HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2021 October 09.

Published in final edited form as:

Cell Rep. 2021 July 20; 36(3): 109407. doi:10.1016/j.celrep.2021.109407.

Alternative 3′ UTRs play a widespread role in translation-independent mRNA association with the endoplasmic reticulum

Larry C. Cheng[1,2,3,6], **Dinghai Zheng**[2,6], **Qiang Zhang**[4], **Aysegul Guvenek**[2,5], **Hong Cheng**[4], **Bin Tian**[1,2,3,5,7,*]

[1]Program in Gene Expression and Regulation and Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA 19104, USA

[2]Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School, Newark, NJ 07103, USA

[3]Graduate Program in Quantitative Biomedicine, School of Graduate Studies, Rutgers University, Piscataway, NJ 08854, USA

[4]State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai 200031, China

[5]Rutgers School of Graduate Studies, Newark, NJ 07103, USA

[6]These authors contributed equally

[7]Lead contact

SUMMARY

Transcripts encoding membrane and secreted proteins are known to associate with the endoplasmic reticulum (ER) through translation. Here, using cell fractionation, polysome profiling, and 3′ end sequencing, we show that transcripts differ substantially in translation-independent ER association (TiERA). Genes in certain functional groups, such as cell signaling, tend to have significantly higher TiERA potentials than others, suggesting the importance of ER association for their mRNA metabolism, such as localized translation. The TiERA potential of a transcript is determined largely by size, sequence content, and RNA structures. Alternative polyadenylation (APA) isoforms can have distinct TiERA potentials because of changes in transcript features. The widespread 3′ UTR lengthening in cell differentiation leads to greater transcript association with the ER, especially for genes that are capable of expressing very long

*Correspondence: btian@wistar.org.

AUTHOR CONTRIBUTIONS

L.C.C., D.Z., and B.T. conceived and designed the experiments and analyses. D.Z. and Q.Z. performed the experiments. L.C.C., D.Z., and A.G. analyzed the data. H.C. provided critical experimental materials. L.C.C. and B.T. wrote the paper.

DECLARATION OF INTERESTS

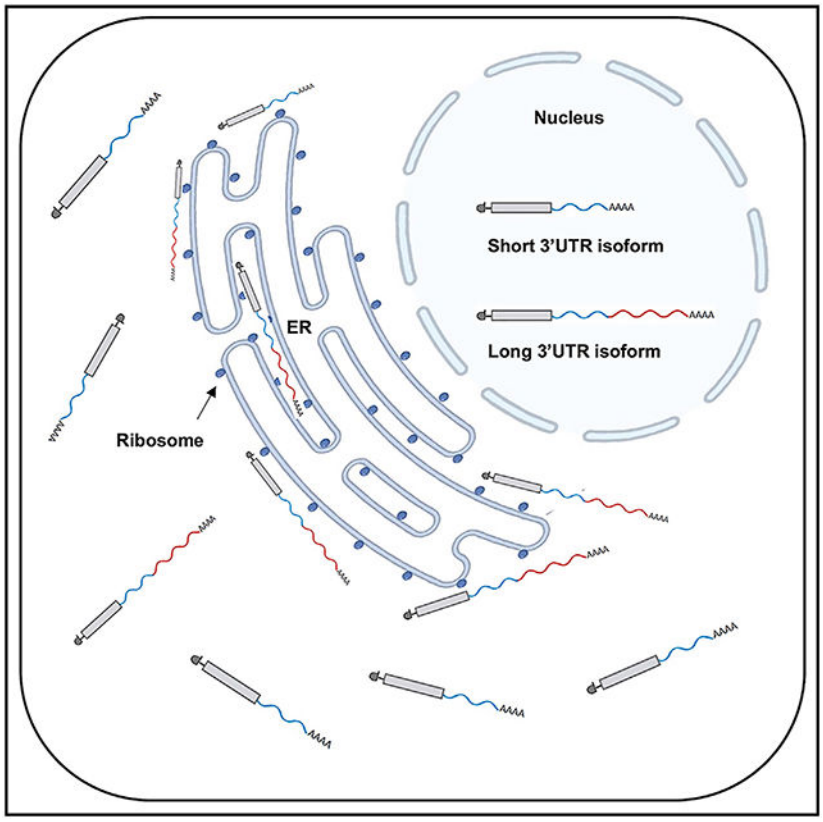
The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109407.

3′ UTRs. Our data also indicate that TiERA is in dynamic competition with translation-dependent ER association, suggesting limited space on the ER for mRNA association.

Graphical abstract



In brief

Cheng et al. show that transcripts differ substantially in translation-independent ER association (TiERA). Transcripts in certain functional groups, such as cell signaling, tend to have high TiERA potentials. Alternative polyadenylation changes transcript features that determine TiERA, leading to distinct ER association of mRNA isoforms.

INTRODUCTION

The 3′ UTR of mRNA plays an important role in post-transcriptional control of gene expression, such as stability, translation, and localization (Kejiou and Palazzo, 2017; Mayr, 2018). Much of its regulatory function is mediated through embedded sequence and structure motifs (Van Nostrand et al., 2020), such as microRNA (miRNA) target sites (Bartel, 2018) and various AU-rich and GU-rich elements (Garneau et al., 2007) for stability and/or translational controls. 3′ UTR size per se also appears to be a feature of mRNA metabolism, such as stability regulation through Upf1 binding (Hogg and Goff, 2010). In addition, the 3′ UTR has been found increasingly to play a role in mRNA localization, especially in polarized cells (Mayr, 2018; Tushev et al., 2018).

Most mammalian mRNA genes harbor multiple cleavage and polyadenylation sites (PASs), leading to alternative polyadenylation (APA) isoforms with different 3′ UTR sizes (Derti et al., 2012; Hoque et al., 2013). The APA isoform expression profile of a gene differs substantially across cell types (Lianoglou et al., 2013; Wang et al., 2008; Zhang et al., 2005) and is dynamically regulated under various conditions, such as cell proliferation (Sandberg et al., 2008), differentiation and development (Ji et al., 2009; Ji and Tian, 2009; Shepard et al., 2011), oncogenesis (Fu et al., 2011; Masamha and Wagner, 2018; Masamha et al., 2014; Mayr and Bartel, 2009; Morris et al., 2012; Singh et al., 2009), cell activation (Berg et al., 2012; Flavell et al., 2008), and cellular stress (Hollerer et al., 2016; Sadek et al., 2019; Zheng et al., 2018). Regulation of APA is an important layer of the gene expression program (Elkon et al., 2013; Gruber and Zavolan, 2019; Tian and Manley, 2017).

mRNAs encoding membrane and secreted proteins typically undergo translation on the endoplasmic reticulum (ER), leading to the distinct subcellular structure of rough ER (Voeltz et al., 2002). The signal recognition particle (SRP), which binds to the nascent polypeptide coming out from a translating ribosome, directs the mRNA to the cytosolic surface of the ER (Akopian et al., 2013). Therefore, ER association of these mRNAs is translation dependent. Similarly, mRNAs have also been found to be associated with other organelles, such as endosomes and mitochondria (Cioni et al., 2019a; Higuchi et al., 2014; Tsuboi et al., 2020; Williams et al., 2014), often together with the ribosome (Béthune et al., 2019). On the other hand, several studies have indicated that not all ER-associated transcripts encode membrane or secreted proteins (Reid and Nicchitta, 2012) and that mRNAs can associate with the ER independent of translation (Cui and Palazzo, 2014; Reid and Nicchitta, 2015). To what extent different transcript features, particularly the 3′ UTR, are involved in these organelle associations is largely unclear. Here, using subcellular fractionation, polysome profiling, and 3′ end sequencing, we examine transcripts associated with the ER in proliferating and differentiated mouse C2C12 cells. We identify transcript features that facilitate translation-independent ER association (TiERA). We analyze functions of encoded proteins associated with transcripts with high or low TiERA potentials. By comparing 3′ UTR isoforms genome wide, we examine the role of APA in regulation of TiERA in proliferating and differentiated cells.

RESULTS

APA isoforms differ in subcellular distribution

We were interested in understanding how APA isoforms are distributed in the cell. To this end, we used the sequential detergent wash protocol developed by Jagannathan et al. (2011) to fractionate proliferating C2C12 myoblast cells into cytosol, membrane, and insoluble parts (Figure 1A; see STAR Methods for details). Western blot analysis indicated enrichment of several marker proteins for the respective fractions (Figure 1B); i.e., the cytoskeleton protein α-tubulin for the cytosol fraction, the ER-resident protein GRP94 for the membrane fraction (with some presence in the insoluble fraction as well), and the histone protein H2A.X for the insoluble fraction.

We next extracted RNAs from these fractions and subjected them to RNA sequencing (RNA-seq) analysis (Figure 1A). As expected, differential gene expression analysis indicated

substantial transcriptome differences between the three fractions (Figure S1A). A much higher proportion of reads in the membrane fraction was aligned to the mitochondrial genome (6%) compared with reads in the other two fractions (<1%), confirming that the membrane fraction contained membranous organelles, such as mitochondria (Figure S1B). In contrast, a much higher proportion of intronic reads was found in the insoluble fraction (20%) compared with other fractions (4%), indicating enrichment of pre-mRNAs, likely chromatin bound, in this fraction (Figure S1B).

Intriguingly, higher proportions of reads in membrane and insoluble fractions were mapped to 3′ UTRs (40% and 34%, respectively; Figure S1B) compared with the cytosol fraction (28%), suggesting that mRNA isoforms with different 3′ UTR sizes might be differentially distributed in the three fractions. To explore this, we analyzed the RNA-seq data using APAlyzer, a computational program our lab developed recently to examine relative expression of 3′ UTR isoforms in different samples (Wang and Tian, 2020). As illustrated in Figure 1C (see STAR Methods for details), a relative expression (RE) value, representing the ratio of RNA-seq read density in the alternative 3′ UTR (aUTR) to that in the common 3′ UTR (cUTR), was calculated for each gene. aUTR and cUTR were demarcated by the first PAS in the 3′ UTR that was conserved in mammals, as annotated in the comprehensive PAS database PolyA_DB v.3 (Wang et al., 2018). Therefore, a larger RE value indicates a higher abundance of long 3′ UTR isoforms relative to that of short 3′ UTR isoforms. Interestingly, we found that RE values were significantly higher in membrane and insoluble fractions compared with the cytosol fraction ($p < 2.2 \times 10^{-16}$; Figure 1D). An example gene, *Nmt1*, which contains two PASs in its 3′ UTR, is shown in Figure 1E. Note the difference in read coverage in its 3′ UTR among the three fractions (Figure 1E, right). Also discernable are higher levels of intronic reads in the insoluble fraction, consistent with presence of pre-mRNAs (Figure 1E, left).

To corroborate the RNA-seq result, we repeated the same fractionation protocol and subjected RNAs to 3′ region extraction and deep sequencing (3′READS), a method our lab developed previously to specifically interrogate 3′ ends of the poly(A)$^+$ transcriptome (Hoque et al., 2013; Zheng et al., 2016; Figure 1F). Because the insoluble fraction contained a mixture of pre-mRNAs and mature mRNAs (RNA-seq data) and included some residual membrane fraction (western blot data), we focused on cytosol and membrane fractions only. Principal-component analysis supported the similarity in gene expression between RNA-seq data and 3′READS data (Figure S1C).

To measure relative distribution between membrane and cytosol fractions for any transcript with a defined PAS, we calculated a ratio, called the membrane localization score (MLS), based on expression of the transcript (using reads per million mapped [RPM]) in membrane versus cytosol fractions (Figure 1G). Using the top two most abundant 3′ UTR APA isoforms of each gene for comparison, we found that genes whose long 3′ UTR isoform had a higher MLS than its short 3′ UTR isoform ($\Delta MLS > \log_2(1.2)$, $p < 0.05$, Fisher's exact test) outnumbered those showing the opposite trend by 2.2-fold (548 versus 251; Figure 1H), indicating a global trend showing that long 3′ UTR isoforms were more enriched in the membrane fraction than short 3′ UTR isoforms.

To further examine how 3′ UTR size affects the MLS, we divided genes into five bins based on the size of aUTR, the sequence between short and long isoforms (Figure 1I). Based on median ΔMLSs (long isoform minus short isoform) of each gene bin, we found that the longer the aUTR, the higher the ΔMLS. For example, genes with an aUTR size of more than 1,634 nt (bin 5, top 20%) showed a much higher ΔMLS than genes with an aUTR size of less than 120 nt (bin 1, bottom 20%, $p = 1.6 \times 10^{-14}$, Wilcoxon test; Figure 1I). Therefore, aUTR size has a positive effect on the MLS, resulting in long 3′ UTR isoforms being more likely than short 3′ UTR isoforms in the membrane fraction. Gene Ontology (GO) analysis identified a few terms enriched for genes in bin 1 or bin 5 (Figure S1D) even though their p values were marginally significant, with the exceptions of the “ribosome” and “cytosolic part” terms that were enriched for bin 1 genes (Figure S1D).

Translating transcripts with long 3′ UTRs are enriched on the ER

Because the membrane fraction could contain RNAs from multiple organelles, including the ER, mitochondria, nucleoplasm, etc., we next wanted to focus on translating RNAs in the membrane fraction. We reasoned that transcripts undergoing translation in the membrane fraction should come mostly from the rough ER, the only known organelle that harbors substantial protein synthesis. To this end, we carried out polysome profiling using cytosol and membrane fractions from C2C12 myoblast cells (Figure 2A). Consistent with a previous report (Lerner and Nicchitta, 2006), our polysome profiles showed a higher poly-some-to-monosome content ratio in the membrane fraction compared with that in the cytosol fraction (Figure 2B), suggesting that translation is more active on the ER than in cytosol. We then extracted RNAs from monosome (M) and polysome (P) portions of each fraction (Figure 2B) and subjected them to 3′READS analysis. As with the total cell extract, we calculated an MLS for each transcript (represented by its PAS) in P and M samples.

Similar to the total cell extract result, P and M samples showed that long 3′ UTR isoforms tended to have a higher MLS than short 3′ UTR isoforms (Figure 2C). The numbers of genes showing a long 3′ UTR isoform > a short 3′UTR isoform were 3.1- and 2.4-fold greater than the numbers of genes showing a short 3′ UTR isoform > a long 3′ UTR isoform, respectively, in P and M samples (significant difference was based on ΔMLS > $\log_2(1.2)$ or < $-\log_2(1.2)$, $p < 0.05$, DEXseq; Figure 2C). A correlation between aUTR size and ΔMLS was readily discernable with M and P data (Figure 2D). Genes in bin 1 (the shortest aUTRs, bottom 20%) and those in bin 5 (the longest aUTRs, top 20%) had significantly different ΔMLSs in both fractions ($p < 2.2 \times 10^{-16}$ and $p = 3.9 \times 10^{-15}$ for P and M samples, respectively; Wilcoxon test; Figure 2D).

Based on the MLSs of all transcripts, we found modest correlations in P versus M ($r = 0.52$, Pearson correlation, Figure 2E, bottom left) and in M versus total cell extract ($r = 0.51$; Figure 2E, top right). In contrast, P and total cell extracts were much better correlated in their MLSs ($r = 0.84$; Figure 2E, bottom right). These results indicate that transcript MLSs in total cell extracts were attributable chiefly to P-associated RNAs. Therefore, the MLS mainly reflects transcript distribution between the rough ER, where membrane-associated P reside, and the cytosol. Taken together, polysome profiling data confirm that differential ER

association between 3′ UTR isoforms are responsible for the difference in distribution in membrane versus cytosol fractions.

Transcriptome-wide identification of TiERA

We next wanted to find out how 3′ UTR-mediated ER association was related to translation. To this end, we treated C2C12 myoblast cells with puromycin, an aminoacyl transfer RNA analog that inhibits translation through ejection of nascent polypeptide from the ribosome (Nathans, 1964), and extracted RNAs from cytosolic and membrane fractions at different time points for 3′READS analysis (Figure 3A). Based on transcript MLSs, the data for 30-min and 60-min puromycin treatments were well correlated ($r = 0.70$, Pearson correlation; Figure S2A), and so were data for non-puromycin and 0-min puromycin treatment ($r = 0.83$, Pearson correlation; Figure S2A). In contrast, a much lower correlation was observed between no puromycin treatment or 0-min treatment samples and 30-min or 60-min treatment samples ($r = 0.31–0.35$, Pearson correlation; Figure S2A). We therefore analyzed the former two as replicates for no puromycin (Puro−) and the latter two as replicates for Puro treatment (Puro+).

Transcriptome-wide comparison of MLSs in Puro+ versus Puro− samples appeared to indicate two transcript populations (Figure 3B). We thus employed a Gaussian mixture model to separate these two based on MLSs in Puro+ and Puro− samples (Figure 3C; see STAR Methods for details). Transcripts in the first population (red dots in Figure 3C), called group I, showed a good MLS correlation between Puro+ and Puro− samples, indicating that their ER association is translatio independent. In fact, their MLS in Puro+ were slightly higher than those in Puro− samples (median = −0.04 and −0.31, respectively; Figures 3C; $p < 2.2 \times 10^{-16}$, K-S test; Figure S2B), indicating that their ER association was enhanced by translational inhibition. In contrast, transcripts in the second population, called group II (blue dots in Figure 3C), showed significantly lower MLSs in Puro+ samples than in Puro− samples (median = −0.03 and 3.54, respectively; Figure 3C; $p < 2.2 \times 10^{-16}$, K-S test; Figure S2B), indicating that their ER association depend on translation. Notably, groups I and II transcripts showed similar MLSs in Puro+ samples (median = −0.04 and −0.03, respectively; Figure 3C), indicating that ER association is comparable between the two groups when translation is inhibited.

We found that 53% and 10% of the group II transcripts encoded membrane proteins and secreted proteins, respectively, based on the secretome and subcellular proteome knowledgebase MetazSecKB (Meinken et al., 2015). In contrast, 10% and 1% of the group I transcripts fell into these two categories, respectively (Figure 3D). Transcripts encoding mitochondrial proteins were depleted from group II (2%; Figure 3D) compared with group I (8%; Figure 3D) or all transcripts (7%; Figure S2C). In addition, GO analysis of these two groups showed that group I transcripts tended to encode proteins localized to the nucleus, microtubules, mitochondria, and cytosol (Table S1), whereas proteins expressed from group II transcripts were more likely to be located at the cell surface, ER membrane, receptor complex, and other membrane structures (Table S2). Thus, the GO result is in good agreement with the result based on MetazSecKB, further supporting the view that group II is enriched with transcripts encoding membrane and secreted proteins. These results

indicate that our Gaussian mixture model using MLS data from Puro− and Puro+ samples effectively separated transcripts that are associated with the ER through translation (group II transcripts), which tend to encode membrane and secreted proteins, and transcripts that are associated with the ER independent of translation (group I transcripts). For simplicity, we call the latter TiERA (Translation-independent ER Association).

3′ UTR isoform analysis corroborates TiERA

We found that genes whose long isoform > short isoform in ER association (ΔMLS > $\log_2(1.2)$, $p < 0.05$, DEXseq) outnumbered those whose long isoform < short isoform in ER association (ΔMLS < $-\log_2(1.2)$, $p < 0.05$, DEXseq) by 2.7-fold in Puro− samples (Figure 3E, left). Interestingly, this bias increased to 7.1-fold in Puro+ samples (Figure 3E, right), indicating that inhibition of translation accentuates 3′ UTR-mediated TiERA, in line with whole-transcriptome analysis result (above).

To further explore the effect of the 3′ UTR on TiERA, we analyzed group I and group II gene transcripts separately in Puro+ and Puro− samples and divided genes in each group into 5 bins based on aUTR size (as in Figure 1I). We found that ΔMLSs of group I genes appeared to be greater in Puro+ samples than in Puro− samples across all aUTR size bins (Figure 3F), confirming that translation inhibition enhances TiERA. For genes with the longest aUTRs (bin 5, top 20%), ΔMLSs between 3′ UTR isoforms was significant higher in Puro+ versus Puro− samples (0.83 versus 0.53, $p = 2.7 \times 10^{-11}$, Wilcoxon test; Figure 3G). In contrast, although ΔMLSs of group II genes followed a similar aUTR size-based increase in Puro+ samples (dashed blue line, Figure 3F), their ΔMLS in Puro− samples did not correlate with aUTR size (solid blue line, Figure 3F). For group II genes in bin 5, ΔMLS was significant only in Puro+ samples ($p < 0.001$ compared with 0, Wilcoxon test; Figure 3G) but not in Puro− samples ($p > 0.05$ compared with 0, Wilcoxon test; Figure 3G). Therefore, 3′ UTR size has an influence on MLSs of group II transcripts only when translation is inhibited. Because there is no significant difference in ΔMLS between group I and group II genes in Puro+ samples (first two bars, $p = 0.58$, Wilcoxon test; Figure 3G), TiERA is equally effective on transcripts from both groups. Together, our results comparing 3′ UTR isoforms corroborate the analysis based on the whole transcriptome and indicate that 3′ UTR-mediated ER association is translation independent.

Transcripts with different TiERA potentials belong to distinct functional groups

We next wanted to know what kinds of transcripts tend to have high or low TiERA potentials. Using GO analysis, we compared top and bottom transcripts based on their MLS in Puro+ samples. For group I genes, top biological process (BP) terms associated with transcripts having high TiERA potentials tended to be related to cell signaling (Table 1), such as “regulation of Rho protein signal transduction” and “regulation of MAPK cascade,” whereas those associated with transcripts having low TiERA potentials tended to be related to metabolism, such as “nucleobase-containing small molecule metabolic process” and “amide biosynthetic process.” Also notable is that several cellular component (CC) terms were found to be significantly associated with group I gene transcripts with low TiERA potentials (Table S4), including “mitochondrial envelope” and “ribosomal subunit.”

Interestingly, signaling-related terms were also found to be significantly associated with group II gene transcripts having high TiERA potentials (Table 1), such as “enzyme linked receptor protein signaling pathway” and “semaphoring-plexin signaling pathway.” Consistently, the top CC term enriched for these transcripts was “receptor complex” (Table S4). In contrast, transcripts having low TiERA potentials in group II genes were enriched with BP terms related to protein localization to organelles (Table 1), such as “protein localization to endoplasmic reticulum” and “ER to Golgi vesicle-mediated transport,” and with CC terms related to organelle structures (Table S4), such as “endoplasmic reticulum-Golgi intermediate compartment” and “coasted vesicle.” GO data thus indicate that some functional gene groups have distinct TiERA potentials. Transcripts encoding proteins involved in cell signaling tend to have high TiERA potentials, whereas those encoding subcellular compartment-related proteins tend to have low TiERA potentials.

Transcript features determine TiERA

We next examined how transcript features might contribute to TiERA. Because our APA isoform analyses indicated that 3′ UTR size enhances TiERA, we first examined size-related features (Figure 4A). Strikingly, the most prominent feature, based on correlation with TiERA potentials (MLS in Puro+ samples), was overall transcript size (R^2 = 0.42; Figure 4A). Consistently, CDS, 3′ UTR, and 5′ UTR sizes had positive contributions to TiERA (R^2 = 0.34, 0.14, and 0.08, respectively; Figure 4A). Intriguingly, although exon number correlated positively with TiERA, likely because of its confounding association with transcript size, the exon-exon junction (EJC) density had a negative effect on TiERA (R^2 = 0.16; Figure 4A). Size-related features altogether explained 50% of TiERA variance (R^2 = 0.50; Figure 4A).

We then examined how sequence motifs are related to TiERA. Using dimer frequencies, we found that A/U-rich dimers, such as AA, AU, UA, and UU, had negative correlations with TiERA (Figure 4B), whereas dimers containing G or C, such as CC, CU, CA, and GG, were correlated positively with TiERA (Figure 4B). Dimer frequencies altogether had an R^2 value of 0.15 (Figure 4A). Thus, although still prominent, sequence content contributed less to TiERA than size features.

Interestingly, when we used dimer counts for analysis, which effectively combined dimer frequency and transcript size features, we found the top three dimers to be GG, GC, and CC in all regions, including the whole transcript, CDS, and 3′ UTR (Figures 4C and S3A). The difference between dimer frequency and dimer counts suggests a synergistic effect between size and sequence content. We reasoned that this could happen when certain sequence features become significant only when they reach a critical length, such as in formation of RNA structures. To explore this, we carried out RNA structure prediction in CDS and 3′ UTR using RNAfold (Lorenz et al., 2011). Using a 100-nt moving window strategy to obtain the average minimum free energy (MFE) of predicted RNA structures (see STAR Methods for details), we found that, indeed, RNA structures in the 3′ UTR and CDS had positive contributions to TiERA, albeit with low R^2 values (R^2 = 0.06 and 0.04, respectively, and R^2 = 0.07 for two regions combined; Figure 4D). Consistently, transcripts with high TiERA potentials (top 20%) tended to have a low MFE in the CDS and 3′ UTR, and those with

low TiERA potentials (bottom 20%) tended to have a high MFE (Figure S3B). In addition, using genes whose 3′ UTR isoforms had significantly different TiERA potentials, we found that aUTRs tended to have lower MFEs when long isoforms had higher MLSs than short isoforms and vice versa (Figure S3C), further supporting the importance of RNA structure to TiERA. We also examined the relationship between TiERA potentials and RNA structures using RNA structures that were identified experimentally by the icSHAPE (in vivo click selective 2-hydroxyl acylation and profiling experiment) method (Figure 4E; see STAR Methods for details; Sun et al., 2019). In line with our RNA structure prediction results, transcripts with high MLSs tended to adopt more RNA structures (high icSHAPE Gini indices; Figure 4E) and transcripts with low MLS tended to contain fewer RNA structures (low icSHAPE Gini indices; Figure 4E) in CDS and 3′ UTR sequences.

We next employed a machine learning method, XGBoost (Chen and Guestrin, 2016), to construct a predictive model for TiERA potentials, using all aforementioned features, including size features, dimer frequencies, and predicted MFE for RNA structures (Figure 4F). Based on correlation of observed MLS (Puro+ samples) and predicted MLS, the model performed well for group I and group II gene transcripts ($r = 0.86$ and 0.81, respectively, Pearson correlation coefficients; Figures 4G and 4H). The good performance of the predictive model for group II transcripts further supports the notion that these transcripts follow the same rules as group I transcripts in TiERA, despite normally being associated with the ER in a translation-dependent manner.

We next tested our predictive model on 3′ UTR isoforms. Overall, the predicted ΔMLSs (long isoform versus short isoform) were correlated with observed values ($r = 0.63$; Figure 4I). The lower correlation coefficient for ΔMLS compared with that for MLS is expected because the former had more data points for calculation and, thus, contained high noise levels. Importantly, when APA genes were divided into five bins based on aUTR size, the correlation between predicted and observed ΔMLS in group I genes was much higher for genes with long aUTRs ($r = 0.64$, bin 5) compared with genes with short aUTRs ($r = 0.35$, bin 1) ($p = 8.1 \times 10^{-25}$, t test; Figure 4J). Our results, therefore, indicate that the TiERA potential of a transcript is its intrinsic property that is governed by size, sequence content, and RNA structures.

Validation of TiERA difference between APA isoforms

We next set out to validate our global findings with detailed analyses of individual genes. The gene *Nmt1* encodes N-myristoyl-transferase 1, which catalyzes myristoylation, a type of lipid modification, of its substrate proteins (Meinnel et al., 2020; Udenwobele et al., 2017). *Nmt1* produces two 3′ UTR isoforms that differ substantially in 3′ UTR size, with the short and long isoforms having 337 nt and 3,144 nt in 3′ UTRs, respectively (Figure 5A). They also differ greatly in MLS (−0.90 and 1.21 for short and long isoforms, respectively, Puro+ samples; Figure 5A). Based on our Gaussian mixture model (Figure 3B), *Nmt1* is a group 1 gene. Interestingly, the long 3′ UTR isoform has many more GG dimers than the short 3′ UTR isoform (347 versus 120), the top dimer in 3′ UTRs for contribution to TiERA based on dimer count (Figure S3A).

Using northern blot analysis with a probe targeting the common region of the two isoforms, we examined the relative abundance of the two isoforms in cytosol and membrane fractions in C2C12 myoblasts (Figures 5B and S4A). In cells without Puro, the relative transcript abundance between the long isoform and short isoform was 3.4-fold higher in the membrane fraction compared with the cytosol (Figures 5B and 5C), supporting the notion that the long isoform is better associated with the ER than the short isoform. The relative abundance increased to 7.2-fold in cells treated with Puro (Figures 5B and 5C), corroborating the notion that translational inhibition enhances TiERA.

To further confirm the contribution of *Nmt1*'s 3′ UTR to TiERA, we cloned the full 3′ UTR of the long isoform into the psiCHECK2 vector, placing it after *Renilla* luciferase CDS (Figure 5D). Northern blot analysis of C2C12 myoblasts transfected with the reporter showed two expected isoforms (Figure 5E). Importantly, the long isoform to short isoform abundance ratio was 9-fold higher in the membrane fraction compared with that in the cytosol. This ratio difference increased to 61-fold when cells were treated with Puro (Figures 5E and 5F). A similar result was obtained when cells were treated with harringtonine, an translational inhibitor that functions at the initiation step (Huang, 1975), indicating that our observations based on Puro treatment are primarily due to translational inhibition as opposed to Puro-induced proteostatic stress (Aviner, 2020; Salomons et al., 2009; Figures S4B and S4C). Our validation data on *Nmt1* isoforms are in good agreement with the conclusions from global analyses.

Distinct TiERA potentials in proliferating and differentiated cells

Proliferating C2C12 myoblasts can differentiate into myotubes, during which APA profiles change substantially (Ji et al., 2009; Li et al., 2015). We next wanted to find out whether the TiERA potential of a transcript might change during cell differentiation. Using the same procedures for cell fractionation, polysome profiling, and 3′READS, we calculated MLSs in P and M portions of membrane and cytosol fractions of differentiated C2C12 myotube cells (Figure S5A). Overall, group I and II gene transcripts showed similar MLS correlation levels between myoblasts and myotubes in P samples (Pearson correlation $r = 0.67$ and 0.71, respectively; Figure 6A). Slightly more modest correlations were found in M samples (Pearson correlation $r = 0.62$ and 0.64, respectively; Figure 6A). As expected, the MLSs of group II genes were generally higher than those of group I genes in P samples, and this difference was much subdued in the M sample (compare blue and red dot distributions in the P and M scatterplots; Figure 6A). This result is consistent with the notion that translation-dependent ER association for group II gene transcripts are more established after initiation of translation or the monosome stage.

We next examined how APA isoforms differed in ER association in myoblasts versus myotubes. To this end, we analyzed ΔMLS between long and short 3′ UTR isoforms across genes (group I only) with different aUTR sizes (Figure 6B). Interestingly, although ΔMLS increased along with aUTR size changes in myoblasts and myotubes (Figure 6B), this trend was more pronounced in myoblasts than in myotubes (Figure 6B). Focusing on genes with a long aUTR (top 20%, bin 5), we found that ΔMLSs between long and short 3′ UTR isoforms were significantly greater in myoblasts than in myotubes in P ($p = 2.0$

× 10^{-5}, Wilcoxon test; Figure 6C) and M ($p = 8.8 \times 10^{-6}$, Wilcoxon test; Figure 6C) samples. Therefore, 3′ UTR-mediated TiERA is more potent in proliferating myoblasts than in differentiated myotubes. Interestingly, using our previously generated 3′READS data based on total RNAs from C2C12 myoblasts and myotubes (Wang et al., 2019), we found that group II genes were generally upregulated during C2C12 differentiation ($p = 5.3 \times 10^{-12}$; Figure 6D), whereas group I genes were downregulated slightly during differentiation ($p = 1.0 \times 10^{-4}$; Figure 6D). Therefore, the decreased TiERA of group I transcripts in differentiated cells could be attributable to increased translation-dependent ER association of group II gene transcripts.

Because 3′ UTRs generally lengthen during C2C12 cell differentiation (Ji et al., 2009; Li et al., 2015; Figure S5B), we wanted to find out how the 3′ UTR size change would affect ER association. Using 3′READS data for proliferating and differentiated C2C12 cells (Wang et al., 2019), we calculated RE difference (RED) scores (long isoform versus short isoform in myotubes versus myoblasts; see formula in Figure 6E) to quantitatively measure APA changes. We found that genes whose long 3′ UTR isoform had a higher TiERA potential than its short 3′ UTR isoform (based on Puro+ samples; Figure 3E) showed much higher RED scores, indicating greater 3′ UTR lengthening in cell differentiation than genes whose isoforms did not differ significantly in TiERA potentials (median RED = 0.33 versus 0.18, $p = 1.3 \times 10^{-3}$, K-S (Kolmogorov–Smirnov) test; Figure 6E). Notably, genes whose short isoform had a greater TiERA potential than the long isoform displayed lower RED scores than other two gene groups (median RED = 0.05) even though the difference was not statistically significant (p = 0.12 versus genes whose isoforms did not differ in TiERA, K-S test; Figure 6E). These results indicate that gene transcripts could increase ER association through 3′ UTR lengthening during cell differentiation.

DISCUSSION

In this study, we examine TiERA for mRNAs in mouse myoblast and myotube cell line models. We show that TiERA is widespread and that transcripts with high or low TiERA potentials are enriched for distinct biological functions. The TiERA potential is an intrinsic property of mRNA, governed by its size, sequence content, and RNA structures. Consistently, 3′ UTR APA isoforms, especially those with substantial differences in these features, differ in TiERA potentials. Therefore, APA in cell differentiation alters ER association of transcripts through 3′ UTR size changes, potentially affecting the localization and functions of their encoded proteins.

The ER is a large and complex subcellular network, having close contact with the nucleus (English and Voeltz, 2013) and other organelles. For group I transcripts, whose protein products are generally not secreted or membrane bound, a high TiERA potential would restrict free movement of the transcript in the cell, leading to localized regulation of mRNA translation and degradation. Our finding that transcripts with high TiERA potentials tend to have functions in signaling pathways may indicate that TiERA could be a mechanism by which proteins involved in signaling pathways are produced locally at certain sites and poised for signaling events. Conversely, transcripts with low TiERA potentials, enriched

for functions of metabolic processes and localization to mitochondria and other subcellular compartments, may be able to move freely in the cell by avoiding ER association.

A case in point is the *Nmt1* gene, whose protein product plays important roles in cellular signaling, protein-protein interaction, and membrane targeting (Meinnel et al., 2020; Udenwobele et al., 2017). Its two 3′ UTR isoforms differ substantially in 3′ UTR size. In addition, the relative size of 3′ UTR to the whole transcript is strikingly different for the two isoforms, accounting for 18% of the short isoform but 68% of the long isoform. Therefore, even though CDS size in general is also important for TiERA, the 3′ UTR of *Nmt1* appears to play a dominant role in the long isoform. The high TiERA potential of the *Nmt1* long 3′ UTR isoform could facilitate translation of encoded proteins to be in proximity to its action site. Regulation of the relative abundance of the two 3′ UTR isoforms of *Nmt1* could have a widespread effect on protein myristoylation and, thus, their localization and interactions with other proteins. In this vein, it is noteworthy that spatial distribution of myristoylated proteins changes drastically in myogenesis (Witten et al., 2017). The biological ramifications of APA of *Nmt1* in cell differentiation and other biological conditions needs to be explored further.

Our analysis also revealed that transcripts of group II genes, which encode secreted or membrane-bound proteins, can have substantial differences in TiERA as well. They are associated with the ER through the SRP-mediated, translation-dependent mechanism, which is more potent for ER association than TiERA, as shown in our analysis. Having high or low TiERA potentials for these transcripts may help them localize when they are not in translation. Interestingly, proteins encoded by group II gene transcripts with high TiERA potentials also tend to have roles in cell signaling. The interplay between TiERA and translation-dependent ER association for these transcripts would be an interesting subject for further analysis.

The bias in transcript distribution between membrane and cytosol has been reported previously using similar cell fractionation methods (Benoit Bouvrette et al., 2018; Jagannathan et al., 2011). However, additional analysis with polysome profiling allowed us to attribute membrane enrichment to the ER more definitely. On the other hand, we cannot rule out the possibility that some membrane association is through other organelles, such as endosomes or mitochondria (Cioni et al., 2019a; Higuchi et al., 2014; Tsuboi et al., 2020; Williams et al., 2014). Future studies using more sophisticated cell fractionation methods (Williamson et al., 2015) or microscopy techniques (Wu and Palazzo, 2020) could provide more precise data. Also notable is that 3′ UTR size in the insoluble fraction appears to be similar to that in the membrane fraction. The insoluble fraction contains a mixture of nuclear and cytoplasmic structures, such as chromatin and the cytoskeleton (Wang et al., 2012). It is therefore possible that long 3′ UTRs may have intrinsic properties to associate with various organelles and subcellular structures, facilitating localized protein production and mRNA decay. How different organelles and subcellular structures implement specificity for transcript association is an interesting question, especially with respect to how RNA-binding proteins are involved in transcript localization (Béthune et al., 2019; Cui et al., 2012; Ma and Mayr, 2018).

Our analyses also suggest that TiERA is in competition with SRP-mediated mechanisms for ER association. First, group I gene transcripts increased ER association when translation was inhibited. Second, 3′ UTR-mediated TiERA decreased during differentiation of C2C12 cells, which coincided with upregulation of group II genes. While this notion awaits more direct experimental validation, our current data appear to argue that the ER space for transcript association is generally limited. In this vein, we recently found that professional secretory cells, which have a high protein production activity on the ER, tend to express shorter 3′ UTRs in general compared with their precursor cells (Cheng et al., 2020). One possibility is that expression of shorter 3′ UTRs could avoid ER association, mitigating transcript crowding on the ER and, thus, avoiding ER stress and related mRNA degradation (Hollien et al., 2009; Reid et al., 2014). Conversely, some cells, such as neurons, tend to express long 3′ UTRs (Guvenek and Tian, 2018; Miura et al., 2013; Zhang et al., 2005). How the TiERA mechanism is employed in these cells for localized translation and degradation (Cioni et al., 2019b; Hyde et al., 2002) requires further investigation.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bin Tian (btian@wistar.org).

Materials availability—All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Data and code availability—Sequencing datasets generated in this study have been deposited into the GEO database with the accession number GSE162971. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

C2C12 myblast cells were cultured in DMEM with 10% fetal bovine serum. C2C12 cell differentiation to myotube cells was induced by culturing cells in DMEM with 2% horse serum when cell confluency reached > 95%. Culture media were supplemented with 100 I.U./mL penicillin and 100 μg/mL streptomycin.

METHOD DETAILS

Cell fractionation—We followed the fractionation by sequential detergent extraction protocol developed by Jagannathan et al. (2011) with minor modifications. Briefly, C2C12 cells (with or without 100 μg/ml Puromycin treatment) on 10-cm dishes were rinsed with pre-chilled PBS and incubated in pre-chilled PBS with 1 mg/$MgCl_2$ on ice at 4°C for 20 min to disrupt microtubules. After removing the PBS/$MgCl_2$ buffer completely, 0.75 mL permeabilization buffer (110 mM KCl, 25 mM K-HEPES pH 7.4, 2.5 mM $MgCl_2$, 0.1 mM EGTA, 0.015% (w/v) digitonin (sigma), 1 mM DTT, 1 x FAST Protease Inhibitor Cocktail (sigma), and 40U/mL SuperaseIn RNase Inhibitor (Thermo Fisher Scientific) was added to each dish to evenly cover all the cells. After 10 min incubation on ice, the dishes were tilted

on ice to allow collection of the cytosolic fraction. The dishes were then gently washed with 8 mL pre-chilled wash buffer (110 mM KCl, 25 mM K-HEPES pH 7.4, 2.5 mM $MgCl_2$, 0.1 mM EGTA, 0.004% (w/v) digitonin (sigma), 1 mM DTT, 1 x FAST Protease Inhibitor Cocktail (sigma), and 40U/mL SuperaseIn RNase Inhibitor (Thermo Fisher Scientific)). After removing all the wash buffer, 0.75 mL lysis buffer (200 mM KCl, 25 mM K-HEPES pH 7.4, 10 mM $MgCl_2$, 1% (v/v) NP40, 0.5% (w/v) sodium deoxycholate, 1 mM DTT, 1 x FAST Protease Inhibitor Cocktail (sigma), and 40U/mL SuperaseIn RNase Inhibitor (Thermo Fisher Scientific)) was added to each dish to evenly cover all the cells. After 10 min incubation on ice, the dishes were tilted on ice to allow collection of the membrane fraction. The dishes were then gently washed with 8 mL pre-chilled PBS buffer. After removing all the PBS, the insoluble fraction on the dish was saved at −80°C for RNA or protein extraction. The cytosolic and membrane fractions were centrifuged at 700 x g and subsequently 1400 x g for 5 min in a micro-centrifuge to remove any debris. RNA from the lysate was extracted using 10x volume TRIZol reagent (Thermo Fisher Scientific).

Polysome profiling—We followed the polysome profiling protocol described in Esposito et al. (2010) with some minor modifications. Specifically, about 20% of the supernatant was saved for cytoplasmic RNA extraction. The rest of the supernatant was transferred to a new tube and was centrifuged at 14,000 x g for 5 min to remove mitochondria. The resulting supernatant was layered onto a 10 mL 10%–50% linear sucrose gradient in a polyallomer tube (Beckman Coulter), containing 20 mM HEPES-KOH pH 7.5, 15 mM MgCl2, 80 mM KCl, 2mM DTT, and 100 μg/ml cycloheximide, followed by centrifugation in an SW-41 Ti rotor at 39,000 rpm at 4°C for 2 hr. The gradient was then fractionated using a system comprising a syringe pump (Harvard Apparatus model 11), a density gradient fractionator (Brandel), and an ISCO UA-6 UV/VIS detector. The lysate was partitioned into three fractions based on UV absorbance, i.e., ribosome-free, monosome, and polysome fractions. Monosome and polysome fractions were incubated at 65°C for 5 min with 25 mM EDTA (pH 8.0), 10 mM Tris-HCl (pH7.0), and 1% SDS, followed by extraction of RNA with phenol-chloroform and Ethanol precipitation.

3′READS—The 3′READS protocol (3′READS+ version) was used in this study as previously described (Zheng et al., 2016). Briefly, poly(A)$^+$ RNA was captured with oligo d$(T)_{25}$ magnetic beads (NEB) and fragmented on-beads by RNase III. After washing away free RNA fragments, the poly(A)-containing RNA fragments were eluted from the beads and precipitated with ethanol and ligated to heat-denatured 5′adaptor (5′-CCUUGGCACCCGAGAAUUCCANNNN) with T4 RNA ligase 1 (NEB). The ligated products were then captured biotin-T_{15}-$(+TT)_5$(Exiqon), where +T is locked nucleic acid, bound to Dynabeads MyOne Streptavidin C1 (Thermo Fisher). Bound RNA was digested with RNase H, which also eluted RNA from the beads. Eluted RNA fragments were precipitated with ethanol after washing with RNase H buffer. Purified RNA fragments were then ligated to a 5′-adenylated 3′ adaptor (5′-rApp/NNNNGATCGTCGGACTGTAGAACTCTGAAC/3ddC) with T4 RNA Ligase 2 (truncated KQ, NEB). The ligation products were reverse-transcribed M-MLV reverse transcriptase (Promega), followed by PCR amplification using Phusion high-fidelity DNA polymerase (NEB) and bar-coded PCR primers for 15–18 cycles. PCR products were size-selected twice

with AMPure XP beads (Beckman Coulter). The size and quantity of the libraries were examined on an Agilent Bioanalyzer. The cDNA libraries were purified and sequenced on an Illumina HiSeq (1x150 nt).

Plasmid construction—For psiCHECK2-*Nmt1* full 3′ UTR, used as a reporter plasmid, a PCR product containing *Nmt1* full 3′ UTR was amplified from C2C12 cDNA and was inserted into psiCHECK2 plasmid using XhoI and NotI. For pcDNA3-*Nmt1* E1-4 and pcDNA3-hRluc that were used for generating *Nmt1* and hRluc RNA probes, respectively, PCR products containing the Nmt1 E1-4 (423 bp) and hRluc (698 bp) sequences were amplified from C2C12 cDNA and psiCHECK2 plasmid, respectively, and were inserted into pcDNA3 using BamH I (NEB) based on the Gibson Assembly method. All primers used in this study are listed in Table S7.

Cell transfection—C2C12 cells in a six-well plate (60%–70% confluence) were transfected with 1.5 μg reporter plasmids with Lipofectamine LTX (Invitrogen) according to the manufacturer's protocol. Cell media were replaced after 6 hr of transfection.

Northern blot analysis—Northern blotting was carried out using the Digoxigenin (DIG) Northern Starter Kit (Roche) according to the manufacturer's protocol. Briefly, 20 μg of total RNA from non-transfected C2C12 cells or 1.5 μg from transfected C2C12 cells were resolved on a denaturing agrose gel, and were transferred onto a nylon membrane (Roche), which were then subjected to UV-crosslinking. RNA on the membrane was hybridized with DIG-labeled RNA probes, which were produced from *in vitro* transcription using PCR products as template. Primers used for the probes were listed in Table S7. DIG-labeled probes were detected by anti-DIG antibody conjugated with alkaline phosphatase. Chemiluminescence from ECL reaction kit was detected on X-ray film. Quantification was carried out in ImageJ.

3′READS data analysis—3′READS data were processed as previously described (Zheng et al., 2016). Briefly, the sequence corresponding to 5′ adaptor was first removed from raw reads by using Cutadapt (Martin, 2011). Reads with short inserts (< 23 nucleotides) were discarded. The remaining reads were then mapped to the genome (mm9) by using bowtie2 (local mode) (Langmead and Salzberg, 2012). The 5′ random nucleotides derived from the 3′ adaptor were removed before mapping. Reads with a mapping quality score (MAPQ) ≥ 10 were kept for further analysis. Reads with ≥ 2 non-genomic 5′-Ts after alignment were called PAS reads. PASs within 24 nucleotides from each other were clustered as previously described (Hoque et al., 2013). The PAS read counts mapped to genes were normalized by the median ratio method in the DESeq program (Anders and Huber, 2010). For 3′UTR APA analysis, the two most abundant APA isoforms (based on PAS reads) were selected. We required that both PASs should be in the last exon of gene. They were named proximal PAS (pPAS) and distal PAS (dPAS) isoforms. Significant APA events were those with ΔMLS or Relative Expression Difference (RED) > $\log_2(1.2)$ or < $-\log_2(1.2)$ and *P*-value < 0.05 (Fisher's exact test or DEXSeq analysis) between samples. RED was calculated as the difference in log2(RPM ratio) of the two APA isoforms between two sample sets. The aUTR size was the distance between the two 3′UTR APA sites.

Analysis of APA using RNA-seq data—3′UTR APA analysis using RNA-seq data was carried out by using the APAlyzer program (Wang and Tian, 2020). Relative expression (RE) was calculated as $\log_2$(aUTR read number / cUTR read number) in a sample.

Gene ontology analysis—Gene ontology (GO) analysis was carried out by using the GOstats Bioconductor package (Falcon and Gentleman, 2007). Generic terms (associated with more than 1,000 genes) were discarded. To reduce redundancy in reporting, any GO term with a gene overlap greater than 75% with a more significant term was discarded. p values are based on the hypergeometric test.

Protein subcellular localization analaysis—We used MetaSecKB database (Meinken et al., 2015) to annotate protein localization. Mitochondrial proteins were defined by MetaSecKB. Secreted proteins were those annotated as secreted or highly likely secreted. Membrane proteins were those annotated as ER, Golgi, lysosome, peroxisome, or plasma membrane proteins. Proteins annotated with other locations or unannotated ones were combined and called ‘other’.

Classification of transcripts based on MLS—A Gaussian mixture modeling method, based on the sklearn.mixture package of the Scikit-learn software (Pedregosa et al., 2011), was used to classify transcripts based on MLS values (2 replicates) in Puro+ and Puro− samples. $p < 0.05$ was used to assign transcripts to groups. In APA analysis, a gene was assigned to a group if all of its isoforms are in the same group.

TiERA predictive model—We used the XGBoost method (Chen and Guestrin, 2016) to construct a statistical model for TiERA prediction. Input data for the model were MLS values (2 replicates) in Puro+ samples. Only well expressed group I transcripts (RPM > 5) were used for training. Transcripts were weighted based on number of detected reads for each transcript. Hyperparameters were fine-tuned with RandomizedSearchCV from sklearn.model_selection package of the Scikit-learn software (Pedregosa et al., 2011) with a 5-fold cross-validation splitting strategy. Input features to train the model were based on size, dimer frequency, and predicted RNA structure. Features were derived from RefSeq sequences.

RNA structure analysis—CDS and 3′UTR sequences were divided into sub-sequences based on a 100-nt moving window with a 50-nt overlap between adjacent windows. RNAfold (Lorenz et al., 2011) was used to calculate the minimum free energy (MFE, kcal/mol) for each window, and the MFE values of all windows were then averaged to represent the whole sequence. *In vivo* mouse embryonic stem cell icSHAPE (in vivo click selective 2-hydroxyl acylation and profiling experiment) reactivity scores (Sun et al., 2019) were extracted for CDSs or 3′UTRs. A 20-nt sliding window over a transcript region was used to generate Gini indices across the region. The median of Gini indices of all windows was used to represent the whole region.

QUANTIFICATION AND STATISTICAL ANALYSIS

Student's t test was used to determine statistical significance between groups, unless specified otherwise. Significant level of APA changes was assessed by using DEXseq (when there were replicates) or Fisher's exact test (when there were no replicates). Significance of gene expression difference was assessed by using DEseq (when there were replicates) or the Fisher's exact test (when there were no replicates). K-S (Kolmogorov–Smirnov) test was used to compare data distributions in different gene sets. The Wilcoxon test was used to compare aUTR size-based gene bins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Dongming He and Terri G. Kinzy for help with polysome profiling, Wei Wang for help with RNA structure analysis, and Wendy Gilbert, Gideon Dreyfuss, and members of B.T. lab for helpful discussions. This work was funded by an NIH grant (R01 GM129069) to B.T. B.T. is also supported by a recruitment grant to The Wistar Institute from The Pew Charitable Trusts. L.C.C. was supported in part by NIH training grant T32 GM008339 and the Rutgers Presidential Fellowship Award. Work in the H.C. lab was supported by National Natural Science Foundation of China (31925008). We are also grateful to The Wistar Institute's Bioinformatics facility for providing technical support. Funding support for The Wistar Institute core facilities was provided by Cancer Center Support Grant P30 CA010815.

REFERENCES

Akopian D, Shen K, Zhang X, and Shan SO (2013). Signal recognition particle: an essential protein-targeting machine. Annu. Rev. Biochem 82, 693–721. [PubMed: 23414305]

Anders S, and Huber W (2010). Differential expression analysis for sequence count data. Genome Biol. 11, R106. [PubMed: 20979621]

Anders S, Reyes A, and Huber W (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. 22, 2008–2017. [PubMed: 22722343]

Aviner R (2020). The science of puromycin: From studies of ribosome function to applications in biotechnology. Comput. Struct. Biotechnol. J 18, 1074–1083. [PubMed: 32435426]

Bartel DP (2018). Metazoan MicroRNAs. Cell 173, 20–51. [PubMed: 29570994]

Benoit Bouvrette LP, Cody NAL, Bergalet J, Lefebvre FA, Diot C, Wang X, Blanchette M, and Lécuyer E (2018). CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in *Drosophila* and human cells. RNA 24, 98–113. [PubMed: 29079635]

Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, and Dreyfuss G (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell 150, 53–64. [PubMed: 22770214]

Béthune J, Jansen RP, Feldbrügge M, and Zarnack K (2019). Membrane-Associated RNA-Binding Proteins Orchestrate Organelle-Coupled Translation. Trends Cell Biol. 29, 178–188. [PubMed: 30455121]

Chen T, and Guestrin C (2016). XGBoost: A Scalable Tree Boosting System. arXiv, arXiv:1603.02754. https://arxiv.org/abs/1603.02754.

Cheng LC, Zheng D, Baljinnyam E, Sun F, Ogami K, Yeung PL, Hoque M, Lu CW, Manley JL, and Tian B (2020). Widespread transcript shortening through alternative polyadenylation in secretory cell differentiation. Nat. Commun 11, 3182. [PubMed: 32576858]

Cioni JM, Lin JQ, Holtermann AV, Koppers M, Jakobs MAH, Azizi A, Turner-Bridger B, Shigeoka T, Franze K, Harris WA, and Holt CE (2019a). Late Endosomes Act as mRNA Translation Platforms and Sustain Mitochondria in Axons. Cell 176, 56–72.e15. [PubMed: 30612743]

Cioni JM, Lin JQ, Holtermann AV, Koppers M, Jakobs MAH, Azizi A, Turner-Bridger B, Shigeoka T, Franze K, Harris WA, and Holt CE (2019b). Late Endosomes Act as mRNA Translation Platforms and Sustain Mitochondria in Axons. Cell 176, 56–72.e15. [PubMed: 30612743]

Cui XA, and Palazzo AF (2014). Localization of mRNAs to the endoplasmic reticulum. Wiley Interdiscip. Rev. RNA 5, 481–492. [PubMed: 24644132]

Cui XA, Zhang H, and Palazzo AF (2012). p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. PLoS Biol. 10, e1001336. [PubMed: 22679391]

Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, and Babak T (2012). A quantitative atlas of polyadenylation in five mammals. Genome Res. 22, 1173–1183. [PubMed: 22454233]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Elkon R, Ugalde AP, and Agami R (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet 14, 496–506. [PubMed: 23774734]

English AR, and Voeltz GK (2013). Endoplasmic reticulum structure and interconnections with other organelles. Cold Spring Harb. Perspect. Biol 5, a013227. [PubMed: 23545422]

Esposito AM, Mateyak M, He D, Lewis M, Sasikumar AN, Hutton J, Copeland PR, and Kinzy TG (2010). Eukaryotic polyribosome profile analysis. J. Vis. Exp (40), 1948. [PubMed: 20567211]

Falcon S, and Gentleman R (2007). Using GOstats to test gene lists for GO term association. Bioinformatics 23, 257–258. [PubMed: 17098774]

Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, and Greenberg ME (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60, 1022–1038. [PubMed: 19109909]

Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, and Xu A (2011). Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. Genome Res. 21, 741–747. [PubMed: 21474764]

Garneau NL, Wilusz J, and Wilusz CJ (2007). The highways and byways of mRNA decay. Nat. Rev. Mol. Cell Biol 8, 113–126. [PubMed: 17245413]

Gruber AJ, and Zavolan M (2019). Alternative cleavage and polyadenylation in health and disease. Nat. Rev. Genet 20, 599–614. [PubMed: 31267064]

Guvenek A, and Tian B (2018). Analysis of alternative cleavage and polyadenylation in mature and differentiating neurons using RNA-seq data. Quant. Biol 6, 253–266. [PubMed: 31380142]

Higuchi Y, Ashwin P, Roger Y, and Steinberg G (2014). Early endosome motility spatially organizes polysome distribution. J. Cell Biol 204, 343–357. [PubMed: 24493587]

Hogg JR, and Goff SP (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell 143, 379–389. [PubMed: 21029861]

Hollerer I, Curk T, Haase B, Benes V, Hauer C, Neu-Yilik G, Bhuvanagiri M, Hentze MW, and Kulozik AE (2016). The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates mammalian mRNA expression in a quantitative and qualitative fashion. RNA 22, 1441–1453. [PubMed: 27407180]

Hollien J, Lin JH, Li H, Stevens N, Walter P, and Weissman JS (2009). Regulated Ire1-dependent decay of messenger RNAs in mammalian cells. J. Cell Biol 186, 323–331. [PubMed: 19651891]

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, and Tian B (2013). Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. Nat. Methods 10, 133–139. [PubMed: 23241633]

Huang MT (1975). Harringtonine, an inhibitor of initiation of protein biosynthesis. Mol. Pharmacol 11, 511–519. [PubMed: 1237080]

Hyde M, Block-Alper L, Felix J, Webster P, and Meyer DI (2002). Induction of secretory pathway components in yeast is associated with increased stability of their mRNA. J. Cell Biol 156, 993–1001. [PubMed: 11901166]

Jagannathan S, Nwosu C, and Nicchitta CV (2011). Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation. Methods Mol. Biol 714, 301–321. [PubMed: 21431749]

Ji Z, and 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. [PubMed: 20037631]

Ji Z, Lee JY, Pan Z, Jiang B, and Tian B (2009). Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl. Acad. Sci. USA 106, 7028–7033. [PubMed: 19372383]

Kejiou NS, and Palazzo AF (2017). mRNA localization as a rheostat to regulate subcellular gene expression. Wiley Interdiscip. Rev. RNA 8.

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Lerner RS, and Nicchitta CV (2006). mRNA translation is compartmentalized to the endoplasmic reticulum following physiological inhibition of cap-dependent translation. RNA 12, 775–789. [PubMed: 16540694]

Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, and Tian B (2015). Systematic profiling of poly(A)+ transcripts modulated by core 3′ end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS Genet. 11, e1005166. [PubMed: 25906188]

Lianoglou S, Garg V, Yang JL, Leslie CS, and Mayr C (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. 27, 2380–2396. [PubMed: 24145798]

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, and Hofacker IL (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol 6, 26. [PubMed: 22115189]

Ma W, and Mayr C (2018). A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3'UTR-Mediated Protein-Protein Interactions. Cell 175, 1492–1506.e19. [PubMed: 30449617]

Martin M (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnet 17, 10–12.

Masamha CP, and Wagner EJ (2018). The contribution of alternative polyadenylation to the cancer phenotype. Carcinogenesis 39, 2–10. [PubMed: 28968750]

Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, and Wagner EJ (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature 510, 412–416. [PubMed: 24814343]

Mayr C (2018). What Are 3′ UTRs Doing? Cold Spring Harb. Perspect. Biol 11, a034728.

Mayr C, and Bartel DP (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138, 673–684. [PubMed: 19703394]

Meinken J, Walker G, Cooper CR, and Min XJ (2015). MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. Database 2015, bav077. [PubMed: 26255309]

Meinnel T, Dian C, and Giglione C (2020). Myristoylation, an Ancient Protein Modification Mirroring Eukaryogenesis and Evolution. Trends Biochem. Sci 45, 619–632. [PubMed: 32305250]

Miura P, Shenker S, Andreu-Agullo C, Westholm JO, and Lai EC (2013). Widespread and extensive lengthening of 3′ UTRs in the mammalian brain. Genome Res. 23, 812–825. [PubMed: 23520388]

Morris AR, Bos A, Diosdado B, Rooijers K, Elkon R, Bolijn AS, Carvalho B, Meijer GA, and Agami R (2012). Alternative cleavage and polyadenylation during colorectal cancer development. Clin. Cancer Res 18, 5256–5266. [PubMed: 22874640]

Nathans D (1964). Puromycin Inhibition of Protein Synthesis: Incorporation of Puromycin into Peptide Chains. Proc. Natl. Acad. Sci. USA 51, 585–592. [PubMed: 14166766]

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, and Dubourg V (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res 12, 2825–2830.

Reid DW, and Nicchitta CV (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J. Biol. Chem 287, 5518–5527. [PubMed: 22199352]

Reid DW, and Nicchitta CV (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol 16, 221–231. [PubMed: 25735911]

Reid DW, Chen Q, Tay AS, Shenolikar S, and Nicchitta CV (2014). The unfolded protein response triggers selective mRNA release from the endoplasmic reticulum. Cell 158, 1362–1374. [PubMed: 25215492]

Sadek J, Omer A, Hall D, Ashour K, and Gallouzi IE (2019). Alternative polyadenylation and the stress response. Wiley Interdiscip. Rev. RNA 10, e1540. [PubMed: 31050180]

Salomons FA, Menéndez-Benito V, Böttcher C, McCray BA, Taylor JP, and Dantuma NP (2009). Selective accumulation of aggregation-prone proteasome substrates in response to proteotoxic stress. Mol. Cell. Biol 29, 1774–1785. [PubMed: 19158272]

Sandberg R, Neilson JR, Sarma A, Sharp PA, and Burge CB (2008). Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320, 1643–1647. [PubMed: 18566288]

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. [PubMed: 22743772]

Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, and Shi Y (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17, 761–772. [PubMed: 21343387]

Singh P, Alley TL, Wright SM, Kamdar S, Schott W, Wilpan RY, Mills KD, and Graber JH (2009). Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. Cancer Res. 69, 9422–9430. [PubMed: 19934316]

Sun L, Fazal FM, Li P, Broughton JP, Lee B, Tang L, Huang W, Kool ET, Chang HY, and Zhang QC (2019). RNA structure maps across mammalian cellular compartments. Nat. Struct. Mol. Biol 26, 322–330. [PubMed: 30886404]

Tian B, and Manley JL (2017). Alternative polyadenylation of mRNA precursors. Nat. Rev. Mol. Cell Biol 18, 18–30. [PubMed: 27677860]

Tsuboi T, Viana MP, Xu F, Yu J, Chanchani R, Arceo XG, Tutucci E, Choi J, Chen YS, Singer RH, et al. (2020). Mitochondrial volume fraction and translation duration impact mitochondrial mRNA localization and protein synthesis. eLife 9, e57814. [PubMed: 32762840]

Tushev G, Glock C, Heumüller M, Biever A, Jovanovic M, and Schuman EM (2018). Alternative 3′ UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. Neuron 98, 495–511.e6. [PubMed: 29656876]

Udenwobele DI, Su RC, Good SV, Ball TB, Varma Shrivastav S, and Shrivastav A (2017). Myristoylation: An Important Protein Modification in the Immune Response. Front. Immunol 8, 751. [PubMed: 28713376]

Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, Blue SM, Chen JY, Cody NAL, Dominguez D, et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature 583, 711–719. [PubMed: 32728246]

Voeltz GK, Rolls MM, and Rapoport TA (2002). Structural organization of the endoplasmic reticulum. EMBO Rep. 3, 944–950. [PubMed: 12370207]

Wang R, and Tian B (2020). APAlyzer: a bioinformatics package for analysis of alternative polyadenylation isoforms. Bioinformatics 36, 3907–3909. [PubMed: 32321166]

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, and Burge CB (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. [PubMed: 18978772]

Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al. (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell 750, 710–724.

Wang R, Nambiar R, Zheng D, and Tian B (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. Nucleic Acids Res. 46 (D1), D315–D319. [PubMed: 29069441]

Wang R, Zheng D, Wei L, Ding Q, and Tian B (2019). Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes. Cell Rep. 26, 2766–2778.e6. [PubMed: 30840896]

Williams CC, Jan CH, and Weissman JS (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. Science 346, 748–751. [PubMed: 25378625]

Williamson CD, Wong DS, Bozidis P, Zhang A, and Colberg-Poley AM (2015). Isolation of Endoplasmic Reticulum, Mitochondria, and Mitochondria-Associated Membrane and Detergent Resistant Membrane Fractions from Transfected Cells and from Human Cytomegalovirus-Infected Primary Fibroblasts. Curr. Protoc. Cell Biol 68, 3.27.21–3.27.33. [PubMed: 26331984]

Witten AJ, Ejendal KFK, Gengelbach LM, Traore MA, Wang X, Umulis DM, Calve S, and Kinzer-Ursem TL (2017). Fluorescent imaging of protein myristoylation during cellular differentiation and development. J. Lipid Res 58, 2061–2070. [PubMed: 28754825]

Wu JJ, and Palazzo AF (2020). Visualization of Endoplasmic Reticulum-Associated mRNA in Mammalian Cells. Methods Mol. Biol 2166, 35–49. [PubMed: 32710402]

Zhang H, Lee JY, and Tian B (2005). Biased alternative polyadenylation in human tissues. Genome Biol. 6, R100. [PubMed: 16356263]

Zheng D, Liu X, and Tian B (2016). 3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA. RNA 22, 1631–1639. [PubMed: 27512124]

Zheng D, Wang R, Ding Q, Wang T, Xie B, Wei L, Zhong Z, and Tian B (2018). Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. Nat. Commun 9, 2268. [PubMed: 29891946]

Highlights

- Transcripts differ in translation-independent ER association (TiERA)
- Various intrinsic features of a transcript determine its TiERA potential
- Alternative polyadenylation isoforms have distinct TiERA potentials
- 3′ UTR size change in cell proliferation and differentiation alters TiERA

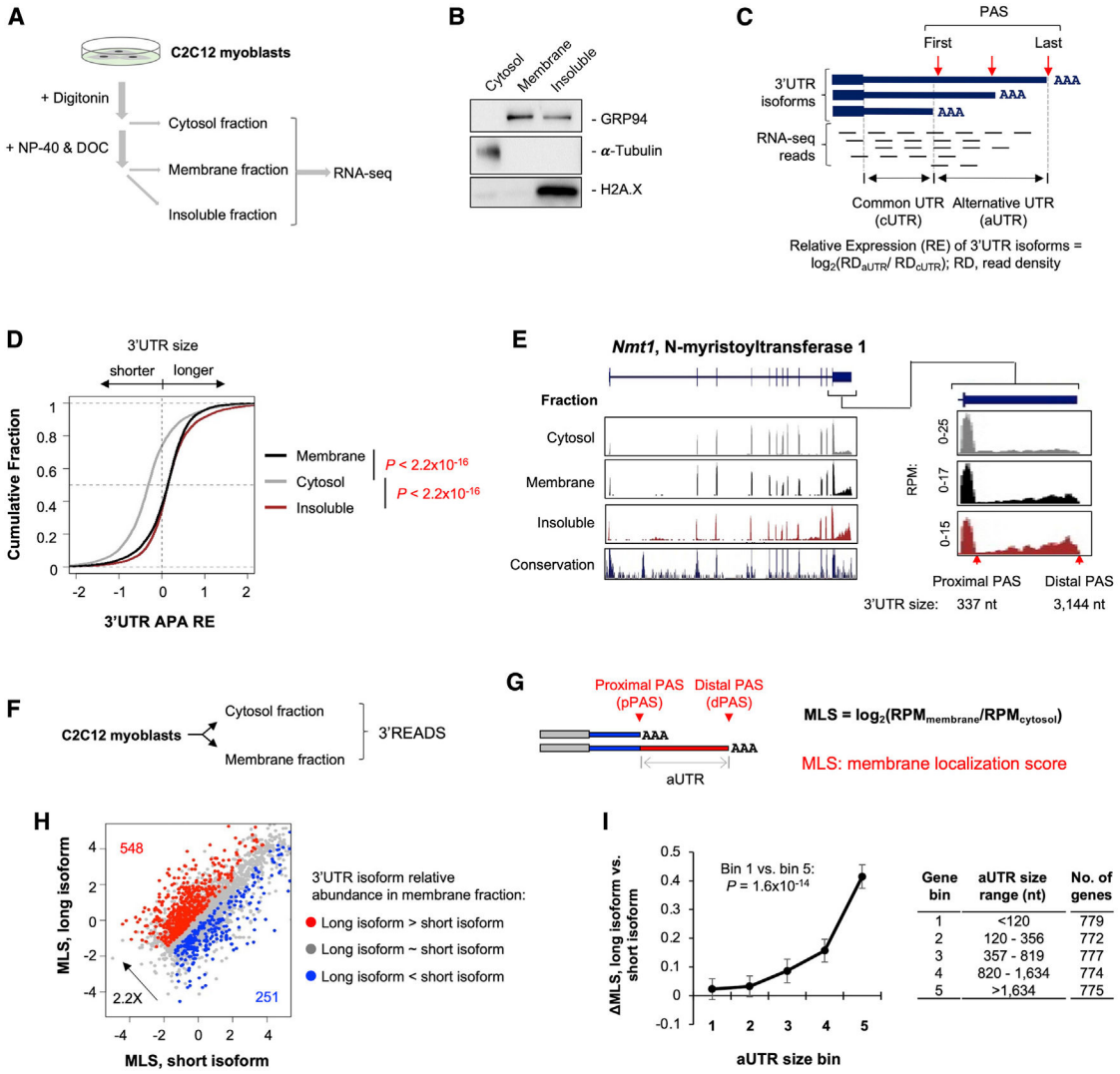


Figure 1. 3′ UTR isoforms are distributed differently in subcellular compartments

(A) Schematic showing subcellular fractionation of mouse C2C12 myoblasts into cytosol, membrane, and insoluble fractions, followed by RNA-seq.

(B) Western blot analysis of several proteins enriched for distinct subcellular compartments.

(C) Schematic showing analysis of 3′ UTR size difference across fractions using RNA-seq reads. The formula for relative expression (RE) is shown.

(D) Cumulative density function (CDF) curves of RE values of cytosolic, membrane, and insoluble fractions. p values are based on K-S test.

(E) UCSC tracks showing an example gene, *Nmt1*. Conservation is based on 100 vertebrates. Pictures of the whole gene (left) and the last exon (right) are shown.

(F) Schematic showing subcellular fractionation of mouse C2C12 myoblasts into cytosol and membrane fractions, followed by 3′ region extraction and deep sequencing (3′READS).

(G) Schematic showing analysis of 3′ UTR isoform abundance in membrane versus cytosol fractions based on 3′READS data. The alternative 3’ UTR (aUTR) is the region present in the long isoform but absent from the short isoform. The MLS formula is indicated. RPM, reads per million mapped (PAS reads only).

(H) Scatterplot comparing the MLS of the long 3′ UTR isoform (y axis) and short 3′ UTR isoform (x axis). Each dot represents a gene. The two most abundant 3′ UTR isoforms (based on 3′READS data) are selected from each gene for comparison. Those with significantly different MLSs ($p < 0.05$, Fisher's exact test) are highlighted in color. (I) Relationship between aUTR size and MLS difference between long and short isoforms. Genes are grouped into five similarly sized bins (~775 genes) based on their aUTR size (size ranges are shown). The p value (Wilcoxon test) comparing bin 1 and bin 5 genes is indicated. The median of each bin is plotted. Error bars indicate SEM.

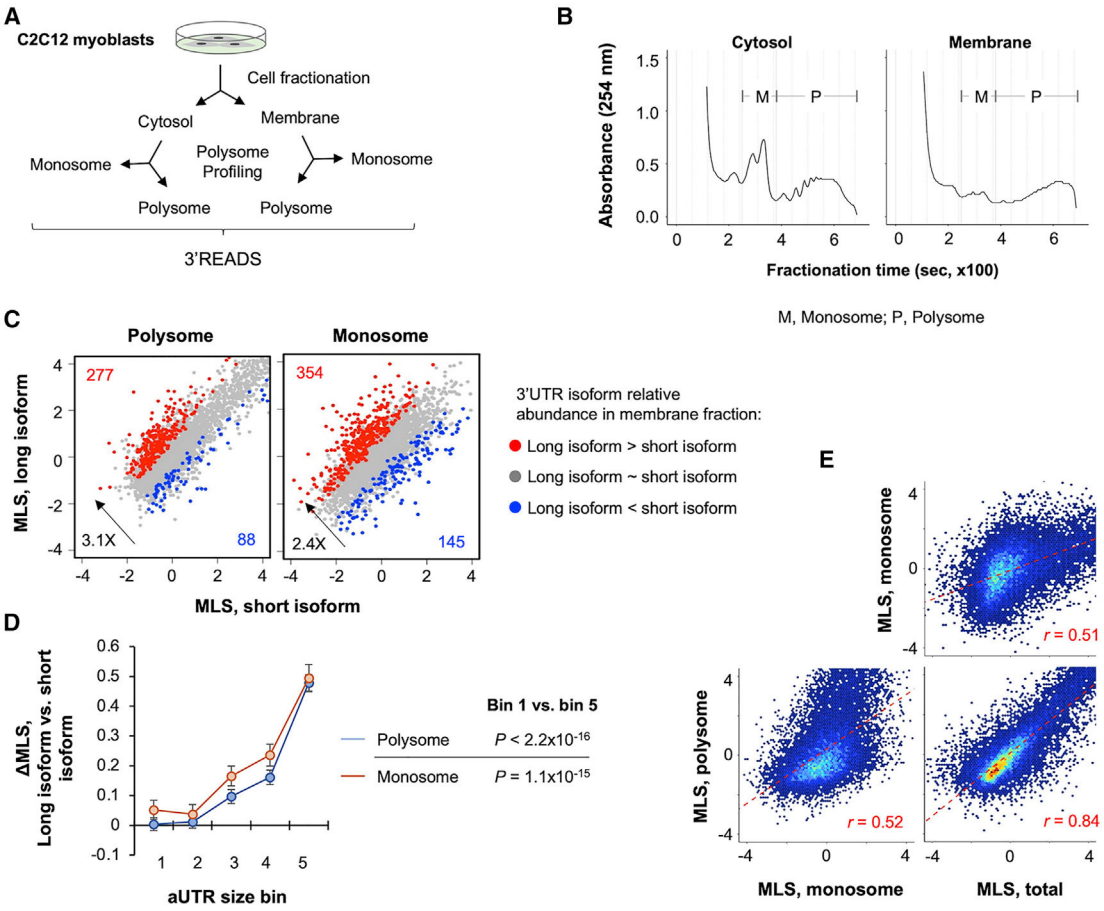


Figure 2. P profiling analysis reveals differences in ER association between 3′ UTR isoforms

(A) Schematic of cell fractionation and P profiling of C2C12 myoblasts. RNAs extracted from Polysome (P) and Monosome (M) fractions were subject to 3′READS analysis.

(B) Polysome profiles of cytosol and membrane fractions. P and M fractions are indicated. Gray lines indicate fractions.

(C) Scatterplot comparing the MLS of the long 3′ UTR isoform (y axis) and short 3′ UTR isoform (x axis). Genes whose isoforms have significantly different MLSs ($p < 0.05$, DEXSeq) are highlighted in color. See Figure 1G for MLS calculation.

(D) Relationship between aUTR size and MLS difference (ΔMLS) between long and short isoforms in P and M samples. Genes are grouped into five similarly sized bins based on their aUTR size (see Figure 1I for details). The p value (Wilcoxon test) comparing bins 1 and 5 is indicated. The median of each bin is plotted. Error bars indicate SEM.

(E) Scatterplots comparing MLSs among total cell extract (total), P, and M samples. Pearson correlation coefficients are indicated.

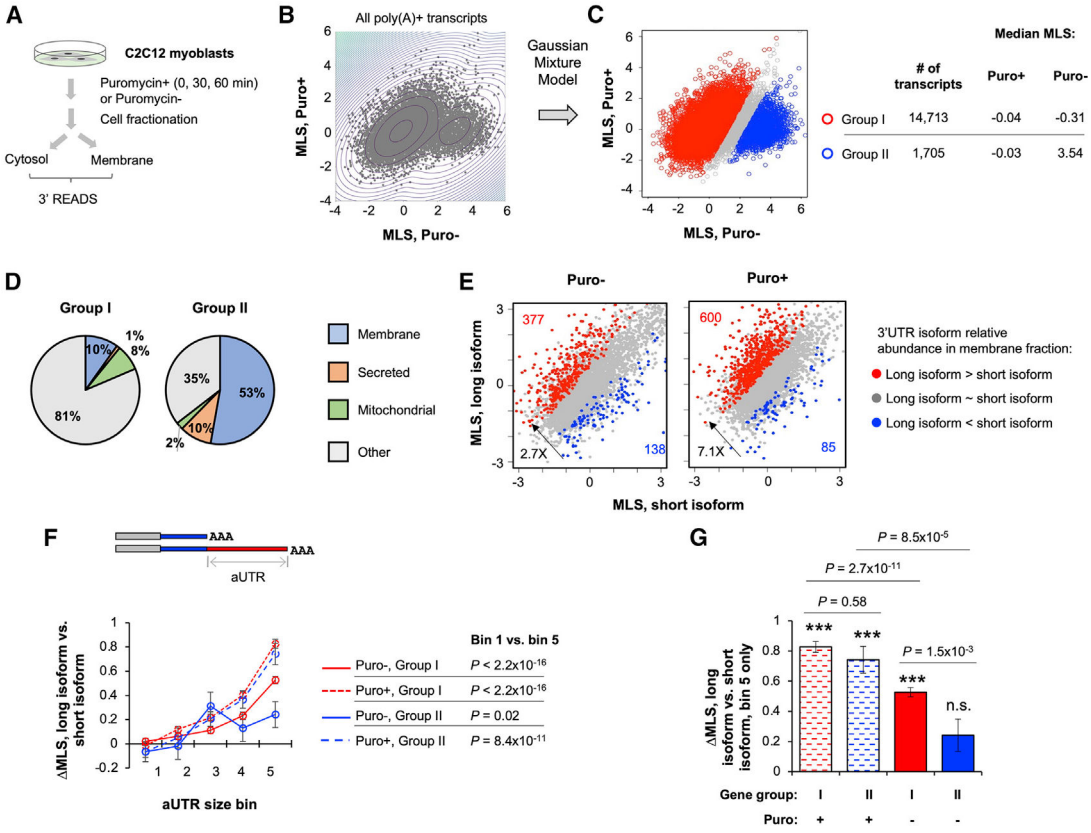


Figure 3. Identification of transcripts with translation-independent ER association (TiERA)

(A) Schematic of the experiment using Puro to inhibit translation in C2C12 myoblasts. Extracted RNAs were subjected to 3′READS analysis.

(B) Scatterplot comparing transcript MLSs between Puro-treated (Puro+) and untreated (Puro−) samples.

(C) Identification of transcripts with translation-independent (group I) or translation-dependent (group II) ER association using Gaussian mixture modeling. The number of transcripts and median MLS for each group in Puro+ or Puro− samples are indicated. There were 1,604 transcripts that were not classified (gray circles).

(D) Pie charts of group I and group II transcripts, displaying proportions of transcripts encoding secreted, membrane, mitochondrial, or other proteins. Protein location annotation was based on the MetazSecKB database.

(E) Scatterplot comparing the MLS of the long 3′ UTR isoform (y axis) and short 3′ UTR isoform (x axis) in Puro− (left) and Puro+ (right) samples. Genes whose isoforms have significantly different MLSs ($p < 0.05$, DEXSeq) are highlighted in color.

(F) Relationship between aUTR size and MLS difference between long and short isoforms (ΔMLS) for group I and group II genes in Puro+ and Puro− samples. Genes are grouped into aUTR size bins as in Figure 1I. The p value (Wilcoxon test) comparing bins 1 and 5 is indicated. The median of each bin is plotted. Error bars indicate SEM.

(G) Bar graph showing ΔMLS of genes with long aUTRs (top 20%, bin 5) in Puro+ and Puro− samples. Group I and II genes are shown separately. Data are from (E). The p value (Wilcoxon test) comparing ΔMLS of each group with the value 0 (no difference) is

indicated. ***$p < 0.001$; n.s., $p > 0.05$. p values (Wilcoxon test) comparing different groups are also shown. The median of each group is shown. Error bars indicate SEM.

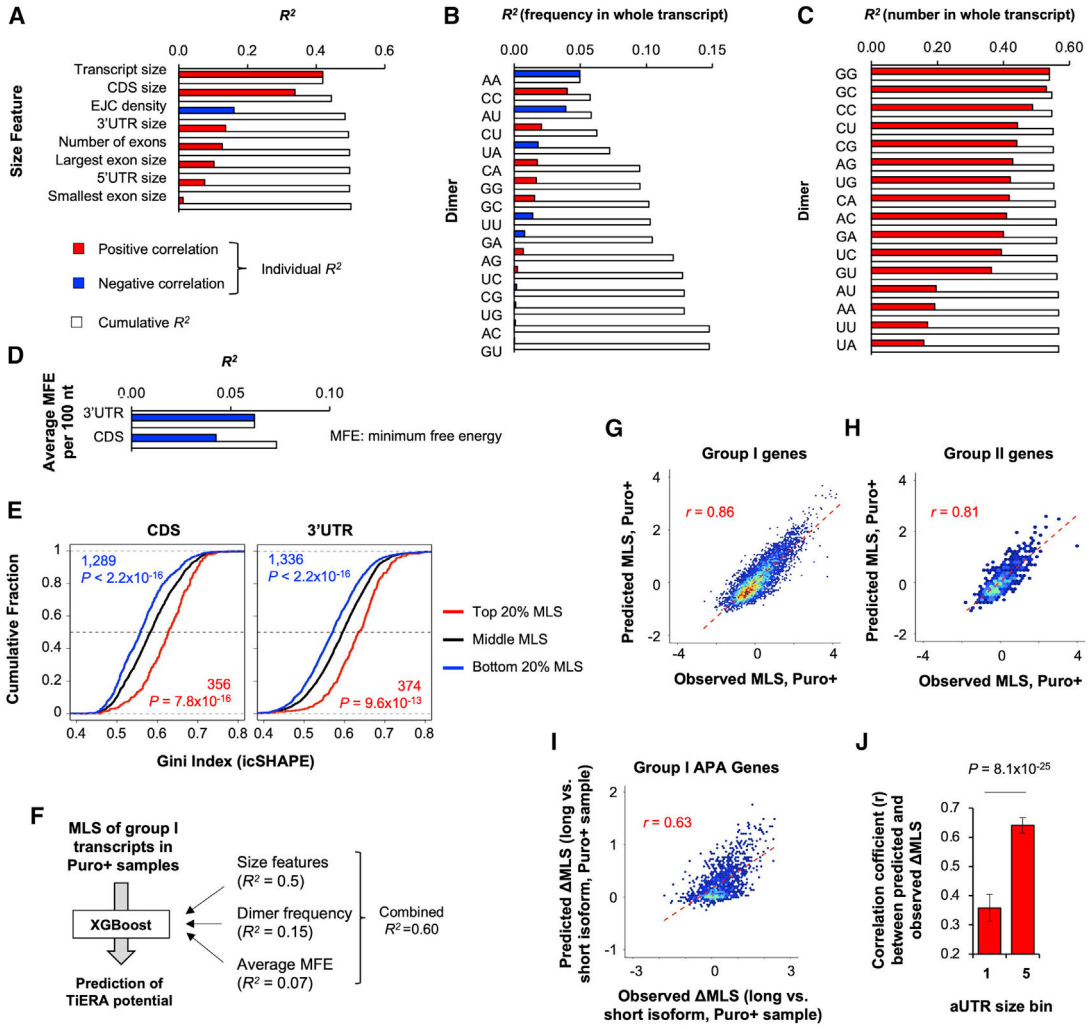


Figure 4. Transcript features determine TiERA

(A) Regression analysis of size features versus MLS in Puro+ samples. Features are sorted according to individual R^2 values. The cumulative R^2 value for a given feature is based on using the feature together with all other features with a better individual R^2 value. Red and blue bars denote positive and negative correlations with MLS, respectively.

(B) As in (A), except that dimer frequencies in the whole transcript are used as features.

(C) As in (A), except that dimer numbers in the whole transcript are used as features.

(D) As in (A), except that the average minimum free energy (MFE) per 100 nt in the 3′ UTR or CDS is used as a feature.

(E) CDF plots of icSHAPE Gini indices of the CDS (left) or 3′ UTR (right). icSHAPE Gini indices are based on a 20-nt moving window across each sequence. Transcripts are divided into high-MLS (top 20%, red line), medium-MLS (middle 60%, black line), and low-MLS (bottom 20%, blue line) groups. p values (K-S test) for significance of the difference between red or blue genes and black genes are indicated. Gene numbers for red and blue genes are also indicated.

(F) Schematic showing construction of a statistical model based on the XGboost program for TiERA potentials. MLSs of group I gene transcripts in Puro+ samples were used for training. Size, dimer frequency, and average MFE features were used.
(G) Scatterplot comparing the predicted MLS (y axis) and observed MLS (x axis, Puro+ samples) of group I gene transcripts. Pearson correlation coefficient is indicated.
(H) Same as in (G), except for group II gene transcripts. Pearson correlation coefficient is indicated.
(I) Scatterplot comparing the predicted ΔMLS between long and short 3′ UTR isoforms (y axis) and their observed ΔMLSs in Puro+ samples (x axis). Pearson correlation coefficient is indicated.
(J) Pearson correlation coefficients for predicted versus observed ΔMLSs (long versus short 3′ UTR isoforms) for genes with short aUTRs (bin 1, bottom 20%) and long aUTRs (bin 5, top 20%). Only group I genes are included. The p value (t test) for comparison of the two bins is indicated. The median of each bin is shown. Error bars indicate standard deviation based on data bootstrapping (20 times).

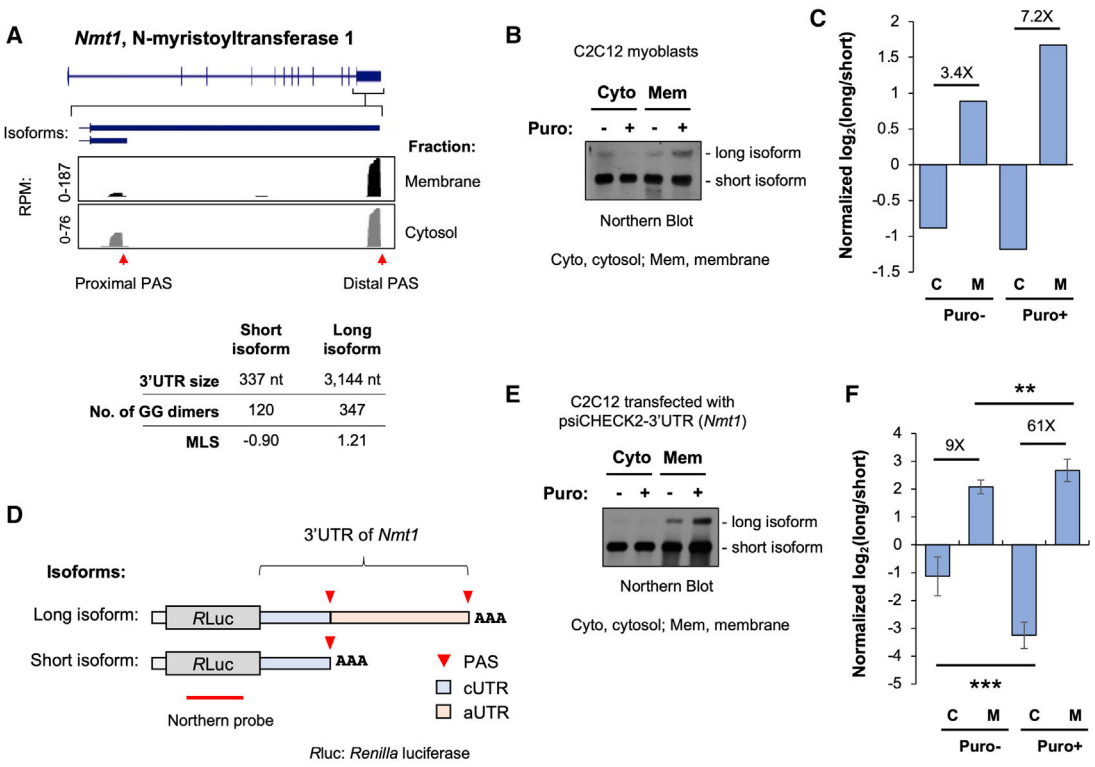


Figure 5. *Nmt1* APA isoforms have different TiERA potentials

(A) UCSC Genome Browser tracks showing 3′READS data for *Nmt1*. 3′ UTR size and MLS for each 3′ UTR isoform are indicated. The number of GG dimers is also indicated.

(B) Northern blot of two *Nmt1* APA isoforms. Cyto, cytosolic fraction; Mem, membrane fraction. Puro treatment is indicated. Two isoforms are indicated.

(C) Normalized $\log_2$(ratio) of the expression level of the long isoform versus short isoform based on the northern blot result in (B). Normalization is based on the mean of all samples.

(D) Schematic of the psiCHECK2-*Nmt1* 3′ UTR vector. *R*Luc, *Renilla* luciferase.

(E) As in (B), except that APA isoforms expressed from the psiCHECK2-*Nmt1* 3′ UTR vector were analyzed by northern blot. A representative image of three replicates is shown.

(F) As in (C), except that quantification is based on (E). Error bars indicate standard deviation of three replicates. The p values (t test) for significance of difference are indicated. **p < 0.05; ***p < 0.01.

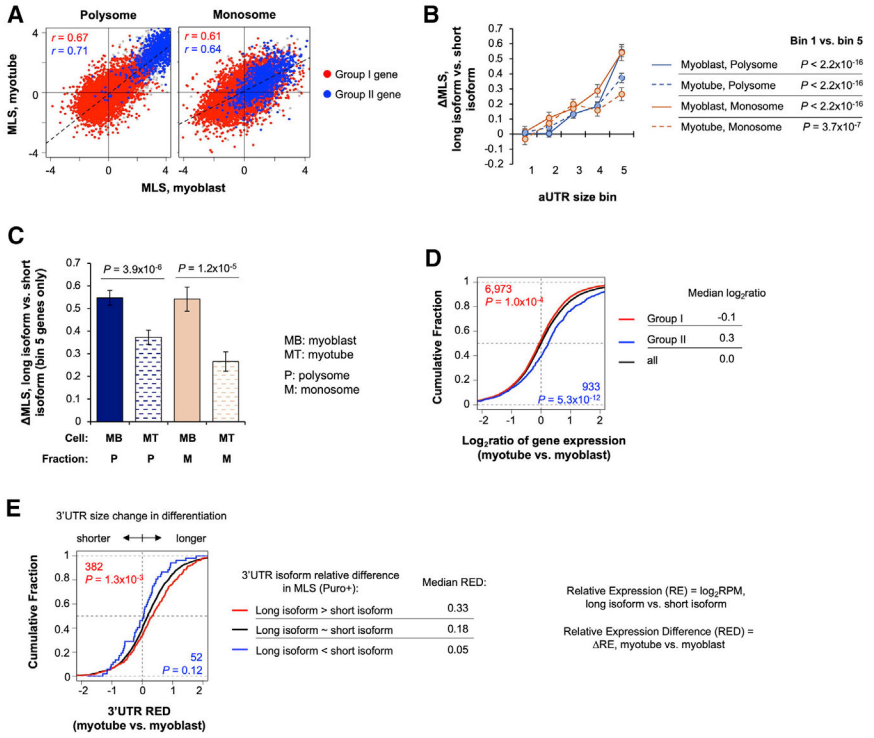


Figure 6. TiERA and APA in cell differentiation

(A) Scatterplot comparing transcript MLS in proliferating C2C12 myoblast cells (x axis) and differentiated myotube cells (y axis). P (left) and M (right) data are shown. Group I and II genes are indicated by different colors. Pearson correlation coefficients are shown for the two gene groups with matching colors.

(B) Relationship between aUTR size and ΔMLS (long versus short isoforms) for group I genes in P and M fractions of myoblasts and myotube samples. Genes are grouped into five similarly sized bins based on their aUTR size, as in Figure 1I. The p value (Wilcoxon test) comparing bins 1 and 5 is indicated. The median of each bin is shown. Error bars indicate SEM.

(C) Bar graph showing MLS of group I genes with long aUTRs (bin 5 in B) in myoblasts (MBs) and myotubes (MTs). *P*-values (Wilcoxon test) comparing transcript groups are indicated. The median of each sample is shown. Error bars are SEM.

(D) CDF plots showing gene expression changes ($\log_2$ ratio of RPM) between proliferating C2C12 MB and differentiated C2C12 MT cells for group I genes (red), group II genes (blue), and all genes (black). The median for each gene set is indicated. p values (K-S test) for significance of difference (red or blue genes versus black genes) are indicated. The gene number for each group is also shown.

(E) CDF plots showing RE difference (RED) of 3′ UTR isoforms in C2C12 differentiation. RED is based on RE of two 3′ UTR isoforms. The formulas for RE and RED are shown. Genes are divided into three groups based on ΔMLS between 3′ UTR isoforms; i.e., long isoform > short isoform (red line), short isoform > long isoform (blue line), and no significant difference between isoforms (black line). MLSs are from Puro+ cells (Figure 3).

Only group I genes are included. p values (K-S test) for significance of difference (red or blue genes versus black genes) are indicated. Gene numbers are also indicated.

Table 1.

Top BPs associated with TiERA potentials of group I and group II gene transcripts

TiERA potential[a]		GO Term	p Value
Group I genes	high	regulation of Rho protein signal transduction	9.9E–09
		regulation of MAPK cascade	1.3E–07
		regulation of GTPase activity	1.9E–07
		regulation of cell-matrix adhesion	4.0E–07
		head development	6.0E–07
	low	nucleobase-containing small molecule metabolic process	1.0E–07
		amide biosynthetic process	2.2E–07
		drug metabolic process	1.5E–06
		ribonucleoprotein complex biogenesis	2.5E–06
		mitochondrion organization	3.2E–05
Group II genes	high	enzyme linked receptor protein signaling pathway	3.6E–06
		semaphorin-plexin signaling pathway	7.7E–06
		regulation of cell size	8.8E–06
		negative regulation of chemotaxis	3.6E–05
		neuron projection guidance	5.5E–05
	low	protein localization to endoplasmic reticulum	6.9E–04
		ER to Golgi vesicle-mediated transport	1.5E–03
		negative regulation of hydrolase activity	3.8E–03
		regulation of cell killing	8.1E–03
		unsaturated fatty acid biosynthetic process	1.4E–02

[a]High and low TiERA potentials are the top 20% and bottom 20% of MLS, respectively, of group I or group II transcripts in Puro+ cells.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-GRP94	Santa Cruz Biotechnology	sc-393402; RRID:AB_2892568
Mouse anti-α-Tubulin	Lab made	N/A
Mouse anti-Histone H2A.X	Santa Cruz Biotechnology	sc-54606; RRID:AB_2114998
Rabbit anti-Histone H2A	Cell Signaling Technology	CST12349; RRID:AB_2687875
Rabbit anti-Calnexin	Santa Cruz Biotechnology	sc-6465-R; RRID:AB_1119918
Mouse anti-α-Tubulin	SIGMA	T5168; RRID:AB_477579
Chemicals, peptides, and recombinant proteins		
Phosphate buffered saline (PBS)	Lab made	N/A
Dulbecco's Modified Eagle Medium (DMEM)	Lab made	N/A
Bovine Calf Serum (CS)	Sigma-Aldrich	12133C-500ML
Horse serum	Sigma-Aldrich	H1138-500ML
Penicillin/Streptomycin	GIBCO	15140-122
Corning 0.05% Trypsin/0.53mM EDTA in HBSS w/o Calcium, Magnesium or Sodium Bicarbonate	Thermo Fisher Scientific	25052CI
Puromycin	Sigma Aldrich	P8833-10MG
Puromycin	GIBCO	2078745
Harringtonine	Abcam	ab141941
Cyclohexamine	Sigma-Aldrich	01810-1G
TRIzol® Reagent	Thermo Fisher Scientific	15596018
SIGMAFAST protease inhibitor cocktail	Sigma-Aldrich	S8820
Oligo $d(T)_{25}$ Magnetic Beads	NEB	S1419S
Adenosine 5′-Triphosphate (ATP)	NEB	P0756S
SUPERase● In RNase Inhibitor (20 U/μL)	Thermo Fisher Scientific	AM2694
NEBNext® RNase III RNA Fragmentation Module	NEB	E6146
T4 RNA Ligase 1	NEB	M0204S
T4 RNA Ligase 2, truncated KQ	NEB	M0373S
TURBO DNA-free Kit	Thermo Fisher Scientific	AM1907
M-MLV Reverse transcriptase	Promega	M1705
Phusion® High-Fidelity DNA Polymerase	NEB	M0530S
Q5® High-Fidelity DNA Polymerase	NEB	M0491S
Dynabeads MyOne Streptavidin C1	Thermo Fisher Scientific	65001
RNase H	Epicenter	R52250
T4 Polynucleotide Kinase	NEB	M0201S
Shrimp Alkaline Phosphatase (rSAP)	NEB	M0371S

REAGENT or RESOURCE	SOURCE	IDENTIFIER
dNTP (25 mM each) ≥ 99% HPLC	GenScript	C01581
AMPure XP beads	Beckman Coulter	A63881
PVDF membrane	Thermo Fisher Scientific	88518
Clarity Western ECL Substrate	Bio-Rad	1705060
Nytran N Nylon Blotting Membrane, 0.45 μm, 30 cm × 3 m, roll	GE Healthcare Life Sciences	10416196
BamHI	NEB	R0136S
Nylon Membranes, positively charged	Roche	11417240001
Lipofectamine LTX Reagent with PLUS Reagent	Invitrogen	1122890
Oligonucleotides		
5′ adaptor: 5′-CCUUGGCACCCGAGAAUUCCANNNN	Sigma-Aldrich	N/A
DNA/LNA oligo: biotin-T_{15}-$(+TT)_5$, where “+T” denotes LNA	Exiqon	N/A
5′ adenylated 3′ blocked 3′ adaptor with degenerated nucleotides: 5′-rApp/ NNNNGATCGTCGGACTGTAGAACTCTGAAC/3ddC	Bioo Scientific	N/A
Reverse transcription primer: 5′-GTTCAGAGTTCTACAGTCCGACGATC	Sigma-Aldrich	N/A
Reverse PCR primer (GX3): 5′-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA	Sigma-Aldrich	N/A
Indexed forward PCR primers (index region in bracket): 5′-CAAGCAGAAGA CGGCATACGAGAT[NNNNNN]GTGACTGGAGTT CCTTGGCACCCGAGAATTCCA	Sigma-Aldrich	N/A
Critical commercial assays		
Agilent RNA 6000 Pico Kit	Agilent Technologies	5067-1513
High Sensitivity DNA analysis kit	Agilent Technologies	5067-4626
High Sensitivity D1000 ScreenTape Assay kit	Agilent Technologies	5067-5584, 5067-5585, 5067-5587
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
NEBNext® Magnesium RNA Fragmentation Module	NEB	E6150S
DC Protein Assay Reagents Package	Bio-Rad	5000116
DIG Northern Starter Kit	Roche	12039672910
Deposited data		
C2C12 myoblast and myotube cell RNA-seq or 3′READS+ (total, fractionated, puromycin-treated, polysome profiling)	Raw data from this study	GEO: GSE162971
C2C12 myoblast & myotube, total cellular RNA (3′READS+)	Wang et al., 2019	GEO: GSE115232
Experimental models: cell lines		
Mouse: C2C12 cells	ATCC	CRL-1658
Recombinant DNA		
Plasmid: psiCHECK2	Promega	C8021
Plasmid: pcDNA3.0	Invitrogen	N/A
Plasmid: psiCHECK2-Nmt1(full 3′UTR)	This study	N/A
Plasmid: pcDNA3-Nmt1(E1-4)	This study	N/A
Plasmid: pcDNA3-hRluc	This study	N/A
Software and algorithms		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cutadapt	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/guide.html
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
DESeq & DEXSeq	Anders and Huber, 2010; Anders et al., 2012	https://bioconductor.org/packages/release/bioc/html/DESeq2.html; https://bioconductor.org/packages/release/bioc/html/DEXSeq.html
STAR (v2.5.2)	Dobin et al., 2013	https://github.com/alexdobin/STAR
GOstats	Falcon and Gentleman, 2007	https://www.bioconductor.org/packages/release/bioc/html/GOstats.html
Scikit-learn	Pedregosa et al., 2011	https://scikit-learn.org/stable/index.html
XGBoost	Chen and Guestrin, 2016	https://xgboost.readthedocs.io/en/latest/
ViennaRNA	Lorenz et al., 2011	https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/index.html
FIJI	Schindelin et al., 2012	https://imagej.net/software/fiji
3′READS analysis code	Li et al., 2015; Zheng et al., 2016	https://github.com/DinghaiZ/3-prime-READS-plus