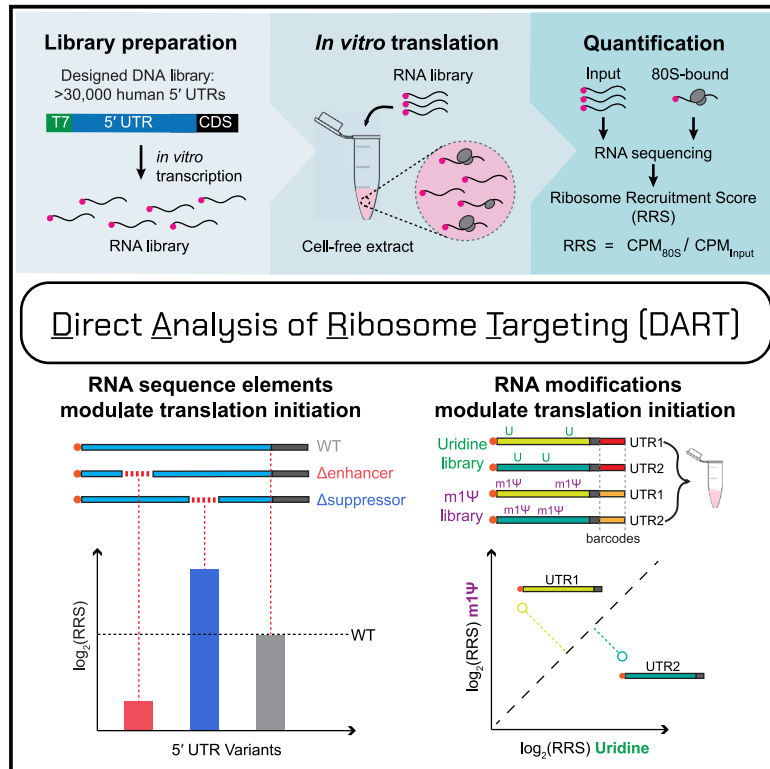**Technology**

Quantitative profiling of human translation initiation reveals elements that potently regulate endogenous and therapeutically modified mRNAs

Graphical abstract

Authors

Cole J.T. Lewis, Li H. Xie (谢力),
Shivani Milind Bhandarkar, ...,
Yixuan Chen (陈逸萱),
Carson C. Thoreen, Wendy V. Gilbert

Correspondence

carson.thoreen@yale.edu (C.C.T.),
wendy.gilbert@yale.edu (W.V.G.)

In brief

Lewis and Xie et al. systematically quantified translation functions of >30,000 human 5′ UTRs, revealing regulatory sequence elements and RNA modifications that impact the initiation process. This approach can aid the design of therapeutic mRNA and facilitate the dissection of the molecular mechanisms of translation initiation.

Highlights

- Measurement of >30,000 human 5′ UTRs reveals a 200-fold range of translation output

- Systematic mutagenesis demonstrates the causality of short regulatory elements

- N1-methylpseudouridine alters translation initiation in a sequence-specific manner

- Optimal modified 5′ UTRs outperform those in the current class of mRNA vaccines

CellPress

Technology

Quantitative profiling of human translation initiation reveals elements that potently regulate endogenous and therapeutically modified mRNAs

Cole J.T. Lewis,[1,3] Li H. Xie (谢力),[1,2,3] Shivani Milind Bhandarkar,[1] Danni Jin (靳丹霓),[1] Kyrillos Abdallah,[1] Austin S. Draycott,[1] Yixuan Chen (陈逸萱),[1] Carson C. Thoreen,[2,*] and Wendy V. Gilbert[1,4,*]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06510, USA
[2]Department of Cellular and Molecular Physiology, Yale School of Medicine, New Haven, CT 06510, USA
[3]These authors contributed equally
[4]Lead contact
*Correspondence: carson.thoreen@yale.edu (C.C.T.), wendy.gilbert@yale.edu (W.V.G.)
https://doi.org/10.1016/j.molcel.2024.11.030

SUMMARY

mRNA therapeutics offer a potentially universal strategy for the efficient development and delivery of therapeutic proteins. Current mRNA vaccines include chemically modified nucleotides to reduce cellular immunogenicity. Here, we develop an efficient, high-throughput method to measure human translation initiation on therapeutically modified as well as endogenous RNAs. Using systems-level biochemistry, we quantify ribosome recruitment to tens of thousands of human 5′ untranslated regions (UTRs) including alternative isoforms and identify sequences that mediate 200-fold effects. We observe widespread effects of coding sequences on translation initiation and identify small regulatory elements of 3–6 nucleotides that are sufficient to potently affect translational output. Incorporation of N1-methylpseudouridine (m1Ψ) selectively enhances translation by specific 5′ UTRs that we demonstrate surpass those of current mRNA vaccines. Our approach is broadly applicable to dissecting mechanisms of human translation initiation and engineering more potent therapeutic mRNAs.

INTRODUCTION

Genetic medicines are a potentially transformative class of therapeutics with diverse applications, including vaccines, immunotherapies, and treatments for genetic disorders. Over the last few years, synthetic mRNAs have emerged as a front-runner among genetic medicine technologies, largely due to the overwhelming success of mRNA-based COVID-19 vaccines. mRNA-based therapeutics are relatively simple to design and produce, can rapidly induce therapeutic protein production, and act transiently without modifying cellular DNA. However, the broader use of mRNA therapeutics is currently limited by the amount of protein that can be produced.

Translation initiation is a rate-limiting process for protein synthesis in human cells. For most messages, translation initiation requires the concerted action of many eukaryotic initiation factors (eIFs) on an mRNA with a 5′ m7G cap.[1] The cap structure is recognized by a complex containing the cap-binding protein eIF4E, the DEAD-box RNA-dependent ATPase eIF4A, and the large scaffold protein eIF4G. In higher eukaryotes, eIF4G mediates ribosome recruitment by binding directly to the eIF3 subunit of 43S complexes that consist of a 40S small ribosomal subunit bound to eIF3, a ternary complex of eIF2·GTP·Met-tRNAi, and

additional factors. The assembled complex scans the 5′ untranslated region (UTR) from 5′ to 3′ to find the appropriate start codon, at which point the 60S large ribosomal subunit joins in a reaction requiring several additional factors. Increasing the rate of ribosome recruitment could substantially increase the potency of mRNA vaccines.

In endogenous mRNAs, translation-enhancing features are generally found in the 5′ untranslated region of mRNAs. Remarkably, changes in 5′ UTR sequences can vary the translation output of an mRNA more than 1,000-fold.[2,3] Some features that distinguish efficiently translated mRNAs include the presence of an m7G cap in an unstructured context at the 5′ end, an AUG initiation codon in a preferred Kozak sequence, shorter 5′ UTRs, lower GC content, and the absence of upstream initiation codons.[4,5] Beyond these minimal attributes, our understanding of the RNA-encoded elements that determine the translation output of mRNAs remains limited. Therefore, translational enhancers cannot yet be designed from first principles.

Naturally occurring translation enhancers may not function in therapeutic mRNAs, which include modified nucleotides to mask the RNA from immune sensors. Unmodified "native" mRNA triggers an intracellular innate immune response through RNA-surveillance mechanisms, including Toll-like receptors

(TLRs), the pattern-recognition receptor RIG-I, and the RNA-dependent kinase PKR.[6–9] Incorporation of modified nucleotides (e.g., N-1-methylpseudouridine) suppresses recognition by TLRs.[10–12] However, modified nucleotides affect RNA-RNA and RNA-protein interactions[13–15] and are therefore likely to disrupt elements that normally promote translation in unmodified mRNAs.

Here, we present direct analysis of ribosome targeting (DART) as a facile approach to quantify translation initiation by more than 35,000 modified 5′ UTRs in human cell extracts. We find that human 5′-UTR-specific ribosome recruitment activity spans over 200-fold and is mediated by non-canonical sequence-dependent mechanisms. Remarkably, the presence of N1-methylpseudouridine affects ribosome recruitment to specific 5′ UTRs by more than 30-fold. DART measurements of ribosome recruitment directly predict protein synthesis from full-length mRNAs, with top-scoring 5′ UTRs outperforming those in the current class of mRNA vaccines. Our results uncover non-canonical regulatory elements that potently affect translation initiation by human 5′ UTRs and establish DART as a powerful approach to engineer optimal 5′ UTRs for therapeutic mRNAs.

DESIGN

Facile, isoform-aware methods to study human translation initiation are currently lacking. Ribosome footprint-based approaches (Ribo-seq,[16] TCP-seq,[17] 40S footprinting,[18] etc.) are powerful methods to quantify ribosome occupancy at the gene level but lack information on the specific transcript isoforms and 5′ UTR sequences that recruited the ribosome. Isoform-specific polysome profiling approaches (TrIP-seq,[2] TL-seq,[19] etc.) and polysome-based massively parallel reporter assays[20–22] (MPRAs) are labor intensive, require large amounts of input material, and do not decouple translation initiation from elongation and mRNA decay. DART, previously developed in budding yeast,[3] appeared to be a promising method to overcome these limitations. However, the original DART protocol required prohibitively large volumes of cell extract and was unable to test multiple conditions in a single experiment. We optimized the DART protocol for use in mammalian systems, reducing hands-on time and decreasing the required input material by two orders of magnitude. We incorporated a barcode-based multiplexing strategy and demonstrated its use by testing different RNA modifications within the same translation reaction and combining multiple reaction conditions onto the same sucrose gradient. These advances massively increase the throughput of DART, enabling researchers to measure the impact of a wide array of genetic, pharmacological, and RNA chemical perturbations on translation initiation. The use of a designed pool of 5′ UTRs in DART further allows direct testing of putative regulatory elements, moving beyond correlation to determine causality.

RESULTS

Development of DART to quantify human 5′-UTR-mediated translational control

Translation initiation culminates in the recruitment of an 80S ribosome positioned at the start codon to begin polypeptide synthesis. 5′ UTRs play a critical role in this process, acting as a platform for binding of initiation factors necessary for ribosome recruitment. However, the features of 5′ UTRs that are responsible for conferring efficient initiation remain largely unknown, and the effects of modified nucleosides (e.g., N1-methylpseudouridine [m1Ψ]) are currently impossible to predict. We therefore sought to develop a high-throughput method to quantify and dissect the effects of 5′ UTR sequences and modifications on translation initiation. DART, which was recently described for measuring synthetic 5′ UTR activity in cell-free translation extracts from budding yeast,[3] appeared to be a promising strategy compatible with testing modified RNAs.

To adapt DART for use in a human system, we began by designing DNA oligonucleotide libraries containing over 35,000 Ensembl-annotated[23] full-length human 5′ UTRs from 14,544 genes. The 5′ UTR sequences range from 10 to 230 nucleotides in length followed by at least 27 nucleotides of coding sequence (CDS) to provide a binding site for the initiating ribosome. Upstream AUGs were removed for simplicity. Oligos include a T7 promoter at the 5′ end for in vitro transcription and a common primer-binding site at the 3′ end for library construction. We transcribed the DNA library and enzymatically added a 5′ methylguanosine cap to produce an RNA pool that reflects endogenous mRNA 5′ ends. RNA pools were incubated in an in vitro translation reaction with HeLa cytoplasmic lysate, which recapitulates cap-stimulated translation over a wide range of mRNA concentrations (Figures S1A and S1B). Translation reactions contained cycloheximide to stabilize recruited ribosomes during sucrose gradient centrifugation, which separated ribosome-bound RNAs from those that failed to recruit a ribosome. Following RNA recovery from the 80S fraction, we prepared Illumina sequencing libraries and calculated a ribosome recruitment score (RRS) as the relative abundance of 80S-bound RNA compared with an input control library (Figure 1A). Human DART reproducibly quantified 5′ UTR activity spanning over a 200-fold range (R^2 = 0.90–0.99; Figures 1B and 1C), highlighting the extensive translational control exerted by human 5′ UTRs.

Systematic testing shows repression by C-rich sequence motifs

We sought to determine 5′ UTR sequence elements that could explain the observed differences in ribosome recruitment. We selected the 100 most active and 100 least active 5′ UTRs from our initial DART analysis and generated a new library in which we systematically deleted 6-nucleotide segments scanning along the full length of each 5′ UTR (Figure 2A). DART analysis on this library of over 6,000 variants of the initial 200 sequences identified hundreds of putative translational enhancer and repressor elements (Figure 2B). An example of a putative translational enhancer within the 5′ UTR of TMSB15B is shown in Figure 2C, where deletion of cap-proximal nucleotides reduced ribosome recruitment more than 4-fold.

To identify common translational regulatory motifs, we quantified the impact of tetramer sequences on ribosome recruitment. Focusing on sequences that were tested at least 30 times across the scanning deletion library, we observed that the removal of C-rich elements significantly enhanced ribosome recruitment (Figure 2D). Accordingly, we noted a striking anticorrelation between cytosine content and ribosome recruitment (Figure 2E)
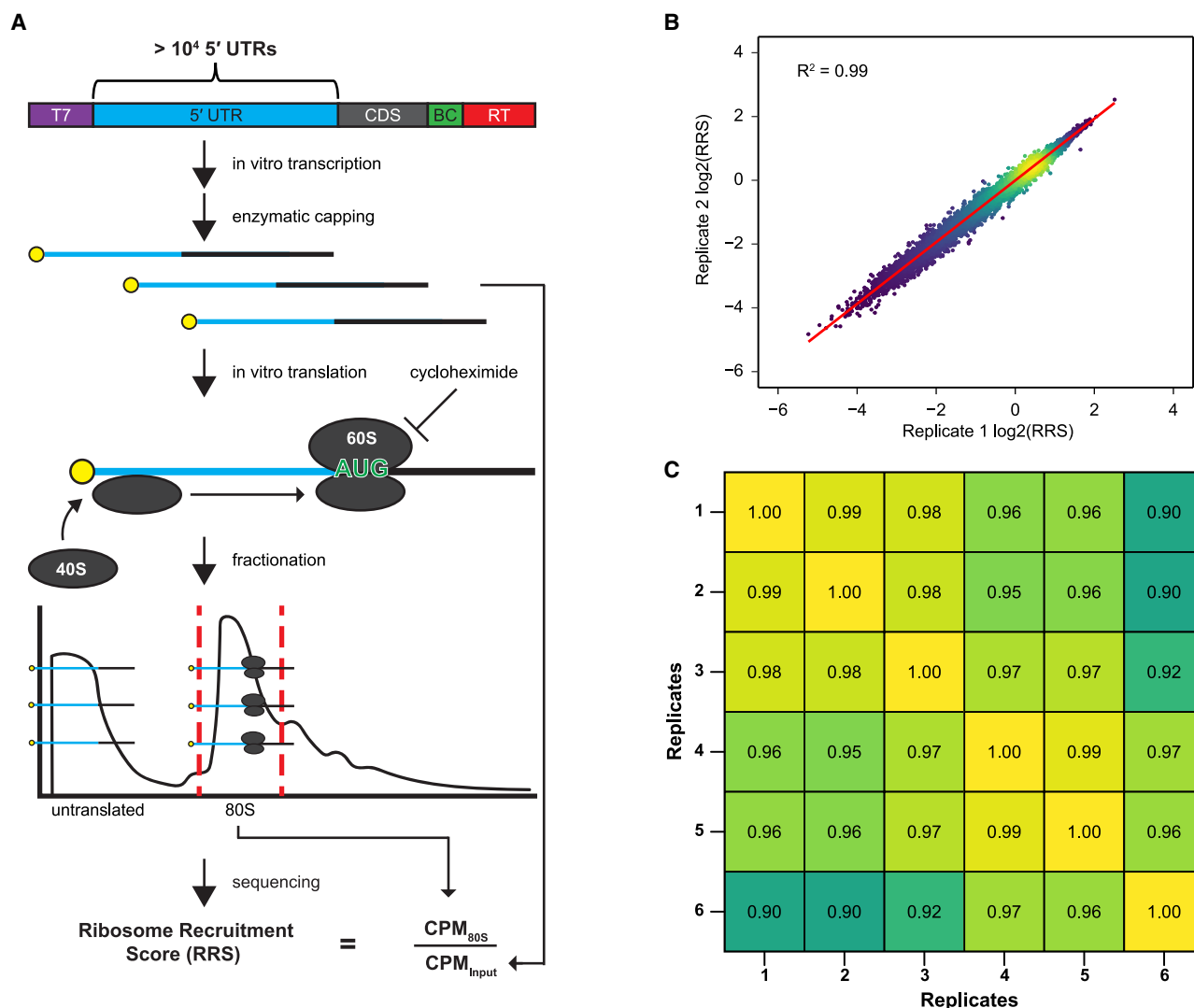
A

B

C

Figure 1. DART quantifies human 5′-UTR-mediated translational control over a 200-fold range
(A) Schematic of the DART workflow. The DNA UTR library sequences contain a T7 promoter, >27 nt of coding sequence, and an RT-binding site for library preparation. Endogenous 5′ UTR sequences were derived from Ensembl annotations.
(B) Human DART reproducibly measures ribosome recruitment over a 200-fold range.
(C) Pearson correlations among six DART replicates.
See also Figure S1.

that was not observed for other nucleotides (Figure S2A). On average, 5′ UTRs with C content below 20% were 5.6-fold more active than 5′ UTRs with greater than 40% C. These results were further corroborated by an unbiased search for enriched motifs within the worst-performing 5′ UTRs. Motif enrichment analysis (DREME,[24] see STAR Methods) of 5′ UTRs in the lowest decile of ribosome recruitment activities identified C-rich motifs (Figure 2F). Interestingly, 3 out of the top 5 enriched motifs contained a CCC trinucleotide element. To determine whether CCC motifs affected ribosome recruitment beyond the overall C content of the 5′ UTRs, we examined 5′ UTRs containing 30%–35% C ($n = 2,295$ UTRs). Within this group of 5′ UTRs, more CCC motifs correlated with less ribosome recruitment (Figure 2G).

Our analysis suggested that CCC elements within 5′ UTRs depress the translational activity of synthetic mRNAs. To test this directly, we generated a pool of 5′ UTR sequences in which up to 5 CCC motifs were iteratively added to 225 endogenous 5′ UTRs and performed DART on these sequences (Figure 2H). We observed a significant and dose-dependent decrease in ribosome recruitment to these 5′ UTRs upon addition of CCC elements (Figures 2I and 2J). We further validated the effect of CCC motifs in repressing translational output of full-length luciferase mRNAs (Figure 2K).

Given that the addition of exogenous CCC motifs was sufficient to repress translation, we tested whether the removal of endogenous CCC elements would conversely increase
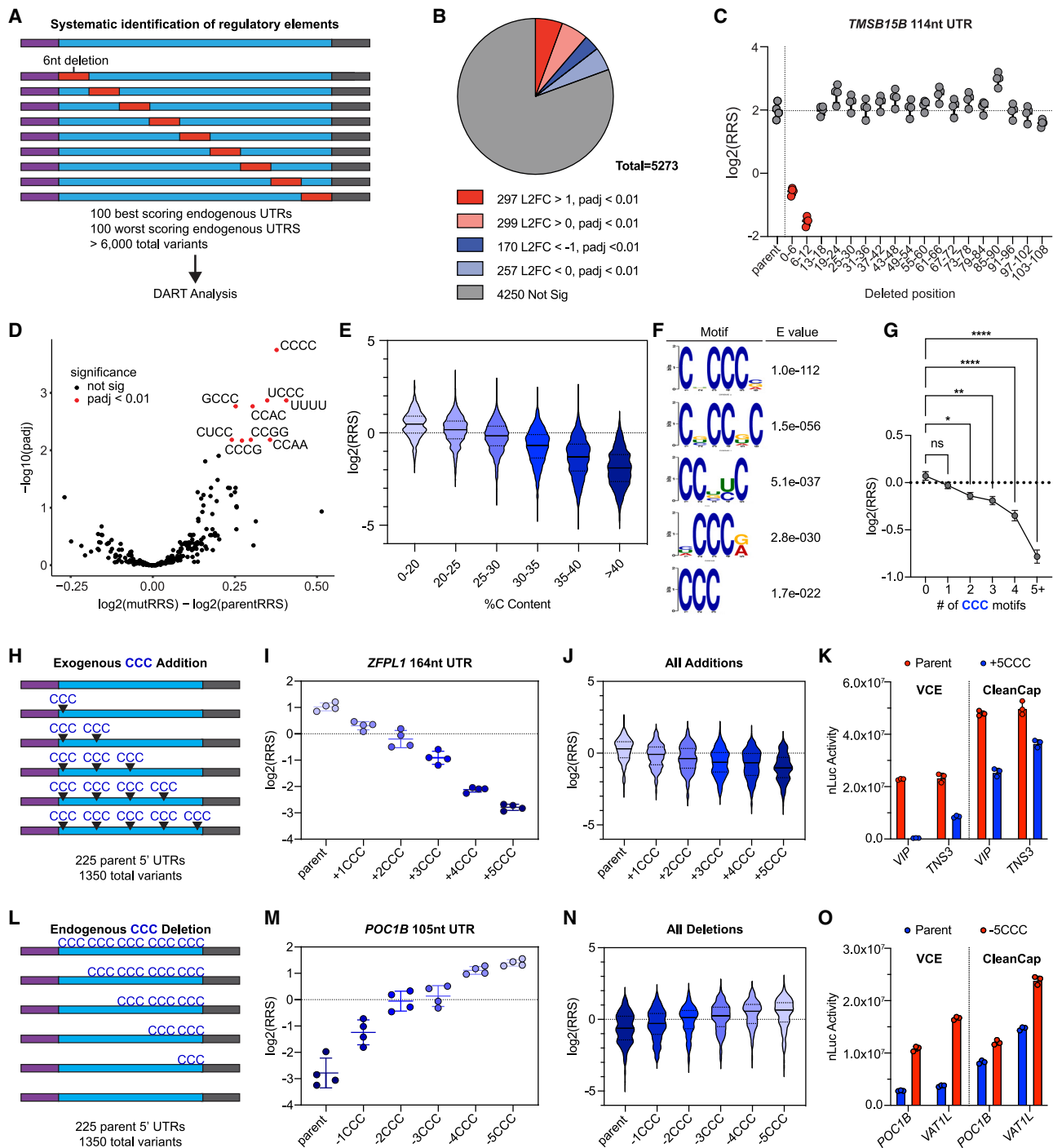
Figure 2. CCC motifs repress vaccinia-capped 5′ UTR activity

(A) Scanning deletion library design for systematic identification of regulatory elements.

(B) Scanning deletion analysis identifies hundreds of hexamers that significantly increase (red) or decrease (blue) RRS ($p$adj < 0.01).

(C) Translational enhancer in the *TMSB15* 5′ UTR. Deletion of nucleotides 1–6 or 7–12 (red) reduces RRS by over 4-fold.

(D) Volcano plot of tetramers deleted ≥30 times covered in the scanning deletion library. Tetramers that significantly altered RRS are highlighted in red ($p$adj < 0.01).

(E) Global trend of reduced RRS with increasing cytidine content.

(F) DREME analysis shows enrichment of C-rich sequence motifs in the bottom 10% of 5′ UTRs by RRS.

translation. Using a similar strategy to the CCC additions, we iteratively deleted up to 5 naturally occurring CCC elements within 225 5′ UTRs and performed DART on these sequences (Figure 2L). Sequential deletion of CCC elements resulted in a dose-dependent increase in ribosome recruitment to these 5′ UTRs (Figures 2M and 2N) and increased protein synthesis in full-length luciferase mRNAs (Figure 2O). These data demonstrate the strength of DART to rapidly iterate through 5′ UTR pools, from unbiased systematic discovery to designed variants that move from correlation to causation.

Pervasive effects of RNA sequence and structure on enzymatic capping

CCC motifs have the potential to base pair with the GGG present at the 5′ end of these RNAs as a part of the optimal T7 RNA polymerase promoter. Vaccinia capping enzyme (VCE) is known to require the 5′ end to be accessible to install an m^7G cap on an RNA substrate.[25] As the addition of a cap increased translation activity by ∼40-fold in these extracts (Figure S1A), variable capping efficiency could significantly impact RRSs. We hypothesized that increased numbers of CCC motifs within 5′ UTRs sequester the 5′ ends into inaccessible secondary structures thereby reducing VCE capping efficiency and decreasing ribosome recruitment. To determine the secondary structure of the 5′ UTR RNA library, we performed chemical probing with dimethyl sulfate (DMS-MaPseq[26]). Briefly, 5′ UTR pools were folded *in vitro*, treated with DMS to probe unpaired A and C residues, and reverse transcribed and sequenced. As expected, DMS treatment specifically increased mutation rates at A and C, and RNA folding led to significant protection compared with denatured controls (Figures S2B and S2C). DMS reactivity was used to constrain 5′ UTR folding *in silico* to determine the pairing probability for regions of interest (STAR Methods, Figure S2D). Consistent with our hypothesis, 5′ UTRs with low activity exhibited significantly more pairing at their 5′ ends which increased with increasing numbers of CCC motifs (Figures S2E–S2G).

To determine if the repressive effect of CCC motifs was due to inhibition of enzymatic RNA capping, we directly measured capping with radiolabeled GTP and found that the addition of CCC motifs significantly reduced capping by VCE (Figure S2H). Accordingly, the addition of 5 CCC elements significantly reduced the stimulatory effect of VCE capping on protein production, whereas the removal of CCC elements enhanced it (Figure S2I). Comparing luciferase reporters with the same 5′ UTR sequences capped two ways, enzymatically (VCE) and co-transcriptionally (CleanCap AG), showed the effects of CCC motifs on translation output were substantially blunted with CleanCap (Figures 2K and 2O). Together, these results establish a widespread inhibitory effect of short CCC motifs on capped mRNAs prepared with VCE, which has implications for research and therapeutic mRNA design.

DART with co-transcriptionally capped 5′ UTRs sensitively detects differences in activity that predict protein output in cells

To determine features of 5′ UTRs that directly impact translation initiation on capped mRNAs, we generated a new library of 51,596 co-transcriptionally capped 5′ UTRs using CleanCap AG, which generates >94% capped RNA,[27,28] and performed DART analysis. We observed an over 200-fold range of RRSs, which were reproducible ($R^2 = 0.83$–0.92; Figure S3A). Ribosome recruitment activities of VCE-capped 5′ UTRs correlated weakly with co-transcriptionally capped ($R^2 = 0.17$; Figure S3B), confirming that VCE sequence bias caused pervasive effects on the apparent activity of 5′ UTRs. The relationship between cap-proximal pairing probability and RRS was substantially weaker for co-transcriptionally capped 5′ UTRs (Figure S3C). The remaining impact of cap-proximal pairing on ribosome recruitment is consistent with a reduced association with the cap-binding complex.[29,30]

Next, we selected 5′ UTRs that spanned the range of RRSs and cloned them upstream of the Firefly luciferase (fluc) coding sequence. These reporters were *in vitro* transcribed with CleanCap AG and transfected into HeLa cells for translation. We observed good agreement between RRSs from the CleanCap DART assay and the amount of protein synthesized ($R^2 = 0.88$; Figure 3A), which was also true when the mRNAs were synthesized in cells from plasmids ($R^2 = 0.72$; Figure S3D). Thus, DART measurements accurately predict ∼200-fold differences in protein synthesis in the context of full-length mRNAs in cells.

DART shares some conceptual similarities with the approach described in Sample et al.,[20] in which *in vitro* transcribed and co-transcriptionally capped mRNA libraries are transfected into cells and fractionated on sucrose gradients. The initiation rates of different 5′ UTRs are compared by calculating the mean ribosome load (MRL) on the mRNAs. We noted that the ∼4-fold dynamic range of ribosome loading (1.89–7.91) observed by this method was mainly contributed by 5′ UTR sequences containing upstream AUGs (uAUGs), and the dispersion of MRL values across 5′ UTR sequences without uAUGs was limited within a 2-fold range. We compared 5,184 endogenous human 5′ UTRs without uAUGs that were analyzed by both methods. We observed that the range of MRL values was compressed

---

(G) The number of CCC trinucleotide motifs correlates with decreased RRS ($n = 2,295$ 5′ UTRs containing 30%–35% cytidine; mean ± 95% CI) *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, one-way ANOVA and Tukey's multiple comparisons test.

(H) Library design testing CCC motif dose dependence using exogenous CCC additions.

(I and J) CCC motifs repress RRS in a dose-dependent manner.

(I) Repressive effect for *ZFPL1* ($n = 4$, mean ± 95% CI) and (J) across all 5′ UTRs tested ($n = 225$ parent 5′ UTRs, 1,350 total variants).

(K) Adding CCC motifs represses translation, with a larger effect on vaccinia-capped than co-transcriptionally capped mRNAs (mean ± SD).

(L) Library design to test CCC motif dose dependence by deleting endogenous CCC motifs.

(M and N) Removal of CCC motifs from 5′ UTRs increases RRS. The effect of CCC removal for the *POC1B* 5′ UTR ($n = 4$, mean ± 95% CI) (M) and all UTRs tested (N, $n = 225$ parent 5′ UTRs, 1,350 total variants).

(O) Deleting CCC motifs within 5′ UTRs increases translation, with a larger effect on vaccinia-capped than co-transcriptionally capped mRNA (mean ± SD).
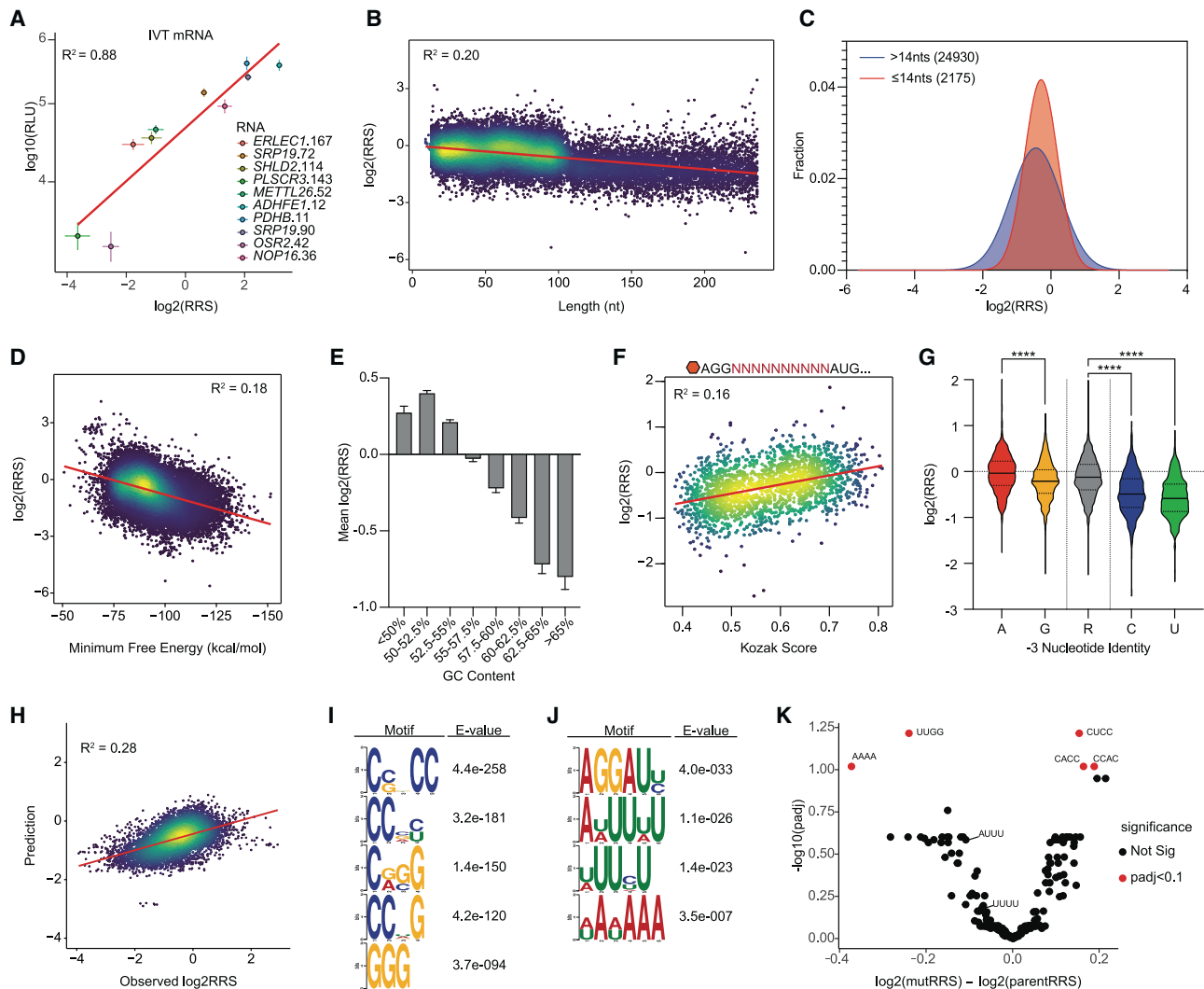
See also Figure S2.

Figure 3. Known regulatory elements explain little of the observed variation in 5′ UTR-specific translation activity

(A) DART scores predict translational output over a 200-fold range. Correlation of RRS with protein production from transfected *in vitro* transcribed mRNAs (mean ± SD).

(B) UTR length modestly negatively correlates with RRS.

(C) Histogram of $\log_2$(RRS) scores for 5′ UTRs shorter than 15 nucleotides (red) or greater than 15 nucleotides (blue).

(D) Predicted minimum free energy modestly negatively correlates with RRS.

(E) Impact of GC content on RRS (mean ± 95% CI).

(F) Kozak sequence strength promotes ribosome recruitment to random 10-nucleotide 5′ UTRs. UTRs scored by conformity to the consensus human Kozak sequence.

(G) Pyrimidines at the −3 position correlate with decreased RRS. ****$p < 0.0001$, unpaired Welch's t test.

(H) A linear regression model incorporating length, predicted minimum free energy, GC content, and Kozak strength explains only 28% of the variance in DART.

(I and J) Sequence motifs enriched in (I) worst 10% and (J) best 10% of 5′ UTRs by RRS.

(K) Volcano plot of tetramers deleted ≥30 times covered in the scanning deletion library. Tetramers that significantly altered RRS are highlighted in red.

See also Figure S3.

compared with RRS (Figure S3E). We hypothesized that MRL is relatively insensitive to translation activity differences between moderately active and very active 5′ UTRs due to the limited resolution of heavy polysomes. To test this, we selected eight 5′ UTRs that exhibited similar MRL and nearly 10-fold differences in RRS and measured protein production based on luciferase reporter activity. Translation in cells showed ∼10-fold differences

in protein synthesis, a range of activity that was better captured by DART (Figure S3F). This result highlights the superior resolution of DART for quantifying translation functions of 5′ UTRs without uAUGs, which includes more than 2/3 of endogenous expressed human mRNAs. Overall, human DART enables sensitive high-throughput quantification across a broad spectrum of 5′ UTR activity.
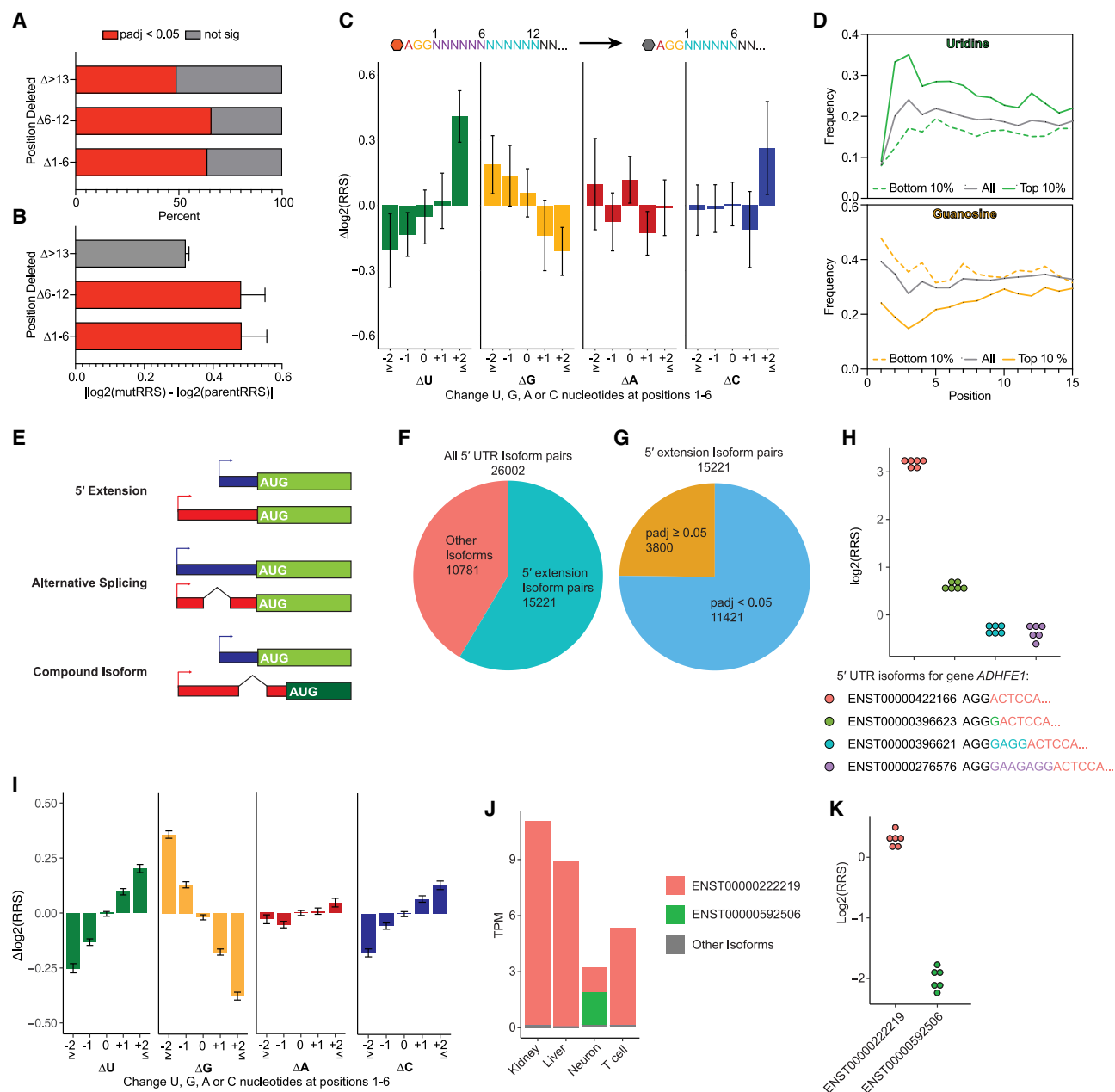
Figure 4. Cap-proximal nucleotide composition affects ribosome recruitment

(A and B) (A) Deletions of nucleotides 1–6 or 7–12 more often caused a significant change (*p*adj < 0.01) and (B) caused larger magnitudes of change in RRS (mean ± 95% CI).

(C) Gain of uridines and loss of guanosines within the first 6 nucleotides increases RRS in a dose-dependent manner. 5′ UTRs were binned based on the change in the number of each nucleotide. Change in RRS relative to parent sequences is plotted on the y axis (mean ± SEM).

(D) Cap-proximal uridines or guanosines are enriched in highly or minimally active 5′ UTRs, respectively. UTRs binned as the top 10% (solid lines) or bottom 10% (dashed lines) by RRS. Plot displays the percent of UTRs containing uridine (top) or guanosine (bottom) at each position.

(E) Illustration of 5′ UTR isoform types.

(F) 5′ extension is the most prevalent isoform type in DART library.

(G) Most 5′ extension isoform pairs differ significantly in RRS.

(H) Example of 5′ extension isoforms of *ADHFE1* that exhibit significantly different RRS. Adding 1–7 nucleotides at the 5′ end is sufficient to alter ribosome recruitment. The changes in the 5′ UTR sequences are labeled below.

(legend continued on next page)

Familiar 5′ UTR determinants exert modest effects on ribosome recruitment with many exceptions

We began our analysis of translational regulatory features within 5′ UTRs by examining general trends in the DART data. We observed a moderate but significant anticorrelation of 5′ UTR length and RRS ($R^2 = -0.20$; Figure 3B). Because a minimum of 12–14 nt are needed to span the distance from the ribosomal P site to the cap,[31,32] it is notable that UTRs less than 14 nt in length recruited ribosomes efficiently (Figure 3C) and promoted high levels of reporter protein synthesis (Figure S3H). RNA-folding stability (Figure 3D) and GC content (Figure 3E) were negatively correlated with ribosome recruitment activity. However, many 5′ UTRs deviated from the global trends. For example, some highly structured 5′ UTRs were among the most efficient ribosome recruiters (Figures S3I and S3J).

As the 43S complex scans along the 5′ UTR, the nucleotides surrounding the AUG start codon, referred to as the Kozak sequence, are thought to play an important role in determining where translation will begin.[33,34] To quantify the impact of the Kozak sequence directly, we generated a library of 2,000 random 10-nucleotide 5′ UTRs and measured their ribosome recruitment activity. We then scored each 10mer UTR based on their similarity to the consensus human Kozak sequence[35] (GCCRCCAUGG, R = purine) and observed a positive correlation between Kozak strength and ribosome recruitment (Figure 3F). Consistent with recognition of a purine at the −3 position,[34,36] 10mers containing a U ($n = 461$) or C ($n = 513$) at this position recruited significantly fewer ribosomes than those with the consensus −3R ($n = 1,011$) (Figure 3G). However, most expressed human 5′ UTRs (84%) contain a purine at −3 indicating that features other than Kozak strength are responsible for determining differences in activity (Figures S3K and S3L).

Overall, most 5′ UTRs showed activities that were not explained by the general trends in length, structure, or Kozak score. A linear regression model incorporating these features was able to predict only 28% of the variability in RRSs measured by DART (Figure 3H). Taken together, these DART results highlight the pervasive contribution of 5′ UTR regulatory elements that remain to be described.

Cap-proximal nucleotide composition affects ribosome recruitment

To identify sequence elements outside the Kozak region that regulate translation, we performed motif enrichment analysis on the least and most active deciles of 5′ UTRs in our library. We observed that C- and G-rich elements were enriched among the worst-performing 5′ UTRs (Figure 3I), indicating that C-rich elements repress translation initiation outside of their negative effect on enzymatic capping. Among the top-performing UTRs, U/A-rich elements were enriched (Figure 3J), suggestive of an enhancing effect. We also systematically correlated the frequency of each trimer in the 5′ UTR sequences with RRS using linear regression.

We found that CC-containing trimers showed negative correlations with translation outcomes in both DART and by MRL using data from Sample et al.[20] (Figure S4A). Additionally, we analyzed ribosome profiling data from Philippe et al.[37] and observed a weak negative trend of translation efficiency (TE) with CC-containing trimers. The TE from ribosome profiling is expected to be less correlated with the other two direct measurements of initiation, as the accumulation of ribosome-protected fragments is affected by both the initiation and elongation rates.

To directly test the impact of short sequence elements on initiation, we repeated DART analysis on the systematic deletion library (Figure 2A) and noted that deletion of individual C-rich tetramers had a significant enhancing effect on ribosome recruitment (Figure 3K). Deletions of cap-proximal nucleotides (Δ1–6 or Δ7–12) were more likely to significantly alter ribosome recruitment (Figure 4A) and caused a larger magnitude of change (Figure 4B) than deletions in the remainder of the UTR sequence (Δ>13). Deletion of the 5′-most 6 nucleotides results in replacement of the cap-adjacent nucleotides, allowing us to directly assess the impact of changing nucleotide composition in this region. We observed that gain of Us and loss of Gs at positions 1 through 6 significantly increased or decreased ribosome recruitment, respectively, in a dose-dependent manner (Figure 4C). Accordingly, across the entire 5′ UTR library, top-performing 5′ UTRs contained more cap-proximal Us and fewer Gs than average, whereas the opposite was true of 5′ UTRs with low activity (Figure 4D).

Next, we examined natural alternative mRNA isoforms to determine whether differences in cap-proximal sequence affect ribosome recruitment by endogenous 5′ UTRs. Alternative 5′ UTR isoforms are frequently generated by alternative transcription start sites and alternative splicing (Figure 4E). We compared the activity of more than 26,000 isoform pairs, which included more than 15,000 pairs that differed only in the cap-proximal sequence due to a 5′ extension in one isoform (Figure 4F). Notably, 75% of 5′ extensions significantly affected ribosome recruitment (Figure 4G), and additions of 1–7 nt were sufficient to cause greater than 4-fold effects (Figure 4H). A weak global correlation between the change in 5′ UTR length and change in RRS ($R^2 = 0.11$; Figure S4B) was driven by a small subset of isoform pairs (2,514/15,221) that differed by more than 50 nt ($R^2 = 0.02$ for 5′ extensions less than 50 nt; Figure S4C). Similar to the artificial deletion constructs, gain of Us and loss of Gs at positions 1 through 6 in natural 5′ UTR isoforms significantly affected ribosome recruitment in a dose-dependent manner (Figure 4I). Surprisingly, gain of Cs in this cap-proximal region also increased RRS nearly as well as Us, suggesting that 5′ pyrimidines more generally may enhance initiation. In some cases, isoforms that were poorly translated in HeLa extracts were preferentially expressed in specific tissues, raising the possibility that some 5′ UTRs are optimized for translation in specific cellular contexts (Figures 4J and 4K). Overall, these analyses

---

(I) Gain of uridines and loss of guanosines in the first 6 nucleotides between 5′ extension pairs increases RRS in a dose-dependent manner (mean ± SEM).

(J and K) An isoform of *DNASE2* that is poorly translated in HeLa extracts is preferentially expressed in neurons.

(J) The RNA expression level for each transcript isoform of gene *DNASE2* in kidney, liver, neuron, and T cells.

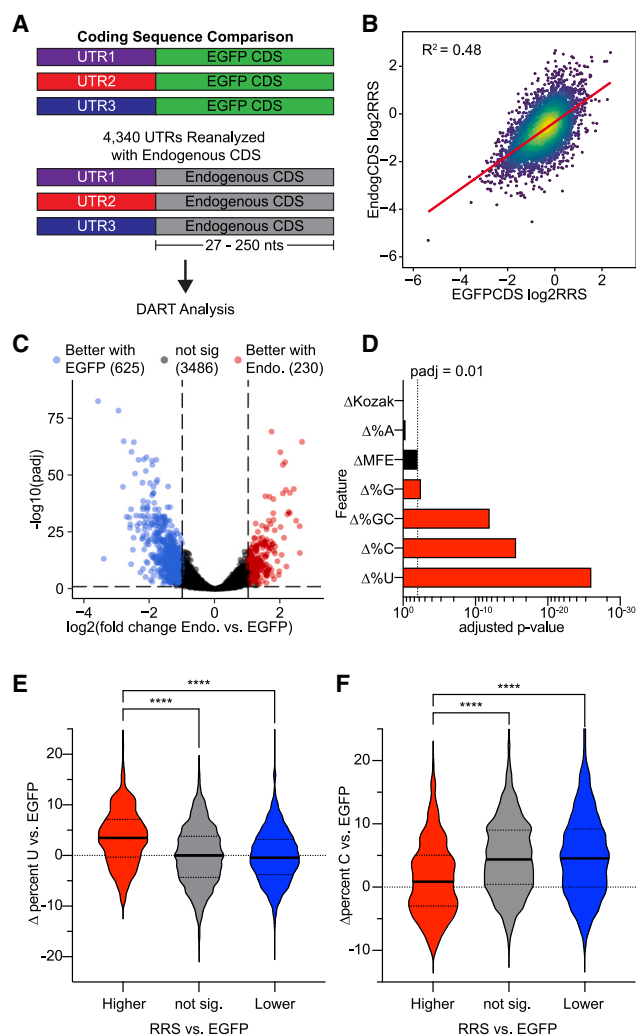(K) RRS scores for two major isoforms of *DNASE2*.

See also Figure S4.

Figure 5. Coding sequence significantly affects ribosome recruitment to many 5′ UTRs

(A) Library design to test the impact of coding sequences on ribosome recruitment.

(B) Coding sequence significantly affects ribosome recruitment. Each 5′ UTR is plotted according to its RRS with EGFP and endogenous coding sequence.

(C) The impact of coding sequence on RRS is context dependent. 5′ UTRs with significantly altered RRS with endogenous compared with EGFP coding sequence (fold change > 2, *p*adj < 0.01) are color labeled.

(D) Changes in uridine, cytidine, guanosine, and GC content within the coding sequence correlate with altered RRS. *p* values determined by Wilcoxon rank test and Benjamini-Hochberg adjustment.

(E and F) Nucleotide content in the coding sequence influences recruitment. (E) Uridine and (F) cytosine. ****p < 0.0001 unpaired Welch's t test.

See also Figure S5.

suggest that short sequences adjacent to the cap exert an outsized influence on the efficiency of ribosome recruitment.

Coding sequences significantly affect ribosome recruitment

The coding region of therapeutic mRNA is largely dictated by the desired protein product, with some room to optimize synony-

mous codons. By contrast, the 5′ UTR can be changed to suit the payload. Given that the scanning 48S complex contacts 6–24 nts of CDS,[17] we wondered whether 5′ UTRs are equally active when paired with different coding sequences. We tested 4,340 5′ UTRs in two contexts: upstream of the enhanced green fluorescent protein (EGFP) coding sequence or the endogenously occurring CDS for each 5′ UTR (Figure 5A). We noted that changing the coding sequence caused substantial differences in ribosome recruitment when comparing identical 5′ UTRs with EGFP or endogenous coding sequence (R^2 = 0.48, Figure 5B). Among the 5′ UTRs tested with both coding sequences, nearly 20% (855) exhibited over a 2-fold change in ribosome recruitment activity (Figure 5C). We observed similar large effects on initiation upon replacement of EGFP with Fluc or SARS-CoV2 spike protein coding sequences (Figure S5A). Among this set, a relatively small percentage of 5′ UTRs maintained similarly high (top 10%) or low (bottom 10%) RRS scores across these different CDSs (Figure S5B). These data indicate that identifying the most optimal 5′ UTR sequence is likely to require screening with the desired coding sequence.

We next analyzed the impact of CDS features on ribosome recruitment and found that nucleotide content within the 5′ region of the coding sequence was significantly correlated with the changes in translation activity (Figure 5D). Specifically, coding sequences that outperformed their EGFP counterparts contained more Us, fewer Cs and Gs, and lower GC content (Figures 5E, 5F, S5C, and S5D), which mirrors the trends in 5′ UTRs (Figures 3I, 3J, and S4A). These differences were not correlated with changes in structure surrounding the start codon (Figures S5E and S5F). It will be interesting to see whether the mechanisms underlying the repressive effects of C-rich and activating effects of U-rich elements are the same when located in CDS and the 5′ UTR. Engineering the 5′ end of coding sequences according to these principles may therefore be a mechanism to increase the protein production of therapeutic mRNAs beyond maximizing codon optimality (see discussion).

DART can be miniaturized and multiplexed for higher throughput with less input

Our results establish DART as a facile method for quantitative analysis of *cis*-regulatory elements in human 5′ UTRs. In principle, DART could dissect 5′ UTR-specific responses to trans factor manipulations including knockdowns, transfections, and small molecule treatments. However, the initial DART conditions used 250 μL of cell extract in 500 μL reactions and required a separate centrifuge bucket for each sample, limiting throughput. We therefore tested DART performance with decreasing input material ranging from 100 μL of cell extract (~10 million cells) to 10 μL (~1 million cells). Barcodes were added prior to pooling reactions for centrifugation (Figure S6A). The smallest scale (10 μL) maintained 87%–89% coverage across the 5′ UTR pool (38,191 sequences) (Figure 6A) and reproduced the results from larger reactions (Figures 6B and S6B). RRSs from miniaturized DART reactions accurately predicted translation output from luciferase reporters (Figure 6C). Thus, DART can be miniaturized to work with limited starting material. Using smaller reaction volumes along with the barcoding strategy also allows pooling multiple samples onto a single sucrose gradient, easily increasing throughput by 10-fold.
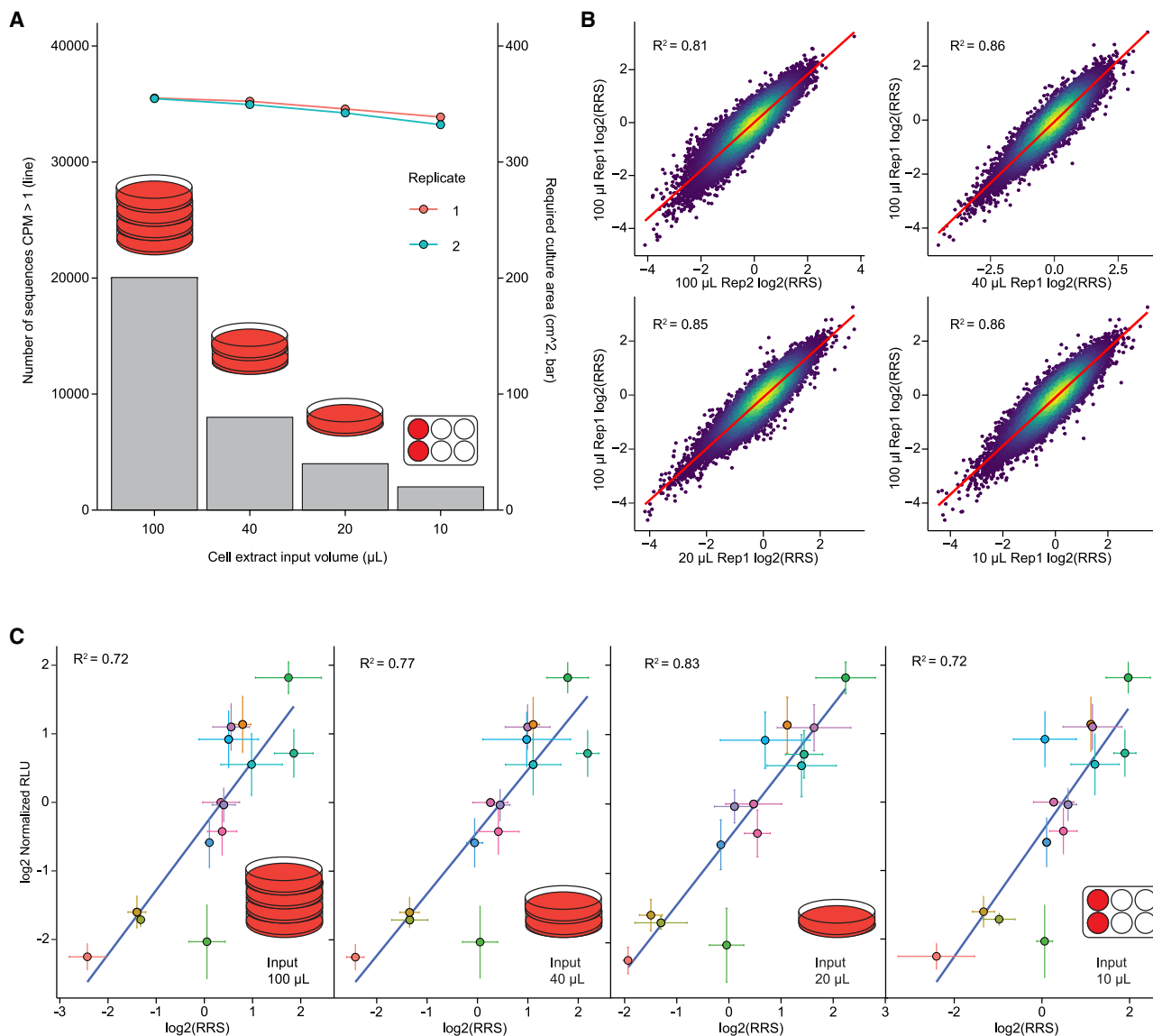
Figure 6. DART assay can be miniaturized to save input material and increase throughput

(A) DART can be scaled down 25-fold without losing sequence representation. Cell lysate input is shown as a bar plot (right y axis), and the number of sequences > 1 CPM is shown as a line plot (left y axis). Plates indicate the cell culture area required for each lysate volume.

(B) DART reproducibility is preserved across volumes. RRS from each input volume are plotted against each other.
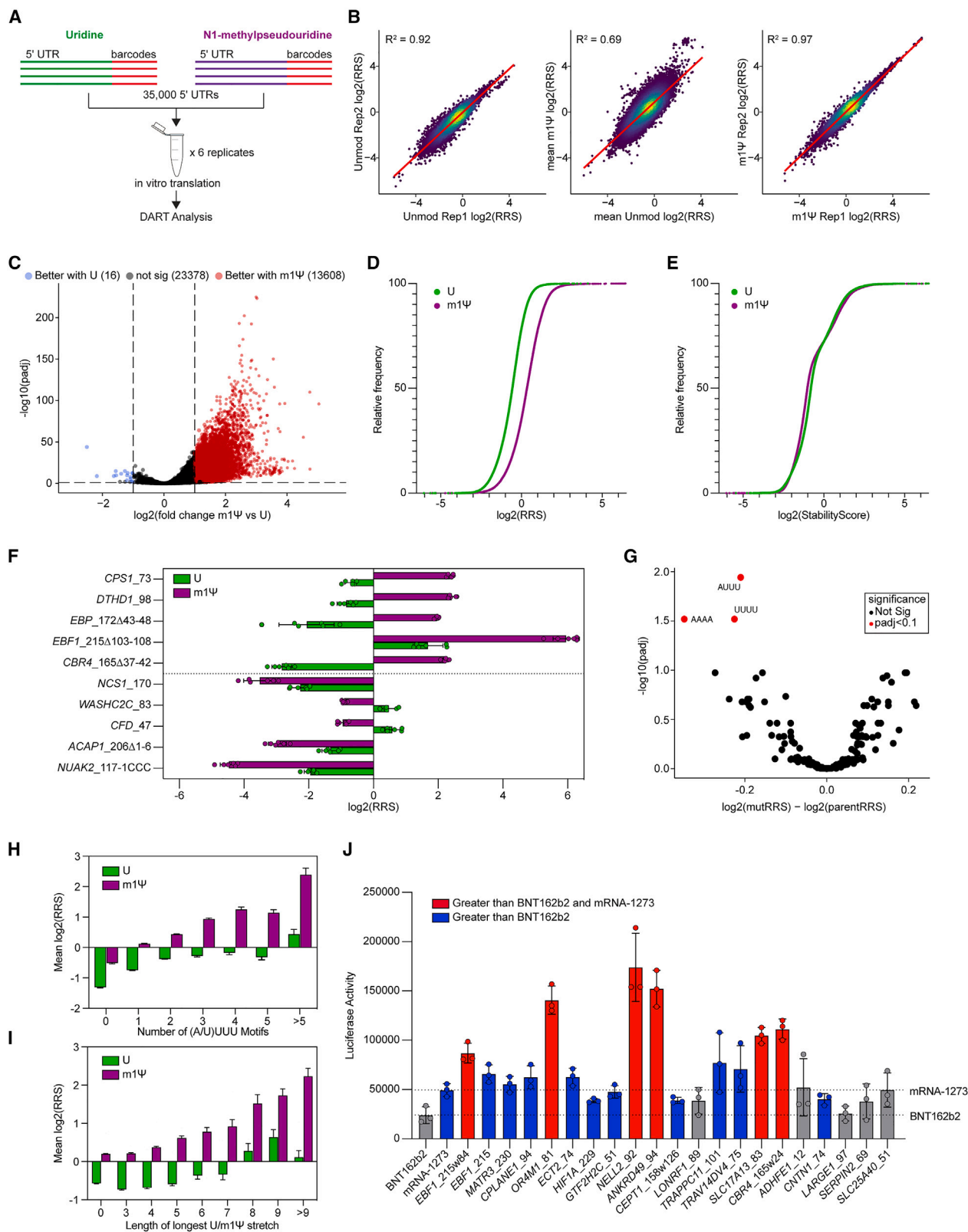
(C) RRS from miniaturized DART predict 5′-UTR-driven translational activity in luciferase reporter mRNAs (mean ± SD).

See also Figure S6.

m1Ψ incorporation into 5′ UTRs significantly alters ribosome recruitment

Current-generation mRNA therapeutics are chemically substituted with m1Ψ in place of uridine to avoid the innate immune response that would otherwise depress protein production *in vivo*.[10,38] Beyond this global effect, it was unclear how m1Ψ within 5′ UTRs would impact translation initiation. We tested this systematically by performing DART on 5′ UTR sequences with full replacement of uridine with m1Ψ. Importantly, uridine and m1Ψ-containing RNAs were co-incubated in the same translation reaction to allow the identification of sequence-dependent direct effects of

m1Ψ substitution separate from any global effects (Figure 7A). Two sets of 4 barcodes were used to distinguish RNAs with uridine or m1Ψ. The barcoded m1Ψ and U libraries had similarly high representation of sequences (Figures S3G and S7A). By contrast to the low inter-replicate variability of ribosome recruitment activity by uridine- and m1Ψ-containing 5′ UTRs (R^2 = 0.83–0.96), we observed large differences in RRS scores when comparing identical sequences with or without m1Ψ substitution (R^2 = 0.69; Figure 7B). The impact of general features (length, secondary structure, GC content, Kozak strength, and the impact of cap-proximal nucleotides) was similar in m1Ψ-substituted 5′ UTRs

(Figures S7A–S7K). Surprisingly, we observed a global stimulatory effect of m1Ψ substitution on ribosome recruitment (1.9-fold on average), with over 36% of 5′ UTRs (*n* = 13,608) exhibiting over a 2-fold increase in initiation activity compared with unmodified sequences (Figures 7C and 7D). We confirmed that m1Ψ incorporation did not significantly alter the stability of the RNA pool during the translation reaction (Figure 7E), demonstrating that m1Ψ directly enhances ribosome recruitment.

Beyond the global effect, the degree to which m1Ψ incorporation altered ribosome recruitment was highly sequence specific with individual sequences ranging from 32-fold stimulation to 5-fold repression (Figure 7F). To determine the sequence elements responsible for the effect of m1Ψ substitution on 5′ UTR activity, we performed DART analysis on the scanning deletion library (Figure 2A) with and without m1Ψ substitution. This analysis revealed that deletion of (A/X)XXX (X=m1Ψ) motifs significantly reduced ribosome recruitment activity (Figure 7G). Notably, deletion of (A/U)UUU motifs did not have a significant impact on unmodified 5′ UTRs (Figure 3K), indicating that these sequence elements are stimulatory only when modified to m1Ψ. Accordingly, across all sequences tested with m1Ψ, increasing numbers of (A/X)XXX motifs correlated with increased ribosome recruitment (Figure 7H). As the (A/X)XXX elements contain consecutive m1Ψ nucleotides, we were curious whether longer stretches of poly(m1Ψ) affected ribosome recruitment. We noted a significant increase in ribosome recruitment to 5′ UTRs containing longer stretches of poly(m1Ψ) (Figure 7I). By contrast, the length of unmodified poly(U) stretches had little effect.

Finally, we sought to validate candidate 5′ UTRs that could drive higher levels of protein synthesis than the 5′ UTRs in current mRNA vaccines. We selected 21 of the top-performing 5′ UTRs from the m1Ψ DART screen, cloned them in front of luciferase reporters, and transfected these reporter mRNAs into HeLa cells. 16/21 and 6/21 outperformed the 5′ UTR sequences from the FDA-approved SARS-CoV-2 vaccines BNT162b2 (Pfizer-BioNTech) and mRNA-1273 (Moderna), respectively (Figure 7J), demonstrating the utility of the DART method to identify highly active sequences for mRNA therapeutic design.

DISCUSSION

Here, we present the development of DART as a high-throughput method to quantify the translational activity of 5′ UTRs in a human system. A major advantage of the DART assay is the ease

with which 5′ UTR libraries can be designed and analyzed. Regulatory features identified in an initial screen can be subsequently interrogated directly, moving from correlation to causation. We use this approach to dissect intrinsic determinants of large differences in translation initiation efficiency, quantifying the impact of known regulatory features, identifying the importance of features whose mechanisms remain to be uncovered, and establishing design principles for therapeutic mRNA development.

A key conclusion from our study is that features of 5′ UTRs typically associated with translation function—length, structure, and AUG sequence context—are relatively minor contributors to ribosome recruitment. Although we confirm that these features influence initiation, they account for only ~28% of the variance across the 200-fold range of differences we observe. We identified other features that also significantly affect ribosome recruitment, including cap-proximal sequences, C-rich repressive motifs, and A/U-rich enhancing motifs. These findings suggest complex and underexplored interactions with the translation initiation machinery that are important contributors to protein output. A second observation at odds with classical models of initiation is the surprising efficiency of very short UTRs, many of which are only 10 nt. This observation is consistent with observations that eukaryotic ribosomes can initiate even on mRNAs that begin with AUG and therefore lack a 5′ UTR altogether.[39] For initiation to occur, these 5′ UTRs must somehow be released from eIF4F to access the ribosome P site. Whether these short UTR mRNAs utilize a common or specialized mode of initiation is still unclear, but further analysis of their sequence requirements may shed light on the underlying mechanism.

Our findings also inform the design of therapeutic mRNAs. Optimization of therapeutic mRNAs typically includes uridine depletion and modification of the remaining uridines to m1Ψ—both aimed at reducing activation of the cellular innate immune response.[38,40] Our results demonstrate a more nuanced approach is necessary. By isolating the initiation step of translation, we find that uridines within the cap-proximal region of 5′ UTRs significantly enhance ribosome recruitment. This effect holds when uridines are substituted with m1Ψ. Surprisingly, however, the substitution of uridine with m1Ψ globally increases ribosome recruitment, indicating unique effects of m1Ψ compared with uridine. We observe that stretches of consecutive m1Ψ—but not uridine—residues within 5′ UTRs have a

---

Figure 7. Global and 5′ UTR-specific effects of m1Ψ on ribosome recruitment

(A) Uridine-containing and m1Ψ-containing RNAs were barcoded and combined in the same reaction for direct comparison of activity.

(B) Identical 5′ UTR sequences containing uridine versus m1Ψ (middle) show widespread differences in RRS (*n* = 6) compared to replicate reproducibility with uridine (U, left) or m1Ψ (right).

(C) Volcano plot of RRS differences with m1Ψ substitution. 5′ UTRs exhibiting significantly changes in RRS (fold change > 2, *p*adj < 0.01) are color labeled.

(D and E) Cumulative distribution plots of (D) RRS and (E) stability scores from UTRs containing uridine (green) or m1Ψ (purple).

(F) The effect of m1Ψ on ribosome recruitment is 5′ UTR specific. Individual examples of 5′ UTRs that are strongly stimulated (top) or repressed (bottom) by m1Ψ substitution (*n* = 6, mean ± SD).

(G) Scanning deletion analysis identifies tetramers that significantly alter ribosome recruitment when deleted in m1Ψ-substituted RNAs (right).

(H and I) (H) Increasing numbers of (A/U)UUU motifs and (I) poly(m1Ψ) stretch length correlate with increased ribosome recruitment in m1Ψ-substituted RNAs, (mean ± 95% CI, *p* < 2.2e−16 by two-way ANOVA test considering the interaction between modifications and number/length of U/m1Ψ stretches.

(J) Optimal m1Ψ-substituted mRNAs produce more protein from luciferase reporter mRNAs than current commercial vaccine 5′ UTRs (mean ± SD, *p* < 0.05 two-tailed Student's t test).

See also Figure S7.

substantial enhancing effect on translation activity. Employing a sweeping uridine depletion strategy is therefore likely to remove potent translational enhancers and decrease protein output from therapeutic mRNAs.

In addition to uridine content optimization, our data reveal that cytosine-rich elements can function as potent translational silencers. Notably, CCC elements repress ribosome recruitment while also inhibiting enzymatic mRNA capping by base pairing with 5′ end and rendering it inaccessible to the VCE. Some current-generation mRNA therapeutics, including mRNA-1273, use an enzymatic capping strategy.[41] Our results indicate that sequence choice can impact mRNA production as well as translational activity in cells. Interestingly, we find that the enhancing and repressive effects of uridine and cytosine, respectively, observed in the 5′ UTR also extend to the coding sequence. This may provide another mechanism to optimize the coding sequence in addition to codon optimality. The surprising impact of CDS on 5′ UTR function additionally suggests that optimal therapeutic mRNA design will require screening of UTR-CDS combinations.

Beyond therapeutic mRNA development, DART has broad applications to examine effects on translation initiation in other contexts. Multiple small molecules in pre-clinical or clinical-stage development for cancer treatment, including silvestrol and zotatafin (eIF4A inhibitors), ribavirin and 4Ei-1 (eIF4E inhibitors), and everolimus and BEZ235 (mTOR inhibitors) are designed to inhibit or alter translation initiation.[42,43] The ability to perform DART with minimal input material and multiplex multiple samples on a single sucrose gradient makes it straightforward to perform dose-escalation studies to define drug-sensitive and -insensitive 5′ UTRs. Moreover, cancerous tissues reprogram their transcriptome, producing an abundance of tumor-specific 5′ UTR isoforms.[44,45] Our results demonstrate that relying on simple principles such as 5′ UTR length, structure formation, etc. is insufficient to predict the translational impact of these isoform changes. DART is a facile method to empirically measure the consequences of UTR variation and test the impact of therapeutic agents on the disease-specific transcriptome.

Limitations of the study

The DART method possesses some limitations. First, DART requires translationally active extract from the cell type or tissue of interest. We have successfully produced active extracts from a wide range of cancerous and immortalized mammalian cell lines. However, producing extracts from less translationally active inputs such as primary cells or tissue samples is likely to require optimization. Second, *in vitro* transcription with T7 and co-transcriptional capping with CleanCap AG yields RNAs that begin with 5′-AG-3′ immediately following the methylguanosine cap. Evaluating the impact of cap-proximal nucleotide positions would require changes to the method of synthesizing the RNA pool. Third, the design of the DART method precludes analysis of how upstream AUG sites affect translation initiation, although the repressive function of these elements is well established. Lastly, in this study, we fully replaced uridine with m1Ψ to match the design of therapeutic mRNAs. In future studies, site-specific modification can be achieved by incubating the pool RNA with recombinant modifying enzymes to uncover the function of endogenous 5′ UTR modifications.

AUTHOR CONTRIBUTIONS

C.J.T.L. and W.V.G. conceived the project. C.J.T.L., L.H.X., S.M.B., D.J., and K.A. performed experiments. C.J.T.L., L.H.X., S.M.B., D.J., A.S.D., Y.C., and C.C.T. analyzed high-throughput sequencing data. W.V.G. and C.C.T. supervised the study. C.J.T.L. wrote the first draft of the manuscript. All authors discussed the results and edited the manuscript.

DECLARATION OF INTERESTS

Yale University has filed a patent application based on this work. C.J.T.L., L.H.X., C.C.T., and W.V.G. are named as co-inventors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines
- METHOD DETAILS
  - 5′ UTR Pool Design and Synthesis
  - *In vitro* translation
  - 80S Isolation and RNA extraction
  - DART Library Preparation
  - Data Processing
  - *In vitro* luciferase assays

- ○ Radiolabeled Vaccinia Capping Assays
- ○ In cell luciferase assays with *in vitro* transcribed mRNA
- ○ In cell luciferase assays with plasmid
- ○ Tetramer analysis from scanning mutagenesis
- ○ DMS-MaPseq library preparation
- ○ DMS-MaPseq analysis
- ○ Linear regression model
- ○ Preparation of HEK293T translation extracts
- ○ Analysis of differential ribosome recruitment
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.11.030.

REFERENCES

1. Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb. Perspect. Biol. *4*, a011544. https://doi.org/10.1101/cshperspect.a011544.

2. Floor, S.N., and Doudna, J.A. (2016). Tunable protein synthesis by transcript isoforms in human cells. eLife *5*, e10921. https://doi.org/10.7554/eLife.10921.

3. Niederer, R.O., Rojas-Duran, M.F., Zinshteyn, B., and Gilbert, W.V. (2022). Direct analysis of ribosome targeting illuminates thousand-fold regulation of translation initiation. Cell Syst. *13*, 256–264.e3. https://doi.org/10.1016/j.cels.2021.12.002.

4. Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N. (2016). Translational control by 5′-untranslated regions of eukaryotic mRNAs. Science *352*, 1413–1416. https://doi.org/10.1126/science.aad9868.

5. Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127. https://doi.org/10.1038/nrm2838.

6. Alexopoulou, L., Holt, A.C., Medzhitov, R., and Flavell, R.A. (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. Nature *413*, 732–738. https://doi.org/10.1038/35099560.

7. Rehwinkel, J., and Gack, M.U. (2020). RIG-I-like receptors: their regulation and roles in RNA sensing. Nat. Rev. Immunol. *20*, 537–551. https://doi.org/10.1038/s41577-020-0288-3.

8. Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. (2004). Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. Science *303*, 1526–1529. https://doi.org/10.1126/science.1093620.

9. Robertson, H.D., and Mathews, M.B. (1996). The regulation of the protein kinase PKR by RNA. Biochimie *78*, 909–914. https://doi.org/10.1016/s0300-9084(97)86712-0.

10. Andries, O., Mc Cafferty, S., De Smedt, S.C., Weiss, R., Sanders, N.N., and Kitada, T. (2015). N(1)-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. J. Control. Release *217*, 337–344. https://doi.org/10.1016/j.jconrel.2015.08.051.

11. Karikó, K., Buckstein, M., Ni, H., and Weissman, D. (2005). Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity *23*, 165–175. https://doi.org/10.1016/j.immuni.2005.06.008.

12. Karikó, K., Muramatsu, H., Welsh, F.A., Ludwig, J., Kato, H., Akira, S., and Weissman, D. (2008). Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Mol. Ther. *16*, 1833–1840. https://doi.org/10.1038/mt.2008.200.

13. Lewis, C.J.T., Pan, T., and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA-protein interactions. Nat. Rev. Mol. Cell Biol. *18*, 202–210. https://doi.org/10.1038/nrm.2016.163.

14. Gilbert, W.V., and Nachtergaele, S. (2023). mRNA Regulation by RNA Modifications. Annu. Rev. Biochem. *92*, 175–198. https://doi.org/10.1146/annurev-biochem-052521-035949.

15. Tanzer, A., Hofacker, I.L., and Lorenz, R. (2019). RNA modifications in structure prediction - status quo and future challenges. Methods *156*, 32–39. https://doi.org/10.1016/j.ymeth.2018.10.019.

16. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223. https://doi.org/10.1126/science.1168978.

17. Archer, S.K., Shirokikh, N.E., Beilharz, T.H., and Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. Nature *535*, 570–574. https://doi.org/10.1038/nature18647.

18. Bohlen, J., Fenzl, K., Kramer, G., Bukau, B., and Teleman, A.A. (2020). Selective 40S Footprinting Reveals Cap-Tethered Ribosome Scanning in Human Cells. Mol. Cell *79*, 561–574.e5. https://doi.org/10.1016/j.molcel.2020.06.005.

19. Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res. *23*, 977–987. https://doi.org/10.1101/gr.150342.112.

20. Sample, P.J., Wang, B., Reid, D.W., Presnyak, V., McFadyen, I.J., Morris, D.R., and Seelig, G. (2019). Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. Nat. Biotechnol. *37*, 803–809. https://doi.org/10.1038/s41587-019-0164-5.

21. Castillo-Hair, S., Fedak, S., Wang, B., Linder, J., Havens, K., Certo, M., and Seelig, G. (2024). Optimizing 5′UTRs for mRNA-delivered gene editing using deep learning. Nat. Commun. *15*, 5284. https://doi.org/10.1038/s41467-024-49508-2.

22. Plassmeyer, S.P., Florian, C.P., Kasper, M.J., Chase, R., Mueller, S., Liu, Y., White, K.M., Jungers, C.F., Djuranovic, S.P., Djuranovic, S., and Dougherty, J.D. (2023). A Massively Parallel Screen of 5′UTR Mutations Identifies Variants Impacting Translation and Protein Production in Neurodevelopmental Disorder Genes. Preprint at medRxiv. https://doi.org/10.1101/2023.11.02.23297961.

23. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. Nucleic Acids Res. *50*, D988–D995. https://doi.org/10.1093/nar/gkab1049.

24. Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659. https://doi.org/10.1093/bioinformatics/btr261.

25. Fuchs, A.L., Neu, A., and Sprangers, R. (2016). A general method for rapid and cost-efficient large-scale production of 5′ capped RNA. RNA *22*, 1454–1466. https://doi.org/10.1261/rna.056614.116.

26. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., and Rouskin, S. (2017). DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat. Methods *14*, 75–82. https://doi.org/10.1038/nmeth.4057.

27. Miller, M., Alvizo, O., Baskerville, S., Chintala, A., Chng, C., Dassie, J., Dorigatti, J., Huisman, G., Jenne, S., Kadam, S., et al. (2024). An engineered T7 RNA polymerase for efficient co-transcriptional capping with reduced dsRNA byproducts in mRNA synthesis. Faraday Discuss. *252*, 431–449. https://doi.org/10.1039/d4fd00023d.

28. Henderson, J.M., Ujita, A., Hill, E., Yousif-Rosales, S., Smith, C., Ko, N., McReynolds, T., Cabral, C.R., Escamilla-Powers, J.R., and Houston, M.E. (2021). Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap((R)) Analog by In Vitro Transcription. Curr. Protoc. 1, e39. https://doi.org/10.1002/cpz1.39.

29. Çetin, B., Song, G.J., and O'Leary, S.E. (2020). Heterogeneous Dynamics of Protein-RNA Interactions across Transcriptome-Derived Messenger RNA Populations. J. Am. Chem. Soc. 142, 21249–21253. https://doi.org/10.1021/jacs.0c09841.

30. Çetin, B., and O'Leary, S.E. (2022). mRNA- and factor-driven dynamic variability controls eIF4F-cap recognition for translation initiation. Nucleic Acids Res. 50, 8240–8261. https://doi.org/10.1093/nar/gkac631.

31. Pisarev, A.V., Kolupaeva, V.G., Yusupov, M.M., Hellen, C.U.T., and Pestova, T.V. (2008). Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. EMBO J. 27, 1609–1621. https://doi.org/10.1038/emboj.2008.90.

32. Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B., and Qian, S.B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc. Natl. Acad. Sci. USA 109, E2424–E2432. https://doi.org/10.1073/pnas.1207846109.

33. Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. 83, 779–812. https://doi.org/10.1146/annurev-biochem-060713-035802.

34. Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283–292. https://doi.org/10.1016/0092-8674(86)90762-2.

35. Gleason, A.C., Ghadge, G., Sonobe, Y., and Roos, R.P. (2022). Kozak Similarity Score Algorithm Identifies Alternative Translation Initiation Codons Implicated in Cancers. Int. J. Mol. Sci. 23, 10564. https://doi.org/10.3390/ijms231810564.

36. Hernández, G., García, A., Weingarten-Gabbay, S., Mishra, R.K., Hussain, T., Amiri, M., Moreno-Hagelsieb, G., Montiel-Dávalos, A., Lasko, P., and Sonenberg, N. (2024). Functional analysis of the AUG initiator codon context reveals novel conserved sequences that disfavor mRNA translation in eukaryotes. Nucleic Acids Res. 52, 1064–1079. https://doi.org/10.1093/nar/gkad1152.

37. Philippe, L., van den Elzen, A.M.G., Watson, M.J., and Thoreen, C.C. (2020). Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5′ TOP motifs. Proc. Natl. Acad. Sci. USA 117, 5319–5328. https://doi.org/10.1073/pnas.1912864117.

38. Nance, K.D., and Meier, J.L. (2021). Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. ACS Cent. Sci. 7, 748–756. https://doi.org/10.1021/acscentsci.1c00197.

39. Kumar, P., Hellen, C.U.T., and Pestova, T.V. (2016). Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. Genes Dev. 30, 1573–1588. https://doi.org/10.1101/gad.282418.116.

40. Vaidyanathan, S., Azizian, K.T., Haque, A.K.M.A., Henderson, J.M., Hendel, A., Shore, S., Antony, J.S., Hogrefe, R.I., Kormann, M.S.D., Porteus, M.H., and McCaffrey, A.P. (2018). Uridine Depletion and Chemical Modification Increase Cas9 mRNA Activity and Reduce Immunogenicity without HPLC Purification. Mol. Ther. Nucleic Acids 12, 530–542. https://doi.org/10.1016/j.omtn.2018.06.010.

41. Corbett, K.S., Edwards, D.K., Leist, S.R., Abiona, O.M., Boyoglu-Barnum, S., Gillespie, R.A., Himansu, S., Schäfer, A., Ziwawo, C.T., DiPiazza, A.T., et al. (2020). SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. Nature 586, 567–571. https://doi.org/10.1038/s41586-020-2622-0.

42. Fan, A., and Sharp, P.P. (2021). Inhibitors of Eukaryotic Translational Machinery as Therapeutic Agents. J. Med. Chem. 64, 2436–2465. https://doi.org/10.1021/acs.jmedchem.0c01746.

43. Hao, P., Yu, J., Ward, R., Liu, Y., Hao, Q., An, S., and Xu, T. (2020). Eukaryotic translation initiation factors as promising targets in cancer therapy. Cell Commun. Signal. 18, 175. https://doi.org/10.1186/s12964-020-00607-9.

44. Demircioğlu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N.A., Kahles, A., Lehmann, K.V., Stegle, O., Brazma, A., et al. (2019). A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. Cell 178, 1465–1477.e17. https://doi.org/10.1016/j.cell.2019.08.018.

45. Weber, R., Ghoshdastider, U., Spies, D., Duré, C., Valdivia-Francia, F., Forny, M., Ormiston, M., Renz, P.F., Taborsky, D., Yigit, M., et al. (2023). Monitoring the 5′UTR landscape reveals isoform switches to drive translational efficiencies in cancer. Oncogene 42, 638–650. https://doi.org/10.1038/s41388-022-02578-2.

46. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

47. Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) 1, 895–905. https://doi.org/10.3390/biology1030895.

48. Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26. https://doi.org/10.1186/1748-7188-6-26.

49. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

50. van den Elzen, A.M.G., Watson, M.J., and Thoreen, C.C. (2022). mRNA 5′ terminal sequences drive 200-fold differences in expression through effects on synthesis, translation and decay. PLoS Genet. 18, e1010532. https://doi.org/10.1371/journal.pgen.1010532.

51. Philippe, L., Vasseur, J.J., Debart, F., and Thoreen, C.C. (2018). La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region. Nucleic Acids Res. 46, 1457–1469. https://doi.org/10.1093/nar/gkx1237.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Hela cytoplasmic extract	Ipracell	Cat#CC-01-40-50
Chemicals, peptides, and recombinant proteins		
Cycloheximide	Millipore Sigma	Cat#01810
Phenylmethanesulfonyl fluoride (PMSF) solution 0.1 M	Millipore Sigma	Cat#93482
RNasin Plus Ribonuclease Inhibitor	Promega	Cat#N2615
TURBO DNase (2 U/μL)	Thermo Fisher	Cat#AM2239
Phenol:chloroform:IAA 25:24:1 pH 6.6	Thermo Fisher	Cat#AM9732
cOmplete Mini Protease Inhibitor Cocktail Tablets	Millipore Sigma	Cat#11836153001
Creatine phosphokinase	Millipore Sigma	Cat#C3755
100 bp DNA ladder	NEB	Cat#N3231L
SYBR Gold (10,000× concentrate in DMSO)	Thermo Fisher	Cat#S11494
T4 RNA Ligase 1, high concentration	NEB	Cat#M0437M
Buffer RLT	Qiagen	Cat#79216
SuperScript III Reverse Transcriptase	Thermo Fisher	Cat#18080-044
Phenol solution	Millipore Sigma	Cat#P4557
GlycoBlue Coprecipitant (15 mg/mL)	Thermo Fisher	Cat#AM9516
Lipofectamine MessengerMax Transfection Reagent	Thermo Fisher	Cat#LMRNA015
Vaccinia Capping System	NEB	Cat#M2080S
Megashortscript T7 Transcription Kit	Thermo Fisher	Cat#AM1354
Megascript T7 Transcription Kit	Thermo Fisher	Cat#AMB13345
Dimethyl sulfate	Millipore Sigma	Cat#D186309
Formamide	Invitrogen	Cat#15515-026
TGIRT-III Reverse Transcriptase	InGex	Cat#NA
L-glutamic acid hemimagnesium salt tetrahydrate	Millipore Sigma	Cat#49605
L-glutamic acid potassium salt monohydrate	Millipore Sigma	Cat#G1149
CleanCap Reagent AG	Trilink	Cat#N-7113
N1-methylpseudouridine triphosphate	Trilink	Cat#N-1081
DMEM, high glucose	Thermo Fisher	Cat#11965092
Fetal Bovine Serum	Millipore Sigma	Cat#F4135
Pen Strep	Thermo Fisher	Cat#15140122
Trypsin-EDTA 0.25%	Thermo Fisher	Cat#2520056
Guanosine 5'-triphosphate, [α-^{32}P]	Revvity	Cat#BLU006H250UC
X-tremeGENE 9 DNA Transfection Reagent	Roche	Cat#06365787001
Protoscript II Reverse Transcriptase	NEB	Cat#M0368S
Phusion High-Fidelity DNA Polymerase	NEB	Cat#M0530L
Critical commercial assays		
Bright-Glo Luciferase Assay System	Promega	Cat#E2620
Nano-Glo Luciferase Assay System	Promega	Cat#N1120
Qubit dsDNA Quantitation Assay, High Sensitivity	Thermo Fisher	Cat#Q32854
MycoStrip Mycoplasma Detection Kit	Invivogen	Cat#rep-mys-50
Deposited data		
Raw and processed DART sequencing data	This paper	GEO: GSE256185

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Human: Hela cells	ATCC	Cat#CCL-2; RRID: CVCL_0030
Human HEK293T cells	ATCC	Cat#CRL-3216; RRID: CVCL_0063
Oligonucleotides		
5 UTR oligonucleotide pools	This paper	Synthesized by Twist Bioscience
See Table S1 for oligonucleotide sequences	This paper	N/A
Software and algorithms		
STAR	Dobin et al.[46]	https://github.com/alexdobin/STAR; RRID: SCR_004463
BBTools Suite	Bushnell B.	https://sourceforge.net/projects/bbmap/; RRID: SCR_016968
Flexbar	Dodt et al.[47]	https://github.com/seqan/flexbar; RRID: SCR_013001
Kozak similarity scoring	Gleason et al.[35]	https://github.com/Agleason1/TIS-Predictor; Zenodo: https://doi.org/10.5281/zenodo.5110255
RNAFold (ViennaRNA)	Lorenz et al.[48]	https://github.com/ViennaRNA/ViennaRNA; RRID: SCR_008550
DREME	Bailey[24]	https://meme-suite.org/meme/doc/download.html; RRID: SCR_001783
DeSeq2	Love et al.[49]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html; RRID: SCR_015687
DART data processing pipeline	This paper	Zenodo: https://doi.org/10.5281/zenodo.14181557
Other		
MaXtract High Density 15 mL Tubes	Qiagen	Cat#129065
MaXtract High Density 1.5 mL Tubes	Qiagen	Cat#129046
Gradient Station Base Unit	Biocomp	Cat#153
Magnabase Holder and Marker Block	Biocomp	Cat#105-914A-IR
10 mm Isopycnic Long Caps	Biocomp	Cat#105-514-6
Marker block for SW41 tubes	Biocomp	Cat# 105-614A
Dynabeads MyOne Silane	Thermo Fisher	Cat#37002D
Ampure XP Reagent	Beckman Coulter	Cat#A63881
DNA LoBind 1.5mL Tubes	Eppendorf	Cat#022431021
SW41 Swining Bucket Rotor	Beckman Coulter	Cat#331362
Ultra-Clear Centrifuge Tubes	Beckman Coulter	Cat#344059
Monarch RNA Cleanup Kit (10μg)	NEB	Cat#T2030
Monarch RNA Cleanup Kit (500μg)	NEB	Cat#T2050
Detailed DART protocol	This paper	Methods S1

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

293T cells were obtained from ATCC (CRL-3216) and maintained in DMEM (Thermo Fisher, 11965092) with 10% heat-inactivated FBS (Millipore Sigma, F4135), 100 U/mL Penicillin, 100 mg/mL Streptomycin (Thermo Fisher 15140122) in cell culture incubators at 5% carbon dioxide at 37 degrees.

Hela cells were obtained from ATCC (CRM-CCL-2) and maintained in DMEM (Thermo Fisher, 11965092) with 10% heat-inactivated FBS (Millipore Sigma, F4135), 100 U/mL Penicillin, 100 mg/mL Streptomycin (Thermo Fisher, 15140122) in cell culture incubators at 5% carbon dioxide at 37 degrees.

Cell lines were routinely monitored for mycoplasma contamination using the MycoStrip Mycoplasma Detection Kit (InvivoGen, rep-mys-50).

METHOD DETAILS

5′ UTR Pool Design and Synthesis

To obtain the 5′ UTRs that are expressed in human tissue, total RNAseq datasets for liver, kidney, neuron, T cell, K562, MCF10a, and HepG2 from the ENCODE project were gathered and aggregated. Transcripts were first filtered by length ranging from 10-230

nucleotides and by TPM > 1 in any dataset. The top 32,355 transcripts with the highest TPM were selected. The top 6,000 transcripts were also chosen for the list of endogenous CDS. The sequences of the selected 5′ UTR and CDS were retrieved from Gencode41. The 5′ UTR sequences were concatenated with EGFP CDS sequences or their endogenous CDS sequences. T7 promoter sequence (GCTAATACGACTCACTATAGGG) was added in front of 5′ UTRs and RT primer binding sequence (CACTCGGGCACCAAGGAC) was added to the 3′ end. CDS sequences were trimmed from 3′ end to make 300-nucleotide total length oligos. All upstream ATGs in the 5′ UTRs were mutated to AGT. The last in-frame codon in CDS was changed to termination codon TAA.

The scanning deletion pool was constructed based on the top 200 and bottom 200 5′ UTR sequences from the initial DART results. For each parental sequence, a scanning deletion of every 6-nt window was made throughout the entire 5′ UTR to generate deletion variants. Upstream ATG generated from the deletion were removed by mutation to AGT.

Designed oligos were purchased as a DNA pool (Twist Bioscience) and PCR amplified with oCL01 (enzymatic capping) or oCL02 (co-transcriptional capping) and the PCR product was gel-purified. Pool RNA was produced by runoff T7 transcription using the MEGAshortscript T7 transcription kit (ThermoFisher AM1354) using the purified DNA template. The pool RNA was then gel-purified and capped using the Vaccinia Capping System (NEB M2080S). For co-transcriptional capping, the T7 promoter sequence was changed to GCTAATACGACTCACTATAAGG and CleanCap AG (TriLink N-7113) was added to the *in vitro* transcription reaction according to manufacturer protocols.

For multiplexed DART reactions, DNA pools were PCR amplified with the common oCL02 forward primer and barcoded reverse primers oCL_Bar1-8. Four barcode sequences were used for uridine and four were used for N1-methylpseudouridine modified RNAs. N1-methylpseudouridine substituted RNAs were produced by the complete replacement of uridine triphosphate in the T7 transcription reaction with N1-methylpseudouridine triphosphate (Trilink N-1081). Uridine and N1-methylpseudouridine RNAs were then mixed in equimolar ratios prior to in vitro translation.

In vitro translation

For standard DART reactions, 40 picomoles of capped pool RNA was added to each 0.5 mL *in vitro* translation reaction containing 0.25 mL of Hela cytoplasmic extract (Ipracell CC-01-40-50), 16 mM HEPES KOH pH 7.4, 40 mM potassium glutamate, 2 mM magnesium glutamate, 0.8 mM ATP, 0.1 mM GTP, 20 mM creatine phosphate, 0.1 mM spermidine, 165 μg creatine phosphokinase, 1.6 mM DTT, 2mM PMSF, 140 units of RNasin Plus, 1X cOmplete EDTA-free protease inhibitor cocktail, and 0.5 mg/mL cycloheximide. The volume of HEK extract and buffers are proportionally reduced as indicated in the miniaturization experiment. Reactions were incubated at 37 °C for 30 minutes in a shaking thermomixer. Following incubation, reactions were immediately loaded onto a sucrose gradient for 80S isolation. Miniaturized DART reactions maintained the same concentrations of components.

80S Isolation and RNA extraction

Translation reactions were loaded onto 10-50% sucrose gradients containing 20 mM HEPES KOH pH 7.4, 2 mM magnesium glutamate, 100 mM potassium glutamate, 0.1% Triton X-100, 3 mM DTT, 0.1 mg/mL cycloheximide and centrifuged at 35,000 RPM for 3 hours at 4°C in a Beckman SW41 rotor. Gradients were fractionated using a Biocomp Gradient Station (Biocomp Instruments) with continual monitoring of absorbance at 260 nm. Fractions corresponding to the 80S peak were collected and pooled. To extract RNA from the pooled fractions, 650 μL of acid phenol and 40 μL of 20% SDS was added per 600 μL of fraction volume. The mixture was then transferred to a 65 °C water bath and incubated for 10 minutes, vortexing every minute. The samples were then cooled for 5 minutes on ice and transferred to a pre-spun 15 mL MaXtract tube (Qiagen 129065) containing an equal volume of chloroform to the acid phenol used. After centrifugation to separate the aqueous and organic phases, the aqueous phase containing RNA was transferred to a new 1.5 mL tube and subjected to a second extraction using 650 μL of phenol:chloroform:IAA (25:24:1) pH 6.8 per 600 μL of aqueous volume. A final extraction was performed using 650 μL of chloroform per 600 μL of aqueous volume and the extracted RNA was precipitated with isopropanol.

DART Library Preparation

RNA from the input pool and 80S fractions was reverse transcribed using oCL05 (Hiseq) or oCL04 (Novaseq) primers using Superscript III (Invitrogen 18080093). The resulting full-length cDNA was gel-purified and ligated to the 5′ adaptor OWG920 (Hiseq) or oCL03 (Novaseq) with High Concentration T4 RNA Ligase (NEB M0437M) overnight on a shaking thermomixer. The ligated cDNA was then purified using 10 μL of MyOne Silane beads (Invitrogen 37002D). Libraries were then amplified using the primer sets RP1 and oCL06 (Hiseq), or i5_Nova and i7_Nova (Novaseq), gel purified, and sequenced on a HiSeq 2500 or Novaseq 6000.

Data Processing

The raw fastq files were first demultiplexed using Flexbar.[47] After demultiplexing, each file was processed with BBMap module to trim adaptors, collapse duplications, and remove UMIs. The processed sequencing results were aligned to pool sequences using STAR,[46] with the following parameters: soloStrand = Forward; alignSJoverhangMin = 999; alignIntronMax = 1; alignIntronMin = 999; outFilterMismatchNmax = 1. The counting of alignment was performed by BBMap pileup. The sequences with < 10 counts in any of the samples were removed. To calculate counts per million (CPM), a total depth for each translation reaction was obtained by summing total counts from all the samples that derived from the same translation reaction, then the raw counts for each sample were normalized by this total depth.

In vitro luciferase assays

5′ UTR sequences were cloned into a plasmid downstream of a T7 promotor and immediately upstream of GFP coding sequence followed by a 15 amino acid glycine-serine linker and either Nano luciferase or Firefly luciferase coding sequence and an encoded 60 nucleotide poly(A) tail. The resulting plasmids were then linearized and used as a template for run-off T7 transcription using the MEGAscript T7 transcription kit (Invitrogen AMB13345). The mRNAs are purified with RNA cleanup columns (NEB T2050). 0.05 pmol of mRNA was incubated in a 10 μL *in vitro* translation reaction containing 4 μL of Hela cytoplasmic extract, 16mM HEPES KOH pH 7.4, 0.1 mM spermidine, 0.8 mM ATP, 0.1 mM GTP, 40 mM potassium glutamate, 2 mM magnesium glutamate, and 20 mM creatine phosphate. Reactions were incubated at 37 °C for 30 minutes in a shaking thermomixer. Following incubation, reactions were immediately halted with ice-cold Passive Lysis Buffer (Promega E1941). Firefly luciferase and nano luciferase levels were then measured using Bright-Glo (Promega E2620) and Nano-Glo (Promega N1130), respectively, on a luminometer.

Radiolabeled Vaccinia Capping Assays

DNA oligonucleotides corresponding to the sequences used in the DART assay were PCR amplified and gel purified on a non-denaturing 8% TBE polyacrylamide gel. The gel-purified products were used as a template for run-off T7 transcription using the MEGAscript T7 transcription kit. 7.5 μg of each RNA were incubated in capping reactions using the Vaccinia capping system (NEB) following the manufacturer's protocol with complete substitution of GTP for alpha-^{32}P-GTP (Revvity, BLU006H250UC). Following capping, RNA was cleaned up using the Monarch 10 μg RNA cleanup kit (NEB T2030L) and washed until no radioactive signal was detectable in the wash buffer. RNA was eluted into 20 μL of water and 4 μL of the eluate was measured on a scintillation counter to detect incorporated alpha-^{32}P-GTP.

In cell luciferase assays with *in vitro* transcribed mRNA

5′ UTR sequences were cloned into a plasmid downstream of a T7 promotor and immediately upstream of GFP coding sequence followed by a 15 amino acid glycine-serine linker and either Nano luciferase or Firefly luciferase coding sequence and an encoded 60 nucleotide poly(A) tail. The resulting plasmids were then linearized and used as a template for run-off T7 transcription using the MEGAscript T7 transcription kit (Invitrogen AMB13345). Hela cells were cultured in DMEM supplemented with 10% FBS. 10,000 cells per well were seeded into a 96-well plate and allowed to adhere overnight. The following day, cells were transfected with 100 nanograms of mRNA using the Lipofectamine MessengerMAX Transfection Reagent (Invitrogen LMRNA015). Cells were lysed in 50μL of Bright-Glo (Promega E2620) plus 50 μL of 1X phosphate-buffered saline 2 hours after transfection and luciferase activity was measured on a luminometer.

In cell luciferase assays with plasmid

5′ UTR sequences were cloned into a plasmid downstream of a CMV promotor, which has been adjusted to initiate at the beginning of the inserted 5′ UTR.[50,51] CDS is the same Nanoluc template used for mRNA *in vitro* transcription. Hela cells were seeded at 150,000 per well into a 12-well plate and allowed to adhere overnight. The following day, cells were transfected with 500 nanograms of mRNA using the X-tremeGENE 9 DNA transfection Reagent (Roche 06365787001) and incubated for 8 hours. Cells were lysed in 100μL of Passive Lysis Buffer (Promega E1941). 10μL of the lysate were mixed with 10μL of Nano-Glo (Promega N1120) for luciferase detection. Remaining lysate was incubated with 1 mL of TRIzol (Thermo 15596026) for RNA purification. After DNA depletion with Turbo DNase (Thermo AM2238), 1 μg of total RNA was reserve transcribed with ProtoScript II (NEB M0368S). The expression levels of Nanoluc and GAPDH were determined by qPCR with the following primer sets: oLX01-oLX02, and oLX03-oLX04.

Tetramer analysis from scanning mutagenesis

The change in ribosome recruitment score (deltaRRS) was obtained by dividing the RRS for each deletion sequence by the RRS for its parent sequences. Tetramers contained within each 6-nucleotide list were associated with a list of all possible tetramer sequences. To reduce noise, only tetramers that were deleted at least 30 times were considered. The impact of each tetramer sequence was calculated as the mean of deltaRRS values. Significance was determined by Wilcoxon test and the p-values were adjusted using the Benjamini-Hochberg correction. P-values less than 0.01 were considered significant.

DMS-MaPseq library preparation

DNA pool was amplified with PCR (forward primer oCL02; reverse primer oCL_Bar2. RNA was produced by T7 *in vitro* translation using the MEGAshortscript kit (ThermoFisher Scientific AM1354), co-transcriptionally capped with CleanCap Reagent AG (TriLink N-7113), treated with TURBO DNase (Invitrogen) and gel-purified. For DMS probing, RNA (2 μg in 6 μL water) was denatured at 95 °C for 2 min. RNA was refolded by adding 88.8 μL 300 mM sodium cacodylate and 1 μL RNasin Plus (Promega N2611) and 10 min incubation at 37 °C, followed by addition of 1.2 μL 500 mM MgCl2 and 20 min incubation at 37 °C. DMS (3 μL, Sigma-Aldrich D186309) was added to the folded RNA solution and allowed for 2 min incubation at 37 °C, then quenched by adding 42.8 μL β-mercaptoethanol (BME). For denatured RNA control, RNA (2 μg in 6 μL water) was mixed with 1 μL RNasin Plus, 39.2 μL water, 50 μL formamide (Invitrogen 15515-026) and 0.8 μL 0.5 M EDTA and incubated at 95 °C for 2 min. The solution was incubated with 3 μL DMS and quenched with 42.8 μL BME. RNAs were purified via ethanol precipitation and reverse-transcribed

using TGIRT-III (InGex) and primer oCL04. The resulting cDNA was gel purified and a 5' adaptor (oCL03) was ligated. Libraries were prepared as described above and sequenced via Novaseq.

DMS-MaPseq analysis

The data processing steps for DMS-MaPseq results, including adapter removal, PCR duplicate removal, and alignment follow the procedure for DART analysis, except using the default setting for STAR: outFilterMismatchNmax. Mismatches to the reference sequence were labeled with MD tags using the SAMtools calmd function. The coverage and mutation rate for each position were calculated and averaged within replicates using a house-developed script. The sequences with a coverage of < 200 reads were removed from further analysis. For each sequence, the mutation rates on A/C positions were normalized to a 0-1 scale, while U/G positions were not considered. The mutation rates were used to predict base pairing probability using RNAfold from the ViennaRNA package.[48]

Linear regression model

The linear regression model was built with the R caret package based on four features: 5' UTR length, GC content in 5' UTR, the minimum free energy, and the Kozak similarity score. DART results were randomly split, 80% (29523 5' UTRs) for the training set and 20% (7396 5' UTRs) for the test set. The accuracy of the linear model was determined by the correlation between the prediction and the RRS scores observed in the test set.

Preparation of HEK293T translation extracts

HEK293T cells were maintained in DMEM (Gibco 11965-092) supplemented with 10% FBS (heat-inactivated, Sigma F4135). The cells were seeded at 17,200 cells/cm^2 and cultured for 3 days. At the moment of harvest, the cells were dissociated with 0.25% trypsin-EDTA (Gibco 2520056), and the collected cell pellet was washed once with 30 mL of ice-cold PBS and once with 10 ml ice-cold isotonic buffer (HEPES-KOH 16 mM, potassium acetate 100 mM, magnesium acetate 0.5 mM, DTT 5 mM). The pellet was resuspended and lysed with an equal volume of hypotonic buffer (HEPES-KOH 16 mM, potassium acetate 10 mM, magnesium acetate 0.5 mM, DTT 5 mM), and incubated on ice for 10 min. The cell lysate was homogenized with 10 passes through a 27-gauge needle and centrifuged at 16,000 x g at 4°C for 1 min. The supernatant was collected and frozen at -80°C until use.

Analysis of differential ribosome recruitment

Differential ribosome recruitment activity between 5' UTRs paired with their endogenous coding sequence and their EGFP counterparts or between unmodified and m1Ψ-substituted RNAs was determined using the R package DESeq2.[49] For comparisons between U and m1Ψ, the read depths for each sample were used to estimate size factors. P-values were adjusted using Benjamini-Hochberg correction and an adjusted p-value of less than 0.01 and an absolute $\log_2$ fold change > 1 were considered significant changes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Detailed information on quantification and statistical analyses are provided in the method details section and figure legends.