of cancer and how U2AF1**a** upregulation plays a role in driving cancer-prone transcriptome reprogramming have not been studied. Together, these establish strong scientific premises for future researchers to probe further how U2AF1 isoform expressions impact cancer biology, particularly in TNBC, through AS regulations.

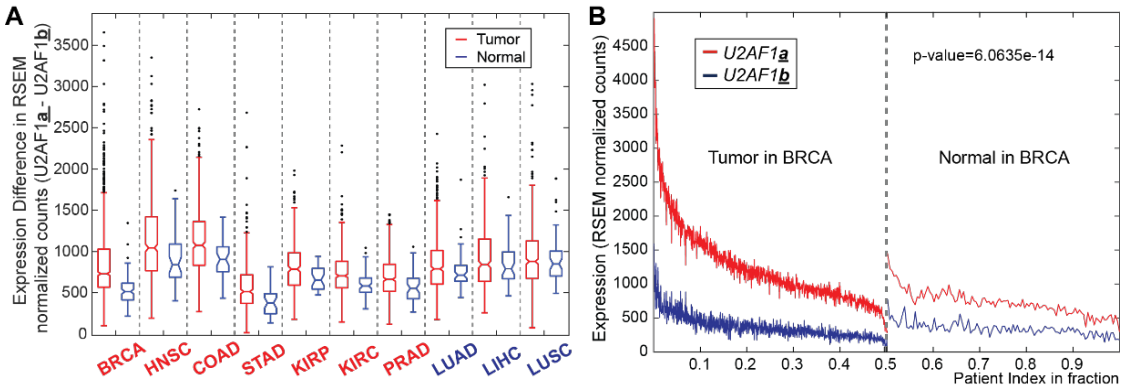


Figure 1. Interrogation of *U2AF1* isoform expressions in TCGA RNA-seq data collections. (A) Expression differences between *U2AF1**a*** and *U2AF1**b*** in tumor and normal tissues are plotted across many types of cancer in TCGA data collections. Red-colored cancer types show a significant increase of *U2AF1**a*** over **b** isoform in tumor. Dark blue-colored cancer types do not show a significant increase of *U2AF1**a*** expression. BRCA dataset shows the most significant differences in *U2AF1* isoform expressions. P-values for each dataset are as follows; BRCA 6.06e-14, HNSC 3.6e-3, COAD 1.5e-3, STAD 5.1e-4, KIRP 0.013, KIRC 2.4e-5, LUAD 0.036, LIHC 0.18, LUSC 0.49. **(B)** Expression of U2AF1**a** and U2AF1**b** in individual tumor and normal tissues of BRCA were plotted. BRCA data were collected from more than 1,000 tumor samples and 106 normal tissues.

Thus, the x-axis is presented as the fraction of patient index. The patient index was sorted by the differences of the two isoform expressions.

The Role of Post-Transcriptional Regulations in mTORC1 Biology and Beyond

As presented in Chapter 1, it has been well-established that mTORC1 regulates a variety of cellular processes through controlling the translation and transcription activities of its downstream effector genes. And our works presented in Chapters 2–3 have shown that mTORC1 signaling also has a role in regulating transcriptome-wide post-transcriptional regulations, both in APA and AS. For mTORC1-activated 3'-UTR APA events, we have shown that in select cases, a significant increase in protein synthesis without significant changes in their mRNA transcript levels was observed. Pathway enrichment analysis on these 3'-UTR-shortened genes revealed that ubiquitin-mediated proteolysis pathway is the most targeted pathway by mTORC1-mediated 3'-UTR shortening[18]. Thus, with UTR-APA analysis on mTORC1-activated transcriptome, a pathway activated by mTORC1 that was previously not associated with mTORC1 activity was revealed. Moreover, with the integration of PAS-Seq in our UTR-APA analysis, we demonstrated that mTORC1 upregulates CCAAT/enhancer binding protein gamma (CEBPG) through 3'-UTR shortening; this upregulation of CEBPG by mTORC1 is critical in protecting cells against endothelial reticulum stress[19]. Furthermore, we showed that cellular mTORC1 activity regulates the expression of U2AF1 isoforms (U2AF1a v. U2AF1b), which in turn changes of alternative splicing in 5'-UTR of many genes that significantly affect translation efficiency (Fig. 2)[20].

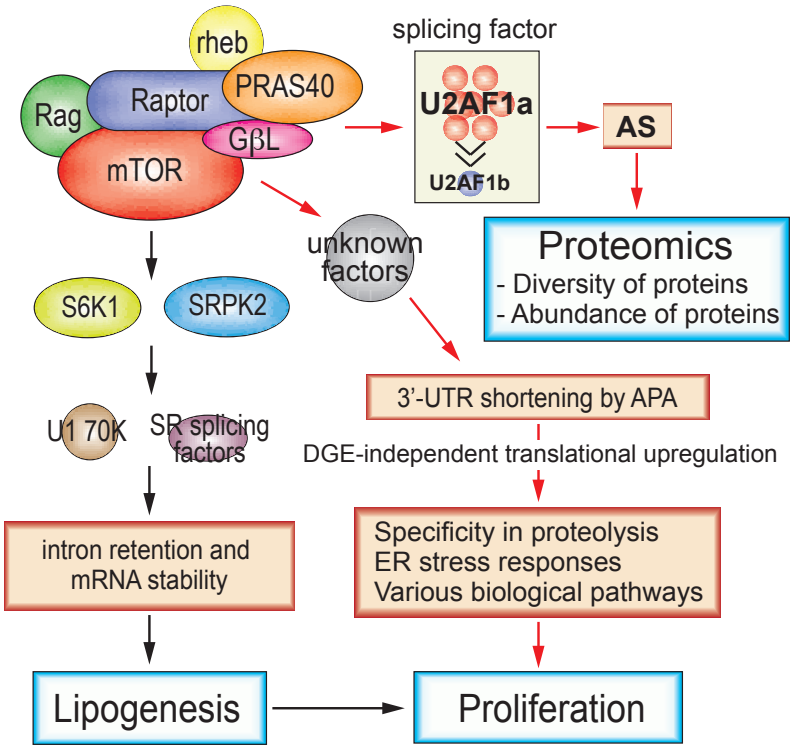


Figure 2. Illustration of how mTORC1-mediated post-transcriptional regulations play a role in controlling various cellular pathways and physiological outcomes. Recent studies have shown that mTORC1 controls the AS and APA of select genes, affecting their expressions. These can lead to changes in cellular biology, *e.g.* proliferation.

On top of our research, others have also reported the role of mTOR in modulating post-transcriptional regulations. For example, Passacantilli *et al*. reported that upon chemical mTOR inhibition, extensive changes in the transcriptome were observed in Ewing sarcoma cells using microarray analysis. Particularly, 1,440 AS events were detected in 918 genes[21]. These data were obtained by microarray technology, which can only identify previously annotated AS events. This suggests that we could expect to observe more mTOR-mediated AS events if unbiased sequencing technologies such as RNA-Seq are used for analyzing these biological samples. Nonetheless, their work demonstrates that mTOR activity indeed regulate transcriptome-wide alternative splicing events. Furthermore, Lee *et al*. also reported that chemical inhibition of mTORC1 leads to malfunction in the splicing of select genes. With improper splicing, introns are retained in these genes, activating nonsense-mediated pathway for the degradation of the transcripts (Fig. 2)[22].

Lee *et al*. demonstrated that, through S6K1 phosphorylation, mTORC1 regulates the activity of SRSF (serine and arginine rich splicing factor) protein kinase 2 (SRPK2), a key regulator of a series of splicing factors. When mTORC1 is downregulated, SRPK2 activity is inhibited, preventing the proper function of downstream splicing factors, causing splicing dysregulation and thus downregulation of a number of lipogenic genes (Fig. 2)[22]. This study showed that mTORC1, through regulating the activities of splicing regulators, can modulate gene expression at the level of splicing regulation. Passacantilli *et al*. also demonstrated that the transcriptome-wide alternative splicing events they observed in ES cells upon mTOR inhibition contribute to drug resistance in that

particular cancer cell line[21]. Thus, these reports not only reveal mTORC1's role in splicing regulation, but that these mTORC1-mediated splicing regulations have significant physiological impacts.

Taken together, these studies, and ours, add post-transcriptional regulation as another layer of gene expression regulation to our understanding of the multi-faceted functions of mTORC1 (Fig. 2). Moreover, analyses of mTORC1-mediated post-transcriptional regulations in various cellular contexts have enabled researchers to not only gain new mechanistic insights into mTORC1's control over previously associated biological processes, but also to discover new cellular signaling pathways that are regulated by mTORC1 and make mechanistic connections that were previously masked due to a lack of analyses focusing on post-transcriptional regulations.

Post-Transcriptional Regulations Beyond mTORC1 Biology

It has been a common knowledge that if we apply the simplistic understanding of the central dogma and the one-gene one-enzyme theory in molecular biology to our understanding of gene expression regulation in cells, we may fail to fully capture the complex landscape of gene expression regulations and the dynamics of functional proteome in cells. As discussed above, post-transcriptional regulations play important roles in the diversity and dynamics of the functional proteomes of biological systems. The physiological impacts of these post-transcriptional regulations can be quite significant and should not be neglected when we study the expression of genes in cells. Nonetheless, most commonly, when researchers perform transcriptome profiling with

technologies such as RNA-Seq, only transcript level analyses are carried out. Moreover, most, if not all, transcript-based biomarkers presently available are designed according to the transcript level profiling of the samples. Indeed, mRNA transcript level changes can suggest physiological outcomes due to protein level changes, yet often times transcript level changes do not lead to corresponding protein level changes. And intriguingly, for many of such cases, post-transcriptional regulations are the reason for these seeming discrepancies[23]. Therefore, given that many genes have been shown to display drastically different functions through AS or APA without changing their transcript levels necessarily, and that multiple user friendly and free bioinformatics tools are currently available online[24], analyses on post-transcriptional regulations with transcriptome profiling datasets should be performed routinely in order to capture the fuller picture of gene expression regulations in the biological systems of interest.

So far, we have discussed the significance of AS and APA events in cellular biology in the context of mTORC1 signaling. We have shown that the regulation of post-transcriptional events by mTORC1 is just as extensive and important to cellular biology as its regulations of transcriptional and translational events. This not only helps cement the role of post-transcriptional regulation as an important layer of gene expression regulation, it also establishes a mechanistic link between a well-studied and high-profile cellular signaling pathway, mTORC1, to the regulation of post-transcriptional events, as well as the physiological outcomes of these events. On the other hand, we believe that post-transcriptional regulations can also play as important and extensive of a role in other cellular signaling contexts as observed in mTORC1 signaling pathway. With the rising

awareness of how post-transcriptional regulations can dictate and steer cellular biology, and the increasing availability and sophistication of technologies to study post-transcriptional events, we expect to see more and more reports on the key roles of post-transcriptional events in various cellular signaling pathways and the pathogenesis of diseases in the near future.

Furthermore, apart from the two types of post-transcriptional events highlighted thus far (AS and APA), there are other types of regulations at the mRNA transcript level that can also be studied with transcriptome-profiling data. They all add to the richness of information one can obtain on top of transcript level analysis with transcriptome data, given that suitable bioinformatics tools are available. One example would be alternative transcription start site events, where the same gene uses different transcription start sites for transcription under different circumstances[25]. The consequences of alternative transcription start site events include alternative 5'-UTR composition, potentially resulting in changes in the fates of the final mRNA transcripts, as well as changes in the N-terminal ends of proteins, potentially leading to the inclusion or exclusion of signal peptides or ubiquitin sites in the final protein products. Another common type of regulation on mRNA transcripts is RNA modification/editing, where the insertion, deletion, and/or substitution of nucleotides is carried out on a transcribed RNA molecule by a set of enzymes. These editing events can lead to changes in the coding of amino acids of the final protein product, or changes in the binding sites of *trans*-acting factors, alternating the regulations on the mRNA transcripts by these factors[26]. Indeed, these regulations on mRNA molecules can have exciting biological consequences. However,

they are not very well explored in the field yet, at least not systematically. Thus, with single-nucleotide resolution sequencing experiments becoming routine for examining the transcriptomes of biological samples nowadays, researchers can and should perform analyses on post-transcriptional regulations and events on top of transcript level analysis to be better informed when studying the biological phenomena of interest.

References

1. Zamore, P. D., Patton, J. G. & Green, M. R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).
2. Singh, R., Valcarcel, J. & Green, M. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173 (1995).
3. Guth, S., Tange, T. O., Kellenberger, E. & Valcárcel, J. Dual Function for $U2AF^{35}$ in AG-Dependent Pre-mRNA Splicing. *Mol. Cell. Biol.* **21**, 7673 (2001).
4. Valcárcel, J., Gaur, R. K., Singh, R. & Green, M. R. Interaction of $U2AF^{65}$ RS Region with Pre-mRNA of Branch Point and Promotion Base Pairing with U2 snRNA. *Science* **273**, 1706 (1996).
5. Shirai, C. L. *et al.* Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell* **27**, 631–643 (2015).
6. Yip, B. H. *et al.* The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J Clin Invest* **127**, 2206–2221 (2017).
7. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
8. Jenkins, J. L. & Kielkopf, C. L. Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. *Trends in Genetics* **33**, 336–348 (2017).
9. Park, S. M. *et al.* U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. *Molecular Cell* **62**, 479–490 (2016).

10. Shirai, C. L. *et al.* Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nature Communications* **8**, 14060 (2017).

11. Hsu, T. Y.-T. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525**, 384–388 (2015).

12. Dayton, T. L., Jacks, T. & Vander Heiden, M. G. PKM2, cancer metabolism, and the road ahead. *EMBO reports* **17**, 1721–1730 (2016).

13. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–3 (2008).

14. Israelsen, W. J. & Vander Heiden, M. G. Pyruvate kinase: Function, regulation and role in cancer. *Seminars in Cell & Developmental Biology* **43**, 43–51 (2015).

15. Warzecha, C. C., Sato, T. K., Nabet, B., Hogenesch, J. B. & Carstens, R. P. ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. *Molecular Cell* **33**, 591–601 (2009).

16. Shirakihara, T. *et al.* TGF-β regulates isoform switching of FGF receptors and epithelial–mesenchymal transition. *The EMBO Journal* **30**, 783–795 (2011).

17. Holzmann, K. *et al.* Alternative Splicing of Fibroblast Growth Factor Receptor IgIII Loops in Cancer. *Journal of nucleic acids* **2012**, 950508 (2012).

18. Chang, J.-W. *et al.* mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. *Nat Commun* **6**, 7218 (2015).

19. Chang, J.-W. *et al.* An integrative model for alternative polyadenylation, IntMAP, delineates mTOR-modulated endoplasmic reticulum stress response. *Nucleic Acids Research* **46**, 5996–6008 (2018).

20. Chang, J.-W. *et al.* mTOR-regulated U2af1 tandem exon splicing specifies transcriptome features for translational control. *Nucleic Acids Research* (2019) doi:10.1093/nar/gkz761.

21. Passacantilli, I., Frisone, P., De Paola, E., Fidaleo, M. & Paronetto, M. P. hnRNPM guides an alternative splicing program in response to inhibition of the PI3K/AKT/mTOR pathway in Ewing sarcoma cells. *Nucleic Acids Res* **45**, 12270–12284 (2017).

22. Lee, G. *et al.* Post-transcriptional Regulation of De Novo Lipogenesis by mTORC1-S6K1-SRPK2 Signaling. *Cell* **171**, 1545-1558.e18 (2017).

23. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**, 227–232 (2012).

24. Hsin-Sung, Y. *et al.* Analyses of alternative polyadenylation: from old school biochemistry to high-throughput technologies. *BMB Reports* **50**, 201–207 (4).

25. Reyes, A. & Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res* **46**, 582–592 (2018).

26. Lerner, T., Papavasiliou, F. N. & Pecori, R. RNA Editors, Cofactors, and mRNA Targets: An Overview of the C-to-U RNA Editing Machinery and Its Implication in Human Disease. *Genes (Basel)* **10**, 13 (2018).

COMPLETE BIBLIOGRAPHY

Abdel-Ghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A., Reddy, A.S.N., 2016. A survey of the sorghum transcriptome using single-molecule long reads. Nature Communications 7, 11706. https://doi.org/10.1038/ncomms11706

Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M., Baltimore, D., 1980. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. Cell 20, 293–301. https://doi.org/10.1016/0092-8674(80)90615-7

Appenzeller-Herzog, C., Hall, M.N., 2012. Bidirectional crosstalk between endoplasmic reticulum stress and mTOR signaling. Trends in Cell Biology 22, 274–282. https://doi.org/10.1016/j.tcb.2012.02.006

Aramburu, J., Ortells, M.C., Tejedor, S., Buxadé, M., López-Rodríguez, C., 2014. Transcriptional regulation of the stress response by mTOR. Sci. Signal. 7, re2. https://doi.org/10.1126/scisignal.2005326

Au, K.F., Sebastiano, V., Afshar, P.T., Durruthy, J.D., Lee, L., Williams, B.A., van Bakel, H., Schadt, E.E., Reijo-Pera, R.A., Underwood, J.G., Wong, W.H., 2013. Characterization of the human ESC transcriptome by hybrid sequencing. Proc Natl Acad Sci USA 110, E4821. https://doi.org/10.1073/pnas.1320101110

Au, K.F., Underwood, J.G., Lee, L., Wong, W.H., 2012. Improving PacBio Long Read Accuracy by Short Read Alignment. PLOS ONE 7, e46679. https://doi.org/10.1371/journal.pone.0046679

Bartke, T., Pohl, C., Pyrowolakis, G., Jentsch, S., 2004. Dual role of BRUCE as an antiapoptotic IAP and a chimeric E2/E3 ubiquitin ligase. Mol. Cell 14. https://doi.org/10.1016/j.molcel.2004.05.018

Bava, F.-A., Eliscovich, C., Ferreira, P.G., Miñana, B., Ben-Dov, C., Guigó, R., Valcárcel, J., Méndez, R., 2013. CPEB1 coordinates alternative 3′-UTR formation with translational regulation. Nature 495, 121. https://doi.org/10.1038/nature11901

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., Gautheret, D., 2000. Patterns of variant polyadenylation signal usage in human genes. Genome Res. 10. https://doi.org/10.1101/gr.10.7.1001

Beck, A.H., Weng, Z., Witten, D.M., Zhu, S., Foley, J.W., Lacroute, P., Smith, C.L., Tibshirani, R., van de Rijn, M., Sidow, A., West, R.B., 2010. 3′-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples. PLOS ONE 5, e8768. https://doi.org/10.1371/journal.pone.0008768

Ben-Sahra, I., Hoxhaj, G., Ricoult, S.J.H., Asara, J.M., Manning, B.D., 2016. mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. Science 351, 728. https://doi.org/10.1126/science.aad0489

Ben-Sahra, I., Manning, B.D., 2017. mTORC1 signaling and the metabolic control of cell growth. Current Opinion in Cell Biology 45, 72–82. https://doi.org/10.1016/j.ceb.2017.02.012

Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. Nature Reviews Genetics 15, 163–175. https://doi.org/10.1038/nrg3662
Berkovits, B.D., Mayr, C., 2015. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature 522, 363. https://doi.org/10.1038/nature14321
Birol, I., Raymond, A., Chiu, R., Nip, K.M., Jackman, S.D., Kreitzman, M., Docking, T.R., Ennis, C.A., Robertson, A.G., Karsan, A., 2015. Kleat: cleavage site analysis of transcriptomes. Pac Symp Biocomput 347–358.
Boutet, S.C., Cheung, T.H., Quach, N.L., Liu, L., Prescott, S.L., Edalati, A., Iori, K., Rando, T.A., 2012. Alternative Polyadenylation Mediates MicroRNA Regulation of Muscle Stem Cell Function. Cell Stem Cell 10, 327–336. https://doi.org/10.1016/j.stem.2012.01.017
Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34, 525–527. https://doi.org/10.1038/nbt.3519
Cao, S.S., Kaufman, R.J., 2014. Endoplasmic Reticulum Stress and Oxidative Stress in Cell Fate Decision and Human Disease. Antioxidants & Redox Signaling 21, 396–413. https://doi.org/10.1089/ars.2014.5851
Carter, M.S., Doskow, J., Morris, P., Li, S., Nhim, R.P., Sandstedt, S., Wilkinson, M.F., 1995. A Regulatory Mechanism That Detects Premature Nonsense Codons in T-cell Receptor Transcripts in Vivo Is Reversed by Protein Synthesis Inhibitors in Vitro. Journal of Biological Chemistry 270, 28995–29003. https://doi.org/10.1074/jbc.270.48.28995
Castets, P., 2013. Sustained activation of mTORC1 in skeletal muscle inhibits constitutive and starvation-induced autophagy and causes a severe, late-onset myopathy. Cell Metab. 17. https://doi.org/10.1016/j.cmet.2013.03.015
Chang, H., Lim, J., Ha, M., Kim, V.N., 2014. TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3′ End Modifications. Molecular Cell 53, 1044–1052. https://doi.org/10.1016/j.molcel.2014.02.007
Chang, J.-W., Yeh, H.-S., Park, M., Erber, L., Sun, J., Cheng, S., Bui, A.M., Fahmi, N.A., Nasti, R., Kuang, R., Chen, Y., Zhang, W., Yong, J., 2019. mTOR-regulated U2af1 tandem exon splicing specifies transcriptome features for translational control. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz761
Chang, J.-W., Zhang, W., Yeh, H.-S., de Jong, E.P., Jun, S., Kim, K.-H., Bae, S.S., Beckman, K., Hwang, T.H., Kim, K.-S., Kim, D.-H., Griffin, T.J., Kuang, R., Yong, J., 2015. mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. Nat Commun 6, 7218. https://doi.org/10.1038/ncomms8218
Chang, J.-W., Zhang, W., Yeh, H.-S., Park, M., Yao, C., Shi, Y., Kuang, R., Yong, J., 2018. An integrative model for alternative polyadenylation, IntMAP, delineates mTOR-modulated endoplasmic reticulum stress response. Nucleic Acids Research 46, 5996–6008. https://doi.org/10.1093/nar/gky340
Chen, D., Shan, J., Zhu, W.G., Qin, J., Gu, W., 2010. Transcription-independent ARF regulation in oncogenic stress-mediated p53 responses. Nature 464. https://doi.org/10.1038/nature08820

Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., Farrow, S., Poudel, P., Burden, F., Jansen, S.B.G., Astle, W.J., Attwood, A., Bariana, T., de Bono, B., Breschi, A., Chambers, J.C., Choudry, F.A., Clarke, L., Coupland, P., van der Ent, M., Erber, W.N., Jansen, J.H., Favier, R., Fenech, M.E., Foad, N., Freson, K., van Geet, C., Gomez, K., Guigo, R., Hampshire, D., Kelly, A.M., Kerstens, H.H.D., Kooner, J.S., Laffan, M., Lentaigne, C., Labalette, C., Martin, T., Meacham, S., Mumford, A., Nürnberg, S.T., Palumbo, E., van der Reijden, B.A., Richardson, D., Sammut, S.J., Slodkowicz, G., Tamuri, A.U., Vasquez, L., Voss, K., Watt, S., Westbury, S., Flicek, P., Loos, R., Goldman, N., Bertone, P., Read, R.J., Richardson, S., Cvejic, A., Soranzo, N., Ouwehand, W.H., Stunnenberg, H.G., Frontini, M., Rendon, A., 2014. Transcriptional diversity during lineage commitment of human blood progenitors. Science 345, 1251033. https://doi.org/10.1126/science.1251033

Chio, I.I., 2012. TRADD contributes to tumour suppression by regulating ULF-dependent p19Arf ubiquitylation. Nat. Cell Biol. 14. https://doi.org/10.1038/ncb2496

CHUSAINOW, J., AJUH, P.M., TRINKLE-MULCAHY, L., SLEEMAN, J.E., ELLENBERG, J., LAMOND, A.I., 2005. FRET analyses of the U2AF complex localize the U2AF35/U2AF65 interaction in vivo and reveal a novel self-interaction of U2AF35. RNA 11, 1201–1214. https://doi.org/10.1261/rna.7277705

Ciolli Mattioli, C., Rom, A., Franke, V., Imami, K., Arrey, G., Terne, M., Woehler, A., Akalin, A., Ulitsky, I., Chekulaeva, M., 2018. Alternative 3′ UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. Nucleic Acids Research 47, 2560–2573. https://doi.org/10.1093/nar/gky1270

Codogno, P., Meijer, A.J., 2005. Autophagy and signaling: their role in cell survival and cell death. Cell Death & Differentiation 12, 1509–1518. https://doi.org/10.1038/sj.cdd.4401751

Cooper, C., Henderson, A., Artandi, S., Avitahl, N., Calame, K., 1995. Ig/EBP (C/EBPγ) is a transdominant negative inhibitor of C/EBP family transcriptional activators. Nucleic Acids Research 23, 4371–4377. https://doi.org/10.1093/nar/23.21.4371

Cunningham, J.T., Rodgers, J.T., Arlow, D.H., Vazquez, F., Mootha, V.K., Puigserver, P., 2007. mTOR controls mitochondrial oxidative function through a YY1–PGC-1α transcriptional complex. Nature 450, 736. https://doi.org/10.1038/nature06322

Darnell, J.E., 2013. Reflections on the history of pre-mRNA processing and highlights of current knowledge: A unified picture. RNA 19, 443–460. https://doi.org/10.1261/rna.038596.113

David, C.J., Boyne, A.R., Millhouse, S.R., Manley, J.L., 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. Genes & Development 25, 972–983. https://doi.org/10.1101/gad.2038011

David, C.J., Manley, J.L., 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes & Development 24, 2343–2364. https://doi.org/10.1101/gad.1973010

Dayton, T.L., Jacks, T., Vander Heiden, M.G., 2016. PKM2, cancer metabolism, and the road ahead. EMBO reports 17, 1721–1730. https://doi.org/10.15252/embr.201643300

Denton, R.M., Pullen, T.J., Armstrong, C.T., Heesom, K.J., Rutter, G.A., 2016. Calcium-insensitive splice variants of mammalian E1 subunit of 2-oxoglutarate dehydrogenase complex with tissue-specific patterns of expression. Biochemical Journal 473, 1165–1178. https://doi.org/10.1042/BCJ20160135

Derti, A., Garrett-Engele, P., MacIsaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., Babak, T., 2012. A quantitative atlas of polyadenylation in five mammals. Genome Research 22, 1173–1183. https://doi.org/10.1101/gr.132563.111

Di Giammartino, D.C., Nishida, K., Manley, J.L., 2011. Mechanisms and consequences of alternative polyadenylation. Mol Cell 43, 853–866. https://doi.org/10.1016/j.molcel.2011.08.017

Durand, S., Lykke-Andersen, J., 2013. Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. Nature Structural & Molecular Biology 20, 702–709. https://doi.org/10.1038/nsmb.2575

Düvel, K., Yecies, J.L., Menon, S., Raman, P., Lipovsky, A.I., Souza, A.L., Triantafellow, E., Ma, Q., Gorski, R., Cleaver, S., Vander Heiden, M.G., MacKeigan, J.P., Finan, P.M., Clish, C.B., Murphy, L.O., Manning, B.D., 2010. Activation of a metabolic gene regulatory network downstream of mTOR complex 1. Mol Cell 39, 171–183. https://doi.org/10.1016/j.molcel.2010.06.022

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., Hood, L., 1980. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell 20, 313–319. https://doi.org/10.1016/0092-8674(80)90617-0

Ekim, B., 2011. mTOR kinase domain phosphorylation promotes mTORC1 signaling, cell growth, and cell cycle progression. Mol. Cell Biol. 31. https://doi.org/10.1128/MCB.05437-11

El Marabti, E., Younis, I., 2018. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. Front Mol Biosci 5, 80–80. https://doi.org/10.3389/fmolb.2018.00080

Elkon, R., Ugalde, A.P., Agami, R., 2013. Alternative cleavage and polyadenylation: extent, regulation and function. Nature Reviews Genetics 14, 496. https://doi.org/10.1038/nrg3482

Fabian, M.R., Sonenberg, N., Filipowicz, W., 2010. Regulation of mRNA Translation and Stability by microRNAs. Annual Review of Biochemistry 79, 351–379. https://doi.org/10.1146/annurev-biochem-060308-103103

Fei, D.L., Zhen, T., Durham, B., Ferrarone, J., Zhang, T., Garrett, L., Yoshimi, A., Abdel-Wahab, O., Bradley, R.K., Liu, P., Varmus, H., 2018. Impaired hematopoiesis and leukemia development in mice with a conditional knock-in allele of a mutant splicing factor gene U2af1. Proc Natl Acad Sci USA 115, E10437. https://doi.org/10.1073/pnas.1812669115

Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9. https://doi.org/10.1038/nrg2290

Fingar, D.C., Salama, S., Tsou, C., Harlow, E., Blenis, J., 2002. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. Genes Dev. 16. https://doi.org/10.1101/gad.995802

Flavell, S.W., Kim, T.-K., Gray, J.M., Harmin, D.A., Hemberg, M., Hong, E.J., Markenscoff-Papadimitriou, E., Bear, D.M., Greenberg, M.E., 2008. Genome-Wide Analysis of MEF2 Transcriptional Program Reveals Synaptic Target Genes and Neuronal Activity-Dependent Polyadenylation Site Selection. Neuron 60, 1022–1038. https://doi.org/10.1016/j.neuron.2008.11.029

Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., Fu, X.-D., 2011. A multiplex RNA-seq strategy to profile poly(A+) RNA: Application to analysis of transcription response and 3′ end formation. Genomics 98, 266–271. https://doi.org/10.1016/j.ygeno.2011.04.003

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 Research 21, 741–747. https://doi.org/10.1101/gr.115295.110

Gandin, V., Masvidal, L., Hulea, L., Gravel, S.-P., Cargnello, M., McLaughlan, S., Cai, Y., Balanathan, P., Morita, M., Rajakumar, A., Furic, L., Pollak, M., Porco, J.A., St-Pierre, J., Pelletier, J., Larsson, O., Topisirovic, I., 2016. nanoCAGE reveals 5′ UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. Genome Research 26, 636–648. https://doi.org/10.1101/gr.197566.115

Graham, R.R., Kyogoku, C., Sigurdsson, S., Vlasova, I.A., Davies, L.R.L., Baechler, E.C., Plenge, R.M., Koeuth, T., Ortmann, W.A., Hom, G., Bauer, J.W., Gillett, C., Burtt, N., Cunninghame Graham, D.S., Onofrio, R., Petri, M., Gunnarsson, I., Svenungsson, E., Rönnblom, L., Nordmark, G., Gregersen, P.K., Moser, K., Gaffney, P.M., Criswell, L.A., Vyse, T.J., Syvänen, A.-C., Bohjanen, P.R., Daly, M.J., Behrens, T.W., Altshuler, D., 2007. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci USA 104, 6758. https://doi.org/10.1073/pnas.0701266104

Grant, M., Boyd, S., 2014. CVX: Matlab Software for Disciplined Convex Programming, version 2.1.

Grant, M., Boyd, S., 2008. Graph implementations for nonsmooth convex programs, in: Blondel, V., Boyd, S., Kimura, H. (Eds.), Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences. Springer-Verlag Limited, pp. 95–110.

Grant, M.C., Boyd, S.P., 2008. Graph Implementations for Nonsmooth Convex Programs, in: Blondel, V.D., Boyd, S.P., Kimura, H. (Eds.), Recent Advances in Learning and Control. Springer London, London, pp. 95–110.

Grassi, E., Mariella, E., Lembo, A., Molineris, I., Provero, P., 2016. Roar: detecting alternative polyadenylation with standard mRNA sequencing libraries. BMC Bioinformatics 17, 423. https://doi.org/10.1186/s12859-016-1254-8

GRAVELEY, B.R., HERTEL, K.J., MANIATIS, T., 2001. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA 7, 806–818. https://doi.org/10.1017/S1355838201010317
Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E., Pesole, G., 2009. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Research 38, D75–D80. https://doi.org/10.1093/nar/gkp902
Gruber, A.R., Martin, G., Keller, W., Zavolan, M., 2014. Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. Wiley Interdiscip. Rev. RNA 5. https://doi.org/10.1002/wrna.1206
Gruber, A.R., Martin, G., Keller, W., Zavolan, M., 2012. Cleavage factor Im is a key regulator of 3′ UTR length. RNA Biology 9, 1405–1412. https://doi.org/10.4161/rna.22570
Guertin, D.A., Sabatini, D.M., 2009. The pharmacology of mTOR inhibition. Sci. Signal. 2. https://doi.org/10.1126/scisignal.267pe24
Guertin, D.A., Sabatini, D.M., 2007. Defining the Role of mTOR in Cancer. Cancer Cell 12, 9–22. https://doi.org/10.1016/j.ccr.2007.05.008
Guo, Y., 2013. TSC1 involvement in bladder cancer: diverse effects and therapeutic implications. J. Pathol. 230. https://doi.org/10.1002/path.4176
Guth, S., Martínez, C., Gaur, R.K., Valcárcel, J., 1999. Evidence for Substrate-Specific Requirement of the Splicing Factor U2AF^{35} and for Its Function after Polypyrimidine Tract Recognition by U2AF^{65}. Mol. Cell. Biol. 19, 8263. https://doi.org/10.1128/MCB.19.12.8263
Guth, S., Tange, T.O., Kellenberger, E., Valcárcel, J., 2001. Dual Function for U2AF^{35} in AG-Dependent Pre-mRNA Splicing. Mol. Cell. Biol. 21, 7673. https://doi.org/10.1128/MCB.21.22.7673-7681.2001
Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., Pena, J.T.G., Nusbaum, J.D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., Tuschl, T., 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA 17, 1697–1712. https://doi.org/10.1261/rna.2799511
Hagting, A., 2002. Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. J. Cell Biol. 157. https://doi.org/10.1083/jcb.200111001
Hardie, D.G., 2007. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. Nature Reviews Molecular Cell Biology 8, 774. https://doi.org/10.1038/nrm2249
Herzel, L., Ottoz, D.S.M., Alpert, T., Neugebauer, K.M., 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nature Reviews Molecular Cell Biology 18, 637. https://doi.org/10.1038/nrm.2017.63
Hicks, M.J., Yang, C.-R., Kotlajich, M.V., Hertel, K.J., 2006. Linking Splicing to Pol II Transcription Stabilizes Pre-mRNAs and Influences Splicing Patterns. PLOS Biology 4, e147. https://doi.org/10.1371/journal.pbio.0040147

Hinnebusch, A.G., Ivanov, I.P., Sonenberg, N., 2016. Translational control by 5′-untranslated regions of eukaryotic mRNAs. Science 352, 1413. https://doi.org/10.1126/science.aad9868
Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G., Tian, B., 2012. Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. Nature Methods 10, 133. https://doi.org/10.1038/nmeth.2288
Howell, J.J., Ricoult, S.J., Ben-Sahra, I., Manning, B.D., 2013. A growing role for mTOR in promoting anabolic metabolism. Biochem. Soc. Trans. 41. https://doi.org/10.1042/BST20130041
Hsieh, A.C., Liu, Y., Edlind, M.P., Ingolia, N.T., Janes, M.R., Sher, A., Shi, E.Y., Stumpf, C.R., Christensen, C., Bonham, M.J., Wang, S., Ren, P., Martin, M., Jessen, K., Feldman, M.E., Weissman, J.S., Shokat, K.M., Rommel, C., Ruggero, D., 2012. The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature 485, 55. https://doi.org/10.1038/nature10912
Hsin-Sung, Y., and Jeongsik, Y., 4. Alternative Polyadenylation of mRNAs: 3′-Untranslated Region Matters in Gene Expression. Mol. Cells 39, 281–285.
Hsin-Sung, Y., lt, sup, gt, lt, sup, gt, Wei, Z., lt, sup, gt, lt, sup, gt, and Jeongsik, Y., lt, sup, gt, lt, sup, gt, 4. Analyses of alternative polyadenylation: from old school biochemistry to high-throughput technologies. BMB Reports 50, 201–207.
Hsu, T.Y.-T., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., Echeverria, G.V., Sun, T., Kurley, S.J., Tyagi, S., Karlin, K.L., Dominguez-Vidaña, R., Hartman, J.D., Renwick, A., Scorsone, K., Bernardi, R.J., Skinner, S.O., Jain, A., Orellana, M., Lagisetti, C., Golding, I., Jung, S.Y., Neilson, J.R., Zhang, X.H.-F., Cooper, T.A., Webb, T.R., Neel, B.G., Shaw, C.A., Westbrook, T.F., 2015. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. Nature 525, 384–388. https://doi.org/10.1038/nature14985
Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., Ares, M., Jr., Yeo, G.W., 2012. Integrative Genome-wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins. Cell Reports 1, 167–178. https://doi.org/10.1016/j.celrep.2012.02.001
Huggins, C.J., Malik, R., Lee, S., Salotti, J., Thomas, S., Martin, N., Quiñones, O.A., Alvord, W.G., Olanich, M.E., Keller, J.R., Johnson, P.F., 2013. C/EBPγ Suppresses Senescence and Inflammatory Gene Expression by Heterodimerizing with C/EBPβ. Mol. Cell. Biol. 33, 3242. https://doi.org/10.1128/MCB.01674-12
Huggins, C.J., Mayekar, M.K., Martin, N., Saylor, K.L., Gonit, M., Jailwala, P., Kasoji, M., Haines, D.C., Quiñones, O.A., Johnson, P.F., 2016. C/EBPγ Is a Critical Regulator of Cellular Stress Response Networks through Heterodimerization with ATF4. Mol. Cell. Biol. 36, 693. https://doi.org/10.1128/MCB.00911-15
Israelsen, W.J., Vander Heiden, M.G., 2015. Pyruvate kinase: Function, regulation and role in cancer. Seminars in Cell & Developmental Biology 43, 43–51. https://doi.org/10.1016/j.semcdb.2015.08.004
Jan, C.H., Friedman, R.C., Ruby, J.G., Bartel, D.P., 2011. Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. Nature 469, 97–101. https://doi.org/10.1038/nature09616

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Vrielink, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K., Rubinsztein, D.C., Agami, R., 2012. The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. Cell 149, 538–553. https://doi.org/10.1016/j.cell.2012.03.022

Jenkins, J.L., Kielkopf, C.L., 2017. Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. Trends in Genetics 33, 336–348. https://doi.org/10.1016/j.tig.2017.03.001

Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., Tian, B., 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl Acad. Sci. USA 106. https://doi.org/10.1073/pnas.0900028106

Ji, Z., Tian, B., 2009. Reprogramming of 3′ Untranslated Regions of mRNAs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. PLOS ONE 4, e8419. https://doi.org/10.1371/journal.pone.0008419

Jia, L., Sun, Y., 2009. RBX1/ROC1-SCF E3 ubiquitin ligase is required for mouse embryogenesis and cancer cell survival. Cell Div. 4. https://doi.org/10.1186/1747-1028-4-16

Jones, R.G., Pearce, E.J., 2017. MenTORing Immunity: mTOR Signaling in the Development and Function of Tissue-Resident Immune Cells. Immunity 46, 730–742. https://doi.org/10.1016/j.immuni.2017.04.028

Karar, J., Maity, A., 2011. PI3K/AKT/mTOR Pathway in Angiogenesis. Front Mol Neurosci 4, 51–51. https://doi.org/10.3389/fnmol.2011.00051

Kielkopf, C.L., Rodionova, N.A., Green, M.R., Burley, S.K., 2001. A Novel Peptide Recognition Mode Revealed by the X-Ray Structure of a Core U2AF35/U2AF65 Heterodimer. Cell 106, 595–605. https://doi.org/10.1016/S0092-8674(01)00480-9

Kim, J., Guan, K.-L., 2019. mTOR as a central hub of nutrient signalling and cell growth. Nature Cell Biology 21, 63–71. https://doi.org/10.1038/s41556-018-0205-1

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., Buratowski, S., 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432, 517–522. https://doi.org/10.1038/nature03041

Kim, M., You, B.-H., Nam, J.-W., 2015. Global estimation of the 3′ untranslated region landscape using RNA sequencing. Methods 83, 111–117. https://doi.org/10.1016/j.ymeth.2015.04.011

Knowles, M.A., Habuchi, T., Kennedy, W., Cuthbert-Heavens, D., 2003. Mutation spectrum of the 9q34 tuberous sclerosis gene TSC1 in transitional cell carcinoma of the bladder. Cancer Res. 63.

Kornblihtt, A.R., 2004. Shortcuts to the end. Nature Structural & Molecular Biology 11, 1156–1157. https://doi.org/10.1038/nsmb1204-1156

Kraft, C., 2003. Mitotic regulation of the human anaphase-promoting complex by phosphorylation. EMBO J. 22. https://doi.org/10.1093/emboj/cdg627

Kralovicova, J., Knut, M., Cross, N.C.P., Vorechovsky, I., 2015. Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3′ splice-site organization and activity of U2AF-related proteins. Nucleic Acids Research 43, 3747–3763. https://doi.org/10.1093/nar/gkv194

Kralovicova, J., Vorechovsky, I., 2016. Alternative splicing of U2AF1 reveals a shared repression mechanism for duplicated exons. Nucleic Acids Research 45, 417–434. https://doi.org/10.1093/nar/gkw733
Kurdi, A., Martinet, W., De Meyer, G.R.Y., 2018. mTOR Inhibition and Cardiovascular Diseases: Dyslipidemia and Atherosclerosis. Transplantation 102.
Kwiatkowski, D.J., Zhang, H., Bandura, J.L., Heiberger, K.M., Glogauer, M., el-Hashemite, N., Onda, H., 2002. A mouse model of TSC1 reveals sex-dependent lethality from liver hemangiomas, and up-regulation of p70S6 kinase activity in Tsc1 null cells. Human Molecular Genetics 11, 525–534. https://doi.org/10.1093/hmg/11.5.525
Lamming, D.W., Sabatini, D.M., 2013. A central role for mTOR in lipid homeostasis. Cell Metab. 18. https://doi.org/10.1016/j.cmet.2013.08.002
Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25
Laplante, M., Sabatini, D.M., 2013. Regulation of mTORC1 and its impact on gene expression at a glance. Journal of Cell Science 126, 1713. https://doi.org/10.1242/jcs.125773
Laplante, M., Sabatini, D.M., 2012. mTOR Signaling in Growth Control and Disease. Cell 149, 274–293. https://doi.org/10.1016/j.cell.2012.03.017
Laplante, M., Sabatini, D.M., 2009. mTOR signaling at a glance. Journal of Cell Science 122, 3589. https://doi.org/10.1242/jcs.051011
Le Pera, L., Mazzapioda, M., Tramontano, A., 2015. 3USS: a web server for detecting alternative 3′UTRs from RNA-seq experiments. Bioinformatics 31, 1845–1847. https://doi.org/10.1093/bioinformatics/btv035
Lee, G., Zheng, Y., Cho, S., Jang, C., England, C., Dempsey, J.M., Yu, Y., Liu, X., He, L., Cavaliere, P.M., Chavez, A., Zhang, E., Isik, M., Couvillon, A., Dephoure, N.E., Blackwell, T.K., Yu, J.J., Rabinowitz, J.D., Cantley, L.C., Blenis, J., 2017. Post-transcriptional Regulation of De Novo Lipogenesis by mTORC1-S6K1-SRPK2 Signaling. Cell 171, 1545-1558.e18. https://doi.org/10.1016/j.cell.2017.10.037
Lee, S.-H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C.S., Mayr, C., 2018. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature 561, 127–131. https://doi.org/10.1038/s41586-018-0465-8
Leppek, K., Das, R., Barna, M., 2018. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nature Reviews Molecular Cell Biology 19, 158–174. https://doi.org/10.1038/nrm.2017.103
Lerner, T., Papavasiliou, F.N., Pecori, R., 2018. RNA Editors, Cofactors, and mRNA Targets: An Overview of the C-to-U RNA Editing Machinery and Its Implication in Human Disease. Genes (Basel) 10, 13. https://doi.org/10.3390/genes10010013
Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., Dewey, C.N., 2009. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26, 493–500. https://doi.org/10.1093/bioinformatics/btp692

Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., Mayr, C., 2013. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. 27. https://doi.org/10.1101/gad.229328.113
Lin-Moshier, Y., 2013. Re-evaluation of the role of calcium homeostasis endoplasmic reticulum protein (CHERP) in cellular calcium signaling. J. Biol. Chem. 288. https://doi.org/10.1074/jbc.M112.405761
Liu, J., Ma, K., Gao, M., Zhang, X., Liu, B., 2011. The activation of mTOR pathway induced by inflammation accelerates the progression of atherosclerosis in hemodialysis patients. International Journal of Cardiology 152, S5–S6. https://doi.org/10.1016/j.ijcard.2011.08.479
Lutz, C.S., Moreira, A., 2011. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdiscip. Rev. RNA 2. https://doi.org/10.1002/wrna.47
Ma, L., Pati, P.K., Liu, M., Li, Q.Q., Hunt, A.G., 2014. High throughput characterizations of poly(A) site choice in plants. Methods 67, 74–83. https://doi.org/10.1016/j.ymeth.2013.06.037
Ma, X.M., Blenis, J., 2009. Molecular mechanisms of mTOR-mediated translational control. Nature Reviews Molecular Cell Biology 10, 307. https://doi.org/10.1038/nrm2672
Mangone, M., Manoharan, A.P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S.D., Mis, E., Zegar, C., Gutwein, M.R., Khivansara, V., Attie, O., Chen, K., Salehi-Ashtiani, K., Vidal, M., Harkins, T.T., Bouffard, P., Suzuki, Y., Sugano, S., Kohara, Y., Rajewsky, N., Piano, F., Gunsalus, K.C., Kim, J.K., 2010. The Landscape of C. elegans 3′UTRs. Science 329, 432. https://doi.org/10.1126/science.1191244
Manley, J.L., 1988. Polyadenylation of mRNA precursors. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 950, 1–12. https://doi.org/10.1016/0167-4781(88)90067-X
Martin, G., Gruber, A.R., Keller, W., Zavolan, M., 2012. Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. Cell Reports 1, 753–763. https://doi.org/10.1016/j.celrep.2012.05.003
Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.-B., Li, W., Wagner, E.J., 2014. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature 510, 412. https://doi.org/10.1038/nature13261
Masvidal, L., Hulea, L., Furic, L., Topisirovic, I., Larsson, O., 2017. mTOR-sensitive translation: Cleared fog reveals more trees. RNA Biology 14, 1299–1305. https://doi.org/10.1080/15476286.2017.1290041
Mata, J., 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. RNA Biology 10, 1407–1414. https://doi.org/10.4161/rna.25758
Mayer, C., Zhao, J., Yuan, X., Grummt, I., 2004. mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. Genes & Development 18, 423–434. https://doi.org/10.1101/gad.285504

Mayr, C., 2016. Evolution and Biological Roles of Alternative 3'UTRs. Trends Cell Biol 26, 227–237. https://doi.org/10.1016/j.tcb.2015.10.012
Mayr, C., Bartel, D.P., 2009. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138. https://doi.org/10.1016/j.cell.2009.06.016
Merendino, L., Guth, S., Bilbao, D., Martínez, C., Valcárcel, J., 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3′ splice site AG. Nature 402, 838–841. https://doi.org/10.1038/45602
Morris, A.R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A.S., Carvalho, B., Meijer, G.A., Agami, R., 2012. Alternative Cleavage and Polyadenylation during Colorectal Cancer Development. Clin Cancer Res 18, 5256. https://doi.org/10.1158/1078-0432.CCR-12-0543
Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. Science 320, 1344. https://doi.org/10.1126/science.1158441
Nguyen, H.D., Leong, W.Y., Li, W., Reddy, P.N.G., Sullivan, J.D., Walter, M.J., Zou, L., Graubert, T.A., 2018. Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. Cancer Res 78, 5363. https://doi.org/10.1158/0008-5472.CAN-17-3970
Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., Zou, L., 2017. Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Molecular Cell 65, 832-847.e4. https://doi.org/10.1016/j.molcel.2017.01.029
O'Grady, T., Wang, X., Höner zu Bentrup, K., Baddoo, M., Concha, M., Flemington, E.K., 2016. Global transcript structure resolution of high gene density genomes through multi-platform data integration. Nucleic Acids Research 44, e145–e145. https://doi.org/10.1093/nar/gkw629
O'Loghlen, A., González, V.M., Piñeiro, D., Pérez-Morgado, M.I., Salinas, M., Martín, M.E., 2004. Identification and molecular characterization of Mnk1b, a splice variant of human MAP kinase-interacting kinase Mnk1. Experimental Cell Research 299, 343–355. https://doi.org/10.1016/j.yexcr.2004.06.006
Oslowski, C.M., Urano, F., 2011. Chapter Four - Measuring ER Stress and the Unfolded Protein Response Using Mammalian Tissue Culture System, in: Conn, P.M. (Ed.), Methods in Enzymology. Academic Press, pp. 71–92. https://doi.org/10.1016/B978-0-12-385114-7.00004-0
Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., Milos, P.M., 2010. Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. Cell 143, 1018–1029. https://doi.org/10.1016/j.cell.2010.11.020
Ozsolak, F., Milos, P.M., 2011. RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics 12, 87–98. https://doi.org/10.1038/nrg2934
Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., Milos, P.M., 2009. Direct RNA sequencing. Nature 461, 814–818. https://doi.org/10.1038/nature08390

Pacheco, Teresa R., Coelho, M.B., Desterro, J.M.P., Mollet, I., Carmo-Fonseca, M., 2006. In Vivo Requirement of the Small Subunit of U2AF for Recognition of a Weak 3′ Splice Site. Mol. Cell. Biol. 26, 8183. https://doi.org/10.1128/MCB.00350-06

Pacheco, T.R., Gomes, A.Q., Barbosa-Morais, N.L., Benes, V., Ansorge, W., Wollerton, M., Smith, C.W., Valcárcel, J., Carmo-Fonseca, M., 2004. Diversity of Vertebrate Splicing Factor U2AF35: IDENTIFICATION OF ALTERNATIVELY SPLICED U2AF1 mRNAs. Journal of Biological Chemistry 279, 27039–27049. https://doi.org/10.1074/jbc.M402136200

Pacheco, Teresa Raquel, Moita, L.F., Gomes, A.Q., Hacohen, N., Carmo-Fonseca, M., 2006. RNA Interference Knockdown of hU2AF35 Impairs Cell Cycle Progression and Modulates Alternative Splicing of Cdc25 Transcripts. MBoC 17, 4187–4199. https://doi.org/10.1091/mbc.e06-01-0036

Pachter, L., 2011. Models for transcript quantification from RNA-Seq. Genomics Methodol.

Palangat, M., Anastasakis, D.G., Fei, D.L., Lindblad, K.E., Bradley, R., Hourigan, C.S., Hafner, M., Larson, D.R., 2019. The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. Genes & Development 33, 482–497. https://doi.org/10.1101/gad.319590.118

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature Genetics 40, 1413. https://doi.org/10.1038/ng.259

Park, S.M., Ou, J., Chamberlain, L., Simone, T.M., Yang, H., Virbasius, C.-M., Ali, A.M., Zhu, L.J., Mukherjee, S., Raza, A., Green, M.R., 2016. U2AF35(S34F) Promotes Transformation by Directing Aberrant ATG7 Pre-mRNA 3′ End Formation. Molecular Cell 62, 479–490. https://doi.org/10.1016/j.molcel.2016.04.011

Parkin, S.E., Baer, M., Copeland, T.D., Schwartz, R.C., Johnson, P.F., 2002. Regulation of CCAAT/Enhancer-binding Protein (C/EBP) Activator Proteins by Heterodimerization with C/EBPγ (Ig/EBP). Journal of Biological Chemistry 277, 23563–23572. https://doi.org/10.1074/jbc.M202184200

Passacantilli, I., Frisone, P., De Paola, E., Fidaleo, M., Paronetto, M.P., 2017. hnRNPM guides an alternative splicing program in response to inhibition of the PI3K/AKT/mTOR pathway in Ewing sarcoma cells. Nucleic Acids Res 45, 12270–12284. https://doi.org/10.1093/nar/gkx831

Pohl, M., Bortfeldt, R.H., Grützmann, K., Schuster, S., 2013. Alternative splicing of mutually exclusive exons—A review. Biosystems 114, 31–38. https://doi.org/10.1016/j.biosystems.2013.07.003

Proudfoot, N.J., 2011. Ending the message: poly(A) signals then and now. Genes Dev 25, 1770–1782. https://doi.org/10.1101/gad.17268411

Reyes, A., Huber, W., 2018. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res 46, 582–592. https://doi.org/10.1093/nar/gkx1165

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics 13, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Richter, J.D., Sonenberg, N., 2005. Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature 433, 477–480. https://doi.org/10.1038/nature03205
Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., Wall, R., 1980. Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell 20, 303–312. https://doi.org/10.1016/0092-8674(80)90616-9
Rosenstiel, P., Huse, K., Franke, A., Hampe, J., Reichwald, K., Platzer, C., Roberts, R.G., Mathew, C.G., Platzer, M., Schreiber, S., 2007. Functional characterization of two novel 5' untranslated exons reveals a complex regulation of NOD2 protein expression. BMC Genomics 8, 472. https://doi.org/10.1186/1471-2164-8-472
Routh, A., Ji, P., Jaworski, E., Xia, Z., Li, W., Wagner, E.J., 2017. Poly(A)-ClickSeq: click-chemistry for next-generation 3′-end sequencing without RNA enrichment or fragmentation. Nucleic Acids Research 45, e112–e112. https://doi.org/10.1093/nar/gkx286
Samali, A., Fitzgerald, U., Deegan, S., Gupta, S., 2010. Methods for Monitoring Endoplasmic Reticulum Stress and the Unfolded Protein Response. International journal of cell biology 2010, 830307. https://doi.org/10.1155/2010/830307
Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., Burge, C.B., 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 320, 1643–1647. https://doi.org/10.1126/science.1155390
Sanfilippo, P., Miura, P., Lai, E.C., 2017. Genome-wide profiling of the 3' ends of polyadenylated RNAs. Methods 126, 86–94. https://doi.org/10.1016/j.ymeth.2017.06.003
Saxton, R.A., Sabatini, D.M., 2017. mTOR Signaling in Growth, Metabolism, and Disease. Cell 168, 960–976. https://doi.org/10.1016/j.cell.2017.02.004
Setzer, D.R., McGrogan, M., Nunberg, J.H., Schimke, R.T., 1980. Size heterogeneity in the 3′ end of dihydrofolate reductase messenger RNAs in mouse cells. Cell 22, 361–370. https://doi.org/10.1016/0092-8674(80)90346-3
Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H., Chen, G., Sun, H., Zhang, Y., Denise, A., Zhang, D.-E., Fu, X.-D., 2014. Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. Nature Structural & Molecular Biology 21, 997–1005. https://doi.org/10.1038/nsmb.2906
Sharon, D., Tilgner, H., Grubert, F., Snyder, M., 2013. A single-molecule long-read survey of the human transcriptome. Nature Biotechnology 31, 1009–1014. https://doi.org/10.1038/nbt.2705
Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proceedings of the National Academy of Sciences 111, E5593. https://doi.org/10.1073/pnas.1419161111
Shenker, S., Miura, P., Sanfilippo, P., Lai, E.C., 2015. IsoSCM: improved and alternative 3′ UTR annotation using multiple change-point inference. RNA 21, 14–27. https://doi.org/10.1261/rna.046037.114

Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., Shi, Y., 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17, 761–772. https://doi.org/10.1261/rna.2581711
Shi, Y., 2017. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nature Reviews Molecular Cell Biology 18, 655–670. https://doi.org/10.1038/nrm.2017.86
Shi, Y., 2012. Alternative polyadenylation: New insights from global analyses. RNA 18, 2105–2117. https://doi.org/10.1261/rna.035899.112
Shimobayashi, M., Hall, M.N., 2014. Making new contacts: the mTOR network in metabolism and signalling crosstalk. Nat. Rev. Mol. Cell Biol. 15. https://doi.org/10.1038/nrm3757
Shirai, C.L., Ley, J.N., White, B.S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E.J., Okeyo-Owuor, T., Liu, T., Griffith, M., McGrath, S., Magrini, V., Fulton, R.S., Fronick, C., O'Laughlin, M., Graubert, T.A., Walter, M.J., 2015. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell 27, 631–643. https://doi.org/10.1016/j.ccell.2015.04.008
Shirai, C.L., White, B.S., Tripathi, M., Tapia, R., Ley, J.N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B., Fulton, R.S., Fronick, C., O'Laughlin, M., Lagisetti, C., Webb, T.R., Graubert, T.A., Walter, M.J., 2017a. Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nat Commun 8, 14060–14060. https://doi.org/10.1038/ncomms14060
Shirai, C.L., White, B.S., Tripathi, M., Tapia, R., Ley, J.N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B., Fulton, R.S., Fronick, C., O'Laughlin, M., Lagisetti, C., Webb, T.R., Graubert, T.A., Walter, M.J., 2017b. Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nature Communications 8, 14060. https://doi.org/10.1038/ncomms14060
Shirakihara, T., Horiguchi, K., Miyazawa, K., Ehata, S., Shibata, T., Morita, I., Miyazono, K., Saitoh, M., 2011. TGF-β regulates isoform switching of FGF receptors and epithelial–mesenchymal transition. The EMBO Journal 30, 783–795. https://doi.org/10.1038/emboj.2010.351
Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., Chang, H.Y., Barna, M., 2017. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. Cell 169, 1051-1065.e18. https://doi.org/10.1016/j.cell.2017.05.022
Singh, N., Sahu, D.K., Chowdhry, R., Mishra, A., Goel, M.M., Faheem, M., Srivastava, C., Ojha, B.K., Gupta, D.K., Kant, R., 2016. IsoSeq analysis and functional annotation of the infratentorial ependymoma tumor tissue on PacBio RSII platform. Meta Gene 7, 70–75. https://doi.org/10.1016/j.mgene.2015.11.004
Singh, P., 2009. Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. Cancer Res. 69. https://doi.org/10.1158/0008-5472.CAN-09-2236
Singh, R., Valcarcel, J., Green, M., 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science 268, 1173. https://doi.org/10.1126/science.7761834

Steijger, T., Abril, J.F., Engström, P.G., Kokocinski, F., Abril, J.F., Akerman, M., Alioto, T., Ambrosini, G., Antonarakis, S.E., Behr, J., Bertone, P., Bohnert, R., Bucher, P., Cloonan, N., Derrien, T., Djebali, S., Du, J., Dudoit, S., Engström, P.G., Gerstein, M., Gingeras, T.R., Gonzalez, D., Grimmond, S.M., Guigó, R., Habegger, L., Harrow, J., Hubbard, T.J., Iseli, C., Jean, G., Kahles, A., Kokocinski, F., Lagarde, J., Leng, J., Lefebvre, G., Lewis, S., Mortazavi, A., Niermann, P., Rätsch, G., Reymond, A., Ribeca, P., Richard, H., Rougemont, J., Rozowsky, J., Sammeth, M., Sboner, A., Schulz, M.H., Searle, S.M.J., Solorzano, N.D., Solovyev, V., Stanke, M., Steijger, T., Stevenson, B.J., Stockinger, H., Valsesia, A., Weese, D., White, S., Wold, B.J., Wu, J., Wu, T.D., Zeller, G., Zerbino, D., Zhang, M.Q., Hubbard, T.J., Guigó, R., Harrow, J., Bertone, P., The RGASP Consortium, 2013. Assessment of transcript reconstruction methods for RNA-seq. Nature Methods 10, 1177–1184. https://doi.org/10.1038/nmeth.2714

Suzuki, K., Bose, P., Leong-Quong, R.Y., Fujita, D.J., Riabowol, K., 2010. REAP: A two minute cell fractionation method. BMC Research Notes 3, 294. https://doi.org/10.1186/1756-0500-3-294

Takagaki, Y., Seipelt, R.L., Peterson, M.L., Manley, J.L., 1996. The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation. Cell 87, 941–952. https://doi.org/10.1016/S0092-8674(00)82000-0

Tavanez, J.P., Madl, T., Kooshapur, H., Sattler, M., Valcárcel, J., 2012. hnRNP A1 Proofreads 3′ Splice Site Recognition by U2AF. Molecular Cell 45, 314–329. https://doi.org/10.1016/j.molcel.2011.11.033

Tee, A.R., Fingar, D.C., Manning, B.D., Kwiatkowski, D.J., Cantley, L.C., Blenis, J., 2002. Tuberous sclerosis complex-1 and -2 gene products function together to inhibit mammalian target of rapamycin (mTOR)-mediated downstream signaling. Proc Natl Acad Sci USA 99, 13571. https://doi.org/10.1073/pnas.202476899

Tee, A.R., Manning, B.D., Roux, P.P., Cantley, L.C., Blenis, J., 2003. Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. Curr. Biol. 13. https://doi.org/10.1016/S0960-9822(03)00506-2

Teixeira, L.K., Reed, S.I., 2013. Ubiquitin ligases and cell cycle control. Annu. Rev. Biochem. 82. https://doi.org/10.1146/annurev-biochem-060410-105307

Thoreen, C.C., 2009. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. J. Biol. Chem. 284. https://doi.org/10.1074/jbc.M900301200

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., Sabatini, D.M., 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. Nature 485, 109. https://doi.org/10.1038/nature11083

Tian, B., Hu, J., Zhang, H., Lutz, C.S., 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Research 33, 201–212. https://doi.org/10.1093/nar/gki158

Tian, B., Manley, J.L., 2013. Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci 38, 312–320. https://doi.org/10.1016/j.tibs.2013.03.005
Tokunaga, F., Brostrom, C., Koide, T., Arvan, P., 2000. Endoplasmic Reticulum (ER)-associated Degradation of Misfolded N-Linked Glycoproteins Is Suppressed upon Inhibition of ER Mannosidase I. Journal of Biological Chemistry 275, 40757–40764. https://doi.org/10.1074/jbc.M001073200
Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25. https://doi.org/10.1093/bioinformatics/btp120
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28, 511–515. https://doi.org/10.1038/nbt.1621
Valcárcel, J., Gaur, R.K., Singh, R., Green, M.R., 1996. Interaction of U2AF^{65} RS Region with Pre-mRNA of Branch Point and Promotion Base Pairing with U2 snRNA. Science 273, 1706. https://doi.org/10.1126/science.273.5282.1706
Van Etten, J., Schagat, T.L., Goldstrohm, A.C., 2013. A guide to design and optimization of reporter assays for 3′ untranslated region mediated regulation of mammalian messenger RNAs. Methods 63, 110–118. https://doi.org/10.1016/j.ymeth.2013.04.020
Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13, 227–232. https://doi.org/10.1038/nrg3185
Voith von Voithenberg, L., Sánchez-Rico, C., Kang, H.-S., Madl, T., Zanier, K., Barth, A., Warner, L.R., Sattler, M., Lamb, D.C., 2016. Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift. Proc Natl Acad Sci USA 113, E7169. https://doi.org/10.1073/pnas.1605873113
Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. https://doi.org/10.1038/nature07509
Wang, F., Song, W., Brancati, G., Segatori, L., 2011. Inhibition of Endoplasmic Reticulum-associated Degradation Rescues Native Folding in Loss of Function Protein Misfolding Diseases. Journal of Biological Chemistry 286, 43454–43464. https://doi.org/10.1074/jbc.M111.274332
Wang, G., Guo, X., Floros, J., 2005. Differences in the translation efficiency and mRNA stability mediated by 5′-UTR splice variants of human SP-A1 and SP-A2 genes. American Journal of Physiology-Lung Cellular and Molecular Physiology 289, L497–L508. https://doi.org/10.1152/ajplung.00100.2005
Wang, W., Wei, Z., Li, H., 2014. A change-point model for identifying 3′UTR switching by next-generation RNA sequencing. Bioinformatics 30, 2162–2170. https://doi.org/10.1093/bioinformatics/btu189
Wang, Y., 2011. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics 11. https://doi.org/10.1002/pmic.201000722

Wang, Y., Liu, J., Huang, B.O., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M., Wang, X.-Z., 2015. Mechanism of alternative splicing and its regulation. Biomed Rep 3, 152–158. https://doi.org/10.3892/br.2014.407
Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B., Carstens, R.P., 2009. ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. Molecular Cell 33, 591–601. https://doi.org/10.1016/j.molcel.2009.01.025
Weirather, J.L., Afshar, P.T., Clark, T.A., Tseng, E., Powers, L.S., Underwood, J.G., Zabner, J., Korlach, J., Wong, W.H., Au, K.F., 2015. Characterization of fusion genes and the significantly expressed fusion isoforms in breast cancer by hybrid sequencing. Nucleic Acids Research 43, e116–e116. https://doi.org/10.1093/nar/gkv562
West, S., Gromak, N., Proudfoot, N.J., 2004. Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432, 522–525. https://doi.org/10.1038/nature03035
Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nat. Methods 6. https://doi.org/10.1038/nmeth.1322
Wu, S., Romfo, C.M., Nilsen, T.W., Green, M.R., 1999. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature 402, 832–835. https://doi.org/10.1038/45590
Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., Li, W., 2014. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. Nature Communications 5, 5274. https://doi.org/10.1038/ncomms6274
Xie, J., Wang, X., Proud, C.G., 2016. mTOR inhibitors in cancer therapy. F1000Res 5, F1000 Faculty Rev-2078. https://doi.org/10.12688/f1000research.9207.1
Xing, Y., Yu, T., Wu, Y.N., Roy, M., Kim, J., Lee, C., 2006. An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. Nucleic Acids Research 34, 3150–3160. https://doi.org/10.1093/nar/gkl396
Yan, J., Marr, T.G., 2005. Computational analysis of 3′-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. Genome Research 15, 369–375. https://doi.org/10.1101/gr.3109605
Yanagiya, A., 2012. Translational homeostasis via the mRNA cap-binding protein, eIF4E. Mol. Cell 46. https://doi.org/10.1016/j.molcel.2012.04.004
Yang, Q., Gilmartin, G.M., Doublié, S., 2010. Structural basis of UGUA recognition by the Nudix protein CFI_m25 and implications for a regulatory role in mRNA 3′ processing. Proceedings of the National Academy of Sciences 107, 10062. https://doi.org/10.1073/pnas.1000848107
Yeh, H.-S., Chang, J.-W., Yong, J., 2016. Ribo-Proteomics Approach to Profile RNA–Protein and Protein–Protein Interaction Networks, in: Lin, R.-J. (Ed.), RNA-Protein Complexes and Interactions: Methods and Protocols. Springer New York, New York, NY, pp. 165–174. https://doi.org/10.1007/978-1-4939-3591-8_14
Yip, B.H., Steeples, V., Repapi, E., Armstrong, R.N., Llorian, M., Roy, S., Shaw, J., Dolatshad, H., Taylor, S., Verma, A., Bartenstein, M., Vyas, P., Cross, N.C.P.,

Malcovati, L., Cazzola, M., Hellström-Lindberg, E., Ogawa, S., Smith, C.W.J., Pellagatti, A., Boultwood, J., 2017. The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. J Clin Invest 127, 2206–2221. https://doi.org/10.1172/JCI91363

Yong, J., Kasim, M., Bachorik, J.L., Wan, L., Dreyfuss, G., 2010. Gemin5 Delivers snRNA Precursors to the SMN Complex for snRNP Biogenesis. Molecular Cell 38, 551–562. https://doi.org/10.1016/j.molcel.2010.03.014

Yoshida, H., Park, S.-Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto, Y., Urano, T., Obayashi, E., 2015. A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit. Genes & Development 29, 1649–1660. https://doi.org/10.1101/gad.267104.115

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W.-K., Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H.P., Shih, L.-Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S., Ogawa, S., 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478, 64–69. https://doi.org/10.1038/nature10496

Zamore, P.D., Patton, J.G., Green, M.R., 1992. Cloning and domain structure of the mammalian splicing factor U2AF. Nature 355, 609–614. https://doi.org/10.1038/355609a0

Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, A.I., Green, M.R., 1992. Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. Proc Natl Acad Sci USA 89, 8769. https://doi.org/10.1073/pnas.89.18.8769

Zhang, W., Chang, J.-W., Lin, L., Minn, K., Wu, B., Chien, J., Yong, J., Zheng, H., Kuang, R., 2015. Network-Based Isoform Quantification with RNA-Seq Data for Cancer Transcriptome Analysis. PLoS Comput Biol 11, e1004465–e1004465. https://doi.org/10.1371/journal.pcbi.1004465

Zhang, X., Virtanen, A., Kleiman, F.E., 2010. To polyadenylate or to deadenylate: that is the question. Cell Cycle 9, 4437–4449. https://doi.org/10.4161/cc.9.22.13887

Zhang, Y., 2014. Coordinated regulation of protein synthesis and degradation by mTORC1. Nature 513. https://doi.org/10.1038/nature13492

Zheng, D., Liu, X., Tian, B., 2016. 3′READS+, a sensitive and accurate method for 3′ end sequencing of polyadenylated RNA. RNA 22, 1631–1639. https://doi.org/10.1261/rna.057075.116

Zheng, X., Boyer, L., Jin, M., Mertens, J., Kim, Y., Ma, L., Ma, L., Hamm, M., Gage, F.H., Hunter, T., 2016. Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation. eLife 5, e13374. https://doi.org/10.7554/eLife.13374

Zhou, L., 2013. The role of RING box protein 1 in mouse oocyte meiotic maturation. PLoS One 8. https://doi.org/10.1371/journal.pone.0068964

Zhou, Z., Fu, X.-D., 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma 122, 191–207. https://doi.org/10.1007/s00412-013-0407-z