2020-10-26

Transcriptome-wide investigation of stop codon readthrough in Saccharomyces cerevisiae [poster]

Kotchaphorn Mangkalaphiban
University of Massachusetts Medical School

Et al.

Transcriptome-wide investigation of stop codon readthrough in *Saccharomyces cerevisiae*

Kotchaphorn Mangkalaphiban, Feng He, Robin Ganesan, Chan Wu, Richard Baker, and Allan Jacobson
Department of Microbiology and Physiological Systems
University of Massachusetts Medical School, Worcester, MA 01655

INTRODUCTION

Translation of mRNA into a polypeptide is terminated when the release factor eRF1 recognizes a stop codon in the ribosomal A site and stimulates nascent peptide release. However, "stop codon readthrough" can occur when a near-cognate tRNA outcompetes eRF1 in decoding the stop codon, resulting in the continuation of elongation into the mRNA 3'-UTR. Previous studies with reporter systems have shown that the efficiency of termination or readthrough is modulated by multiple *cis*-acting elements (Figure 1). It remains to be determined whether these elements are important at a genome-wide level and whether other mRNA features proximal to the stop codon significantly affect termination and readthrough efficiencies *in vivo*. Accordingly, we carried out ribosome profiling analyses of yeast cells expressing defective or depleted eRF1 and developed bioinformatics strategies to calculate readthrough efficiency, and to identify mRNA and peptide features which influence that efficiency.

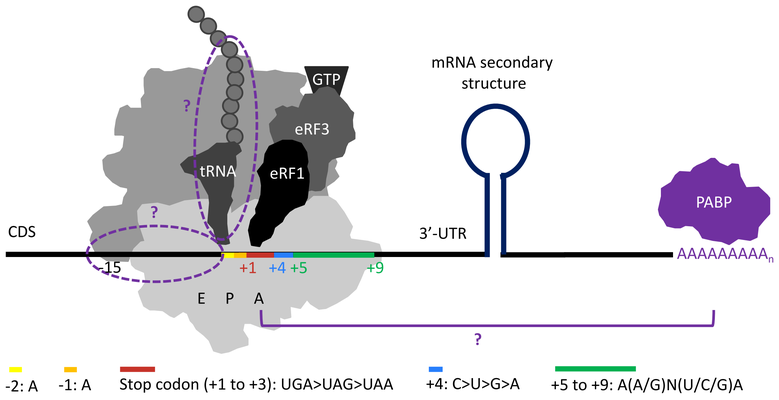


Figure 1. Factors affecting stop codon readthrough in eukaryotes (Adapted from Rodnina *et al.*, Nucleic Acids Res, 2019). eRF = eukaryotic release factor, PABP = poly(A)-binding protein.

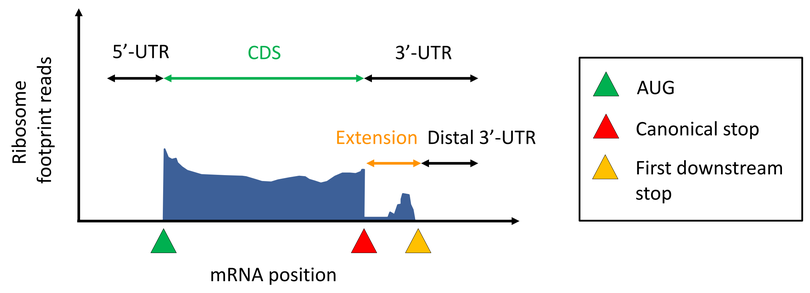
METHODS

Ribosome profiling datasets

eRF1 (sup45 for yeast) mutants:
- This project
  - WT at 25 °C
  - WT at 37 °C
  - *sup45-ts* at 25 °C
  - *sup45-ts* at 37 °C
- Wu *et al.*, Molecular Cell, 2019
  - WT
  - *sup45-d* (depleted)

Recycling factor mutants:
(to rule out effects of re-initiation events)
- Young *et al.*, Cell, 2015
  - WT
  - *rli1-d* (degron, depleted)
- Young and Guydosh, Cell Rep., 2019
  - WT
  - *hcr1Δ*

Calculation of readthrough efficiency



B

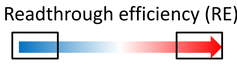
$$\text{Readthrough efficiency} = \log_2\left(\frac{\#\text{ of frame 0 footprints}_{\text{extension}} \,/\, \text{length (kb)}_{\text{extension}}}{\#\text{ of frame 0 footprints}_{\text{CDS–33nt}} \,/\, \text{length (kb)}_{\text{CDS–33nt}}}\right)$$

Figure 2. A. Schematic depicting ribosome footprint count across a gene's regions (Adapted from Brar and Weissman, Nat. Rev. Mol. Cell Biol., 2015). **B.** Formula to calculate readthrough efficiency.

Analysis strategies

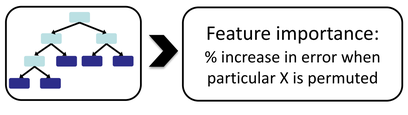
Genes with the following properties were excluded from further analyses in order to minimize noise:
- Lack 3'-UTR annotations
- Overlap > 18 bp on the same strand
- $RPKM_{CDS} < 5$ and $RPKM_{3'\text{-}UTR} < 0.5$

The remaining genes were sorted by their readthrough efficiency values and grouped:



Two machine learning approaches involving the random forest algorithm (Breiman, Mach. Learn., 2001) were trained for each sample:
1. Regression: use X to predict Y_1
2. Classification: use X to predict Y_2

	X				Y_1	Y_2
Gene	Stop codon	nt +4	3'-UTR length	...	RE	Group
Gene1	UGA	C	75	...	1.45323	High
Gene2	UAA	A	143	...	-3.5244	Low
Gene3	UAG	G	97	...	0.2347	High
...	...	...	...	...		...



RESULTS

Mutation of eRF1 promotes readthrough and accumulation of ribosomes at stop codons

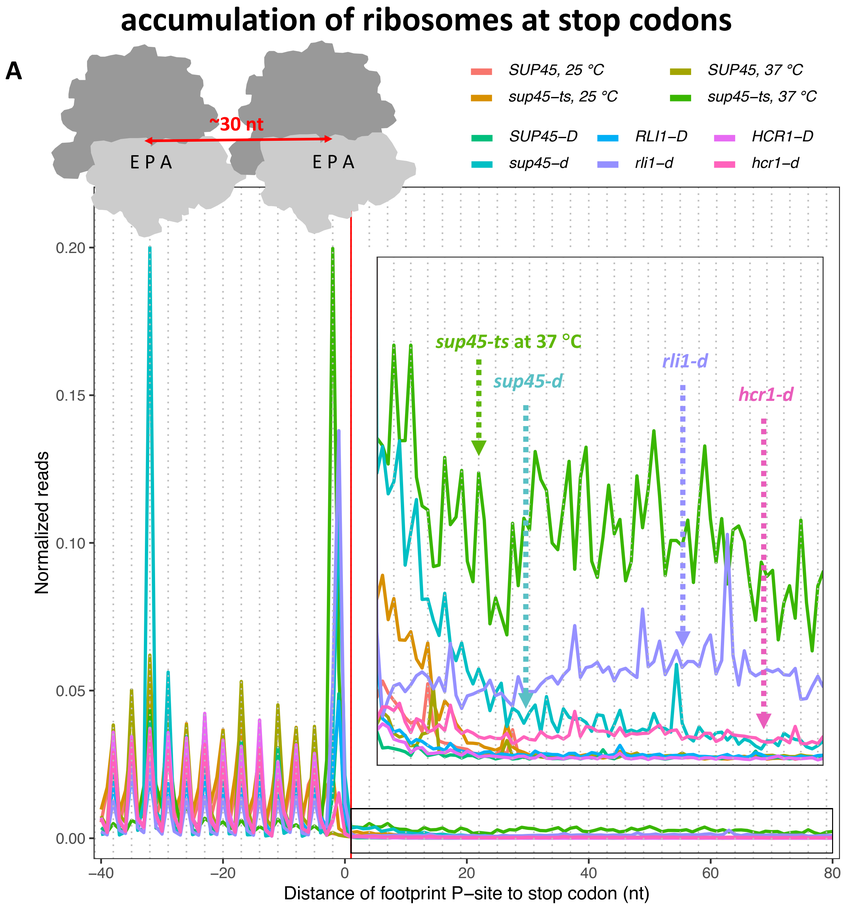


Figure 3. A. Normalized ribosome footprints for all non-overlapping genes that have 3'-UTR annotations aligned at their canonical stop codons. Footprints were plotted by the position of their P-sites. (Inset) Magnified view of the region in the black box, showing increased 3'-UTR ribosome occupancy in *sup45-ts* cells at 37 °C, *sup45-d*, *rli1-d*, and *hcr1-d* strains relative to WT. **B. Fractions of reads of each of the three reading frames in different mRNA regions.** The results of each sample were the average of 1-3 replicates ± SEM. (Red arrows) Compared to their WT counterpart, *sup45-ts* at 37 °C and *sup45-d* showed dominance of frame 0 (in-frame with CDS) in the extension region, indicating readthrough.

Random forest analyses determined that stop codon identity, P-site amino acid identity, and 3'-UTR length are important factors in predicting readthrough efficiency

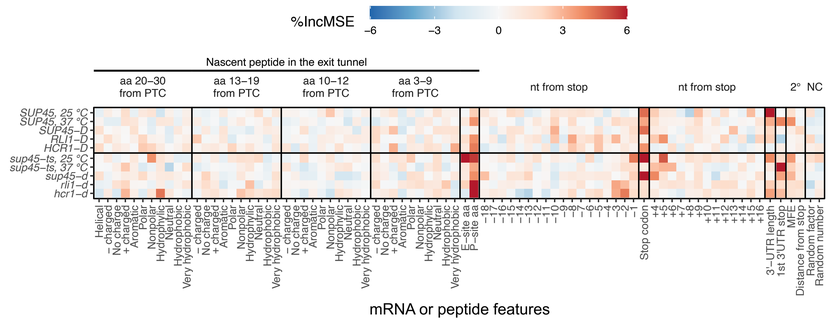


Figure 4. Feature importance extracted from the random forest regression models. The relative importance of a feature is represented by percent increase in mean square error (%IncMSE), which is percent increase in mean square error that results from permuting the feature. The higher the %IncMSE, the more important the feature is in predicting readthrough efficiency. Similar results were obtained from random forest classification models (not shown).

Identities of readthrough-permissive stop codons and nucleotides at positions -2 to +9 were mostly true at a genome-wide level

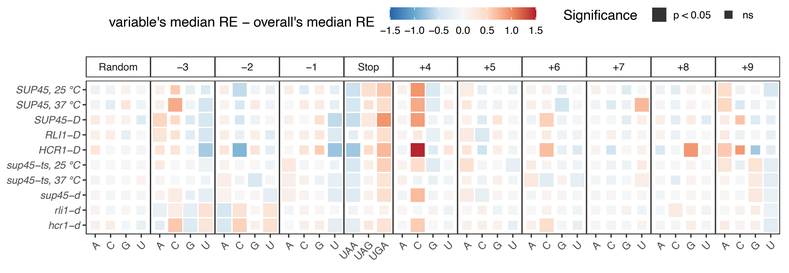


Figure 5. Heatmap of median readthrough efficiency of genes containing particular stop codon or nucleotide relative to median readthrough efficiency of all genes in the sample. Positive values (red) indicate that the group of genes had higher readthrough efficiencies compared to the sample median, while negative values (blue) indicate lower readthrough efficiencies. Wilcoxon's rank sum test was used to determine whether the difference in group and sample median was significant. Significant difference was represented as a larger tile.

Specific codons in the P site were associated with higher or lower readthrough efficiencies

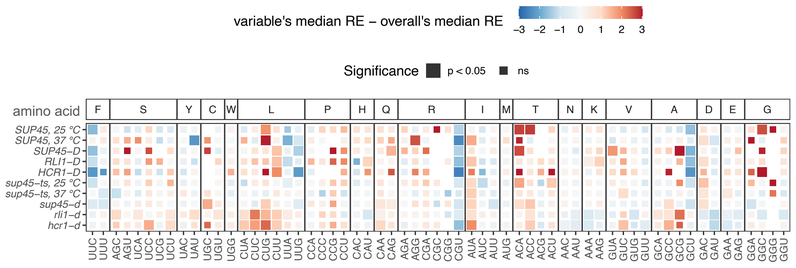


Figure 6. Heatmap of median readthrough efficiency of genes containing particular triplet codon in the P site (See legend of Figure 5 for details of analysis). There were six codons (UUC, UUU, UUA, UUG, CGU, and GCU) associated with lower readthrough efficiencies and five codons (CUG, CUU, AUA, ACA, and ACC) associated with higher readthrough efficiencies. No single property of amino acids or tRNAs was exclusive to either group.

Readthrough efficiency increased with 3'-UTR lengths in cells expressing defective or depleted eRF1, but not in wild-type or recycling factor mutant cells

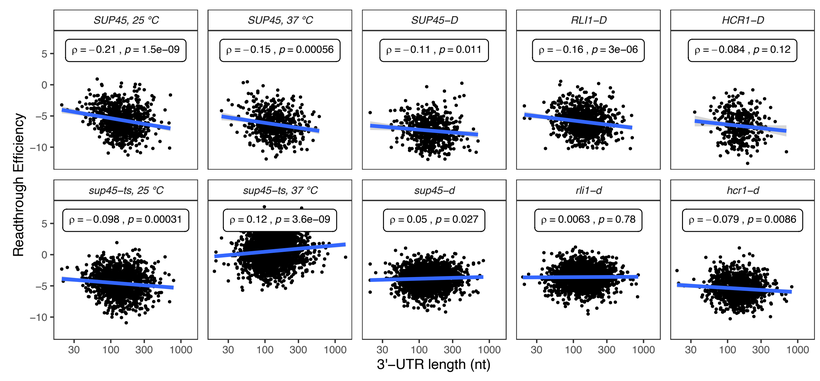


Figure 7. Scatter plot of readthrough efficiency vs. 3'-UTR length. Spearman's correlation coefficient (ρ) and p-value were calculated for each sample. The length of 3'-UTR is used to indicate the distance between the stop codon and poly(A)-binding protein (PABP). Because PABP is known to interact with eRF3 and enhance termination efficiency (Ivanov *et al.*, Nucleic Acids Res, 2016), it has been hypothesized that the proximity of the stop codon to PABP may influence termination/readthrough efficiency. This hypothesis is supported by the eRF1 mutant data, but not WT or recycling mutant data.

CONCLUSIONS

Our results confirmed the general roles of known regulatory elements in genome-wide regulation and identified several new mRNA or peptide features important for translation termination and readthrough:
- Random forest analyses revealed that the most influential features consist of the stop codon, the penultimate codon in the P site, and 3'-UTR length.
- Known effects of stop codon and surrounding nucleotide identities were mostly true at a genome-wide level.
- Specific codons in the P site were associated with higher (CUG, CUU, AUA, ACA, ACC) or lower (UUN, CGU, GCU) readthrough efficiencies.
- Readthrough efficiency increased with 3'-UTR length (implying proximity to PABP) in eRF1 mutant cells, while this trend was reversed in wild-type cells, suggesting that PABP's expected role in aiding termination was prominent only when eRF1 was inefficient. This is consistent with the extents of premature termination or readthrough observed in three different genes in *upf1Δ* cells (Abstract/Talk: Chan Wu, *et al.*).