Cell Systems, Volume *11*

Supplemental Information

Translation Initiation Site Profiling Reveals Widespread Synthesis of Non-AUG-Initiated Protein Isoforms in Yeast

Amy R. Eisenberg, Andrea L. Higdon, Ina Hollerer, Alexander P. Fields, Irwin Jungreis, Paige D. Diamond, Manolis Kellis, Marko Jovanovic, and Gloria A. Brar

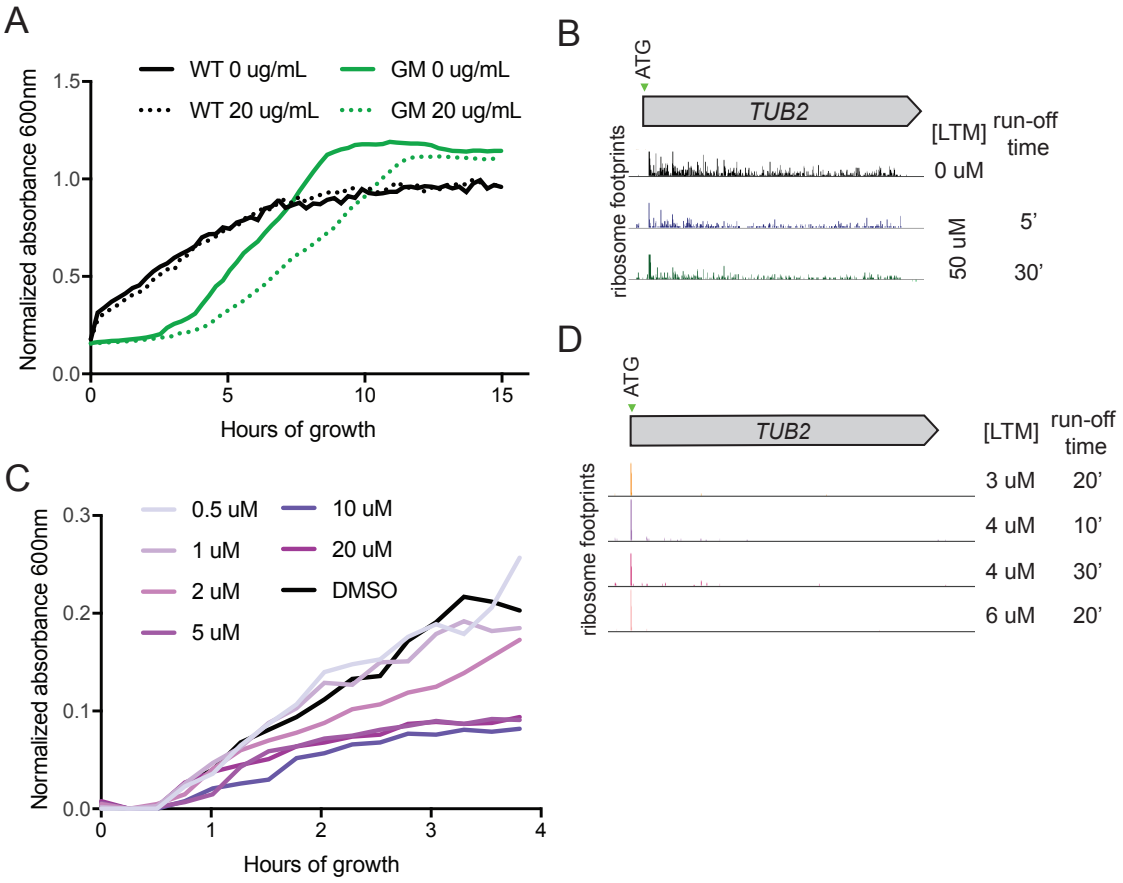

Figure S1: Optimization of TIS-profiling conditions for yeast, Related to Figure 1

(A) Growth curve of WT cells or Green Monster (GM) mutant cells treated with harringtonine. The GM strain lacks 16 ABC transporter drug efflux genes. Solid lines indicate no treatment and dotted lines indicate 20 ug/mL of harringtonine. Absorbance at 600 uM was used to measure growth over 16 hours. Estimated doubling time for WT cells is 3.7 and 3.3 hours for 0 and 20 ug/mL harringtonine respectively, and 1.9 and 2.8 hours for GM cells for 0 and 20 ug/mL harringtonine respectively.

(B) Ribosome profiling reads from cells treated with 0 or 50 μM LTM and either 5 or 30 minutes run-off time for a representative gene, *TUB2*.

(C) Growth curve of WT yeast treated with LTM at concentrations between 0-20 μM. Absorbance at 600 uM was used to measure growth over four hours. Estimtated doubling time for 0 μM LTM was 1.1 hours, and increased to 1.8 hours for 20 μM LTM.

(D) Ribosome profiling reads from cells treated with varying LTM concentration and run off times for a representative gene, *TUB2*.

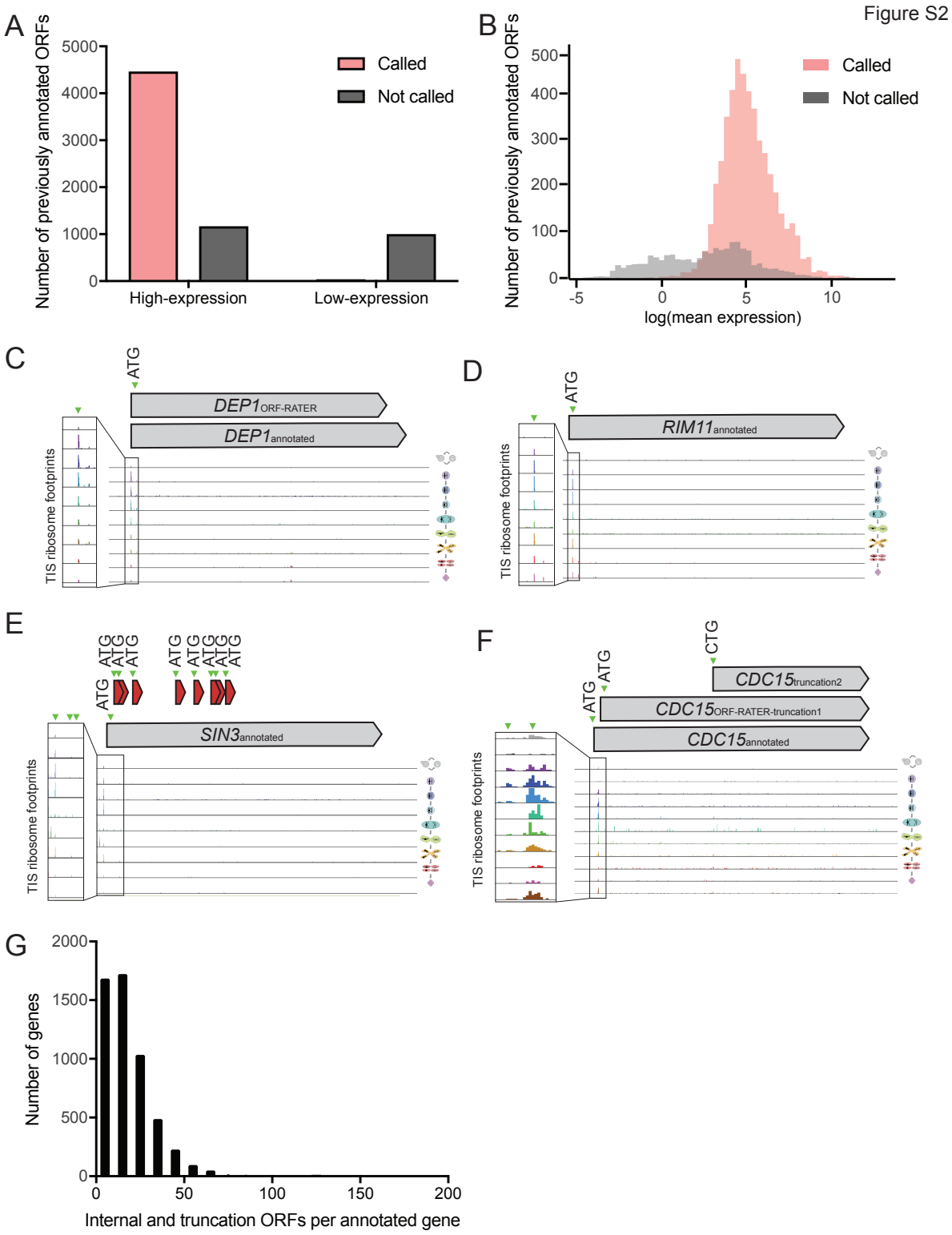

Figure S2: Categories of false positive and false negative ORF-RATER calls, Related to Figure 2

(A) Previously annotated ORFs that are called (pink) or not called (gray), at expression values greater (high-expression) or less than (low-expression) 5 mean RPKM. Approximately half of annotated ORFs that were not called have low expression.
(B) Distribution of expression (mean RPKM of all time points) for annotated ORFs that are called (pink) versus not called (gray).
(C) TIS-profiling for *DEP1,* a gene showing a change in stop codon annotation leading to it not being called as an annotated ORF by ORF-RATER.
(D) TIS-profiling for *RIM11*, a gene that is an example of a false negative, where an apparent peak is present at the annotated ATG but was not identified as a TIS by ORF-RATER.
(E) TIS-profiling for *SIN3*, a gene with many internal ORFs called, most of which are likely false positives.
(F) TIS-profiling for *CDC15*, a gene with two truncated ORFs called, the first of which represents a likely misannotation and the second of which is a likely false positive.
(G) Number of internally initiated ORFs called per annotated gene.

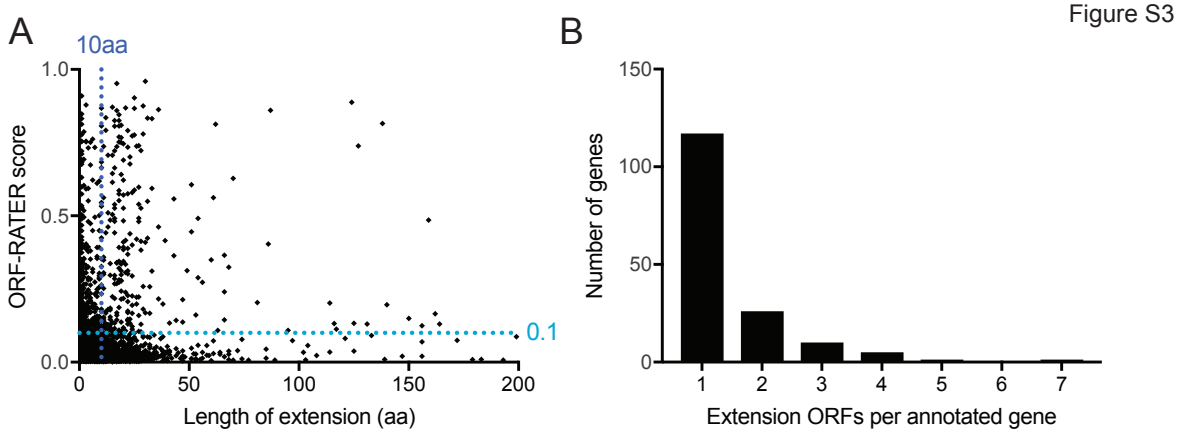

Figure S3: Properties of extension ORFs used for setting cutoffs, Related to Figure 2
(A) Length versus score for all extension ORFs, with a line showing the length cutoff at 10 amino acids and the score cutoff of 0.1.
(B) Number of extension ORFs called per annotated gene.

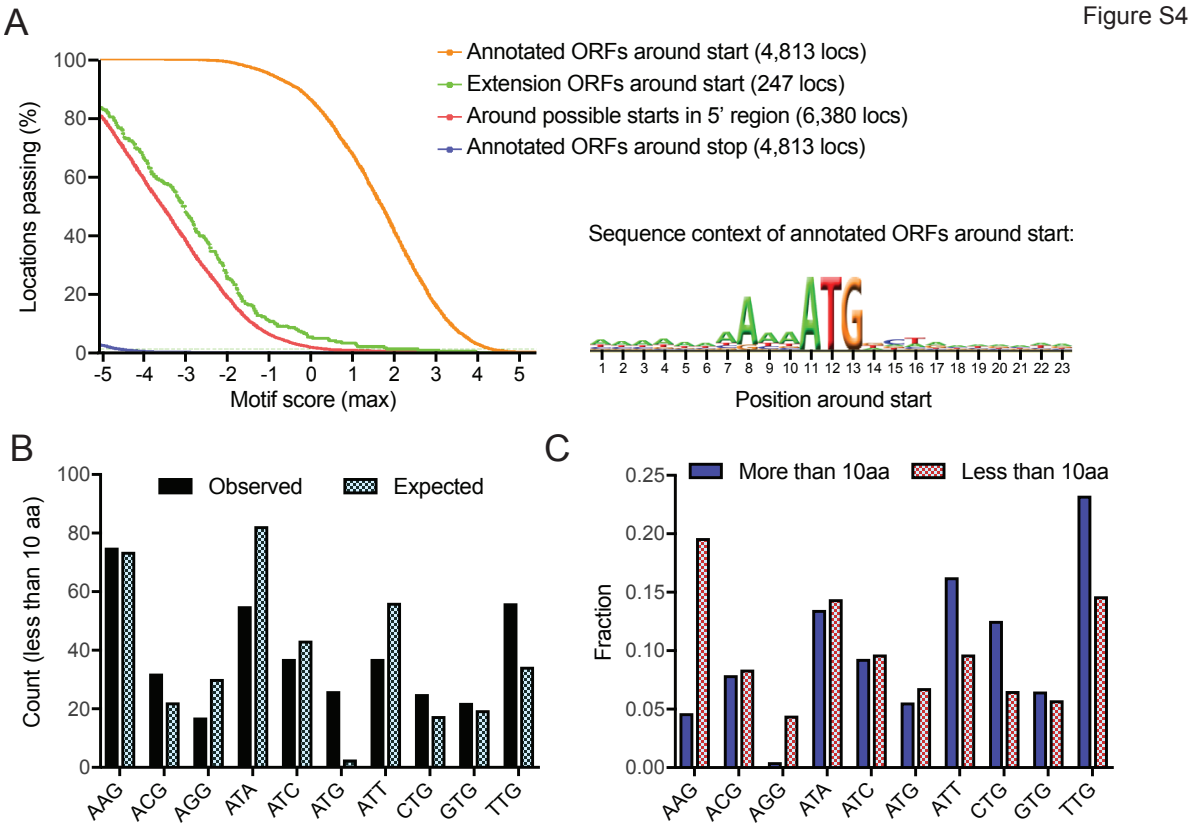

Figure S4: Translated near-cognate-initiated ORFs do not show Kozak sequence context enrichment, Related to Figure 3, Methods

(A) Enrichment plot (left) for yeast Kozak motif in the 10 bp region up and downstream of ORF-RATER called annotated genes (orange), near-cognate extensions (green), all possible in-frame near-cognate start codons (red), and stop codons for annotated genes (blue). Sequence context logo (right) was derived from annotated ORFs.

(B) Comparison of start codon usage for called extensions less than 10aa from canonical start codon (observed) to prevalence within UTR (expected), showing a lack of codon bias relative to what was observed for longer, more likely functional extensions (as seen in Figure 3F).

(C) Comparison of start codon usage between extensions that initiate more than and less than 10 amino acids upstream of the canonical start codon. Longer extensions show a stronger bias toward better start codons and against weaker start codons.

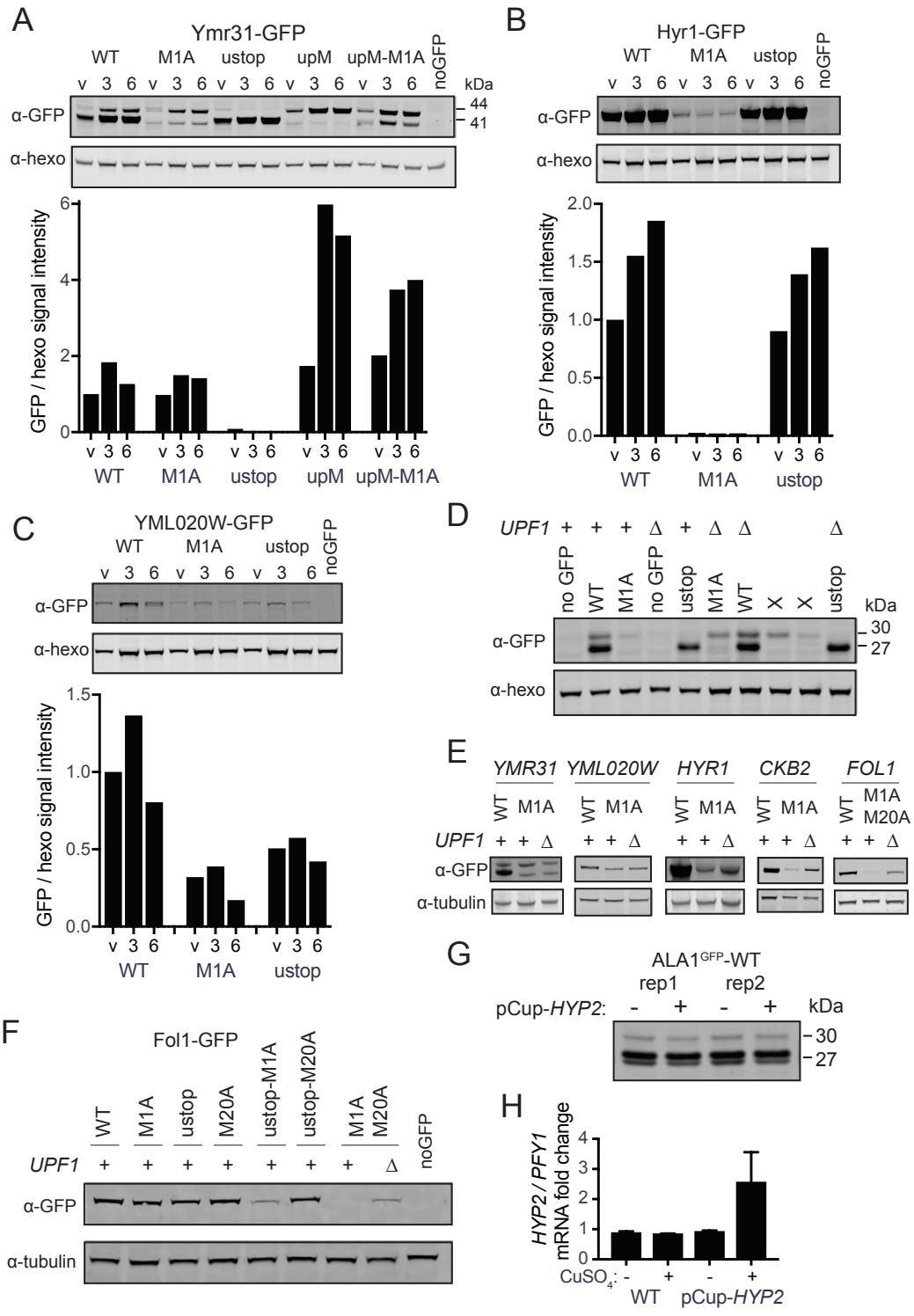

Figure S5: Western blot replicates and quantification for alternate isoforms, Related to Figures 4-7

(A) Replicate western blot of *YMR31-GFP* constructs, as in Figure 4C (top) and quantification of upper GFP band relative to hexokinase loading control for three replicates (bottom).

(B) Replicate western blot of *HYR1-GFP* replicates, as in Figure 5E (top) and quantification of GFP relative to hexokinase loading control for three replicates (bottom).
(C) Replicate western blot of *YML020W-GFP* replicates, as in Figure 5F (top) and quantification of GFP relative to hexokinase loading control for three replicates (bottom).
(D) Replicate western blot of $ALA1^{GFP}$ reporter constructs, as in Figure 6A. Xs indicate samples that were not discussed in this study.
(E) Replicate western blots of *YMR31-GFP, YML020W-GFP, HYR1-GFP, CKB2-GFP and FOL1-GFP* with and without *upf1*Δ, as in Figure 6E.
(F) Replicate western blot of *FOL1-GFP* constructs, as in Figure 6I.
(G) Western blot of $ALA1^{GFP}$*-WT* reporter for cells with and without the *pCup-HYP2* construct with copper ($CuSO_4$) addition leading to overexpression of eIF5A for two replicates, which is quantified in Figure 7C.
(H) qPCR fold change of *HYP2* transcript relative to *PFY1* for cells with and without the *pCup-HYP2* construct with and without copper ($CuSO_4$) addition for three replicates. Related to Figure 7C.

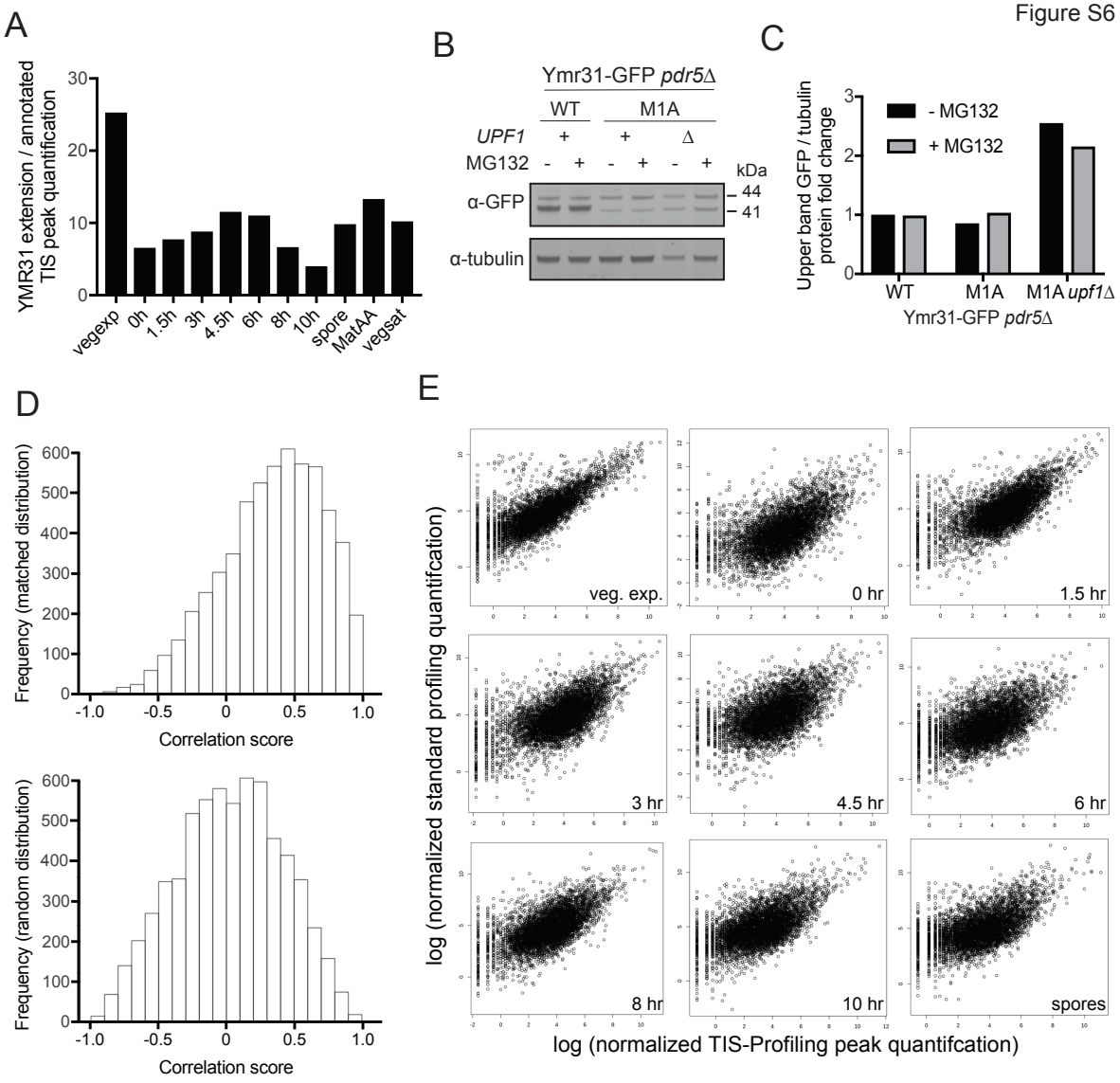

Figure S6: Positive correlation of TIS peaks with gene expression for annotated AUG sites but not near-cognate sites, Related to Figure 4

(A) Quantification of *YMR31* TIS-profiling peaks for the extension peak relative to the annotated peak. For all timepoints, the non-AUG extension peak is higher than the annotated AUG peak.

(B) Western blot of Ymr31-GFP with the proteasome inhibitor MG132. WT, M1A and M1A *upf1Δ* strains were treated with 100 uM MG132 for one hour. All strains are *pdr5Δ* to allow MG132 to enter cells, and samples were taken at 4h in meiosis.

(C) Quantification of the upper GFP band relative to tubulin for Figure S6B.

(D) Distribution of spearman correlation scores for peak height quantification comparing standard and TIS-profiling across all meiotic time points for all annotated genes (top) compared to a matched random distribution set (bottom). The set of annotated genes is significantly enriched for positive correlation scores, as seen by a K.S. test with a p-value of $<2.2 \times 10^{-16}$.

(E) Scatter plots comparing peak quantification of TIS versus standard profiling for each timepoint.

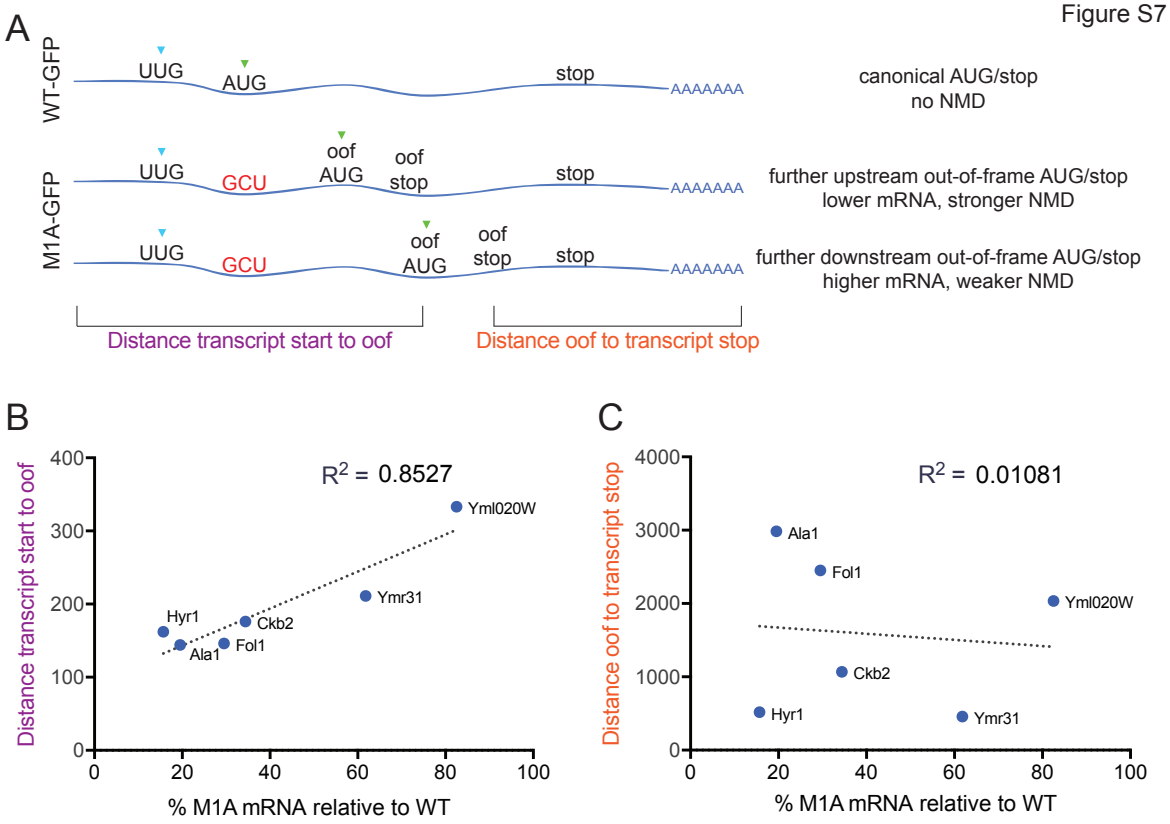

Figure S7: Effect of NMD for M1A transcripts does not correlate with distance from premature stop to transcript end, Related to Figure 6

(A) Diagram of a canonical ORF (*WT-GFP*) compared to two possible *M1A-GFP* constructs where the annotated AUG is mutated, leading to initiation at a later, out-of-frame (oof) AUG. Two different positions of the oof AUG/stop are shown, leading to different outcomes of NMD effect. For the mutated *M1A* construct, two distances are indicated, the distance between the transcript start to the oof AUG/stop (purple), and the distance from the oof AUG/stop to the transcript stop (orange).

(B) Correlation between the distance from the transcript start to the newly created oof ORF relative to the percent of *M1A / WT* mRNA level from Figure 6G, where a lower percentage indicates a stronger NMD effect and a higher percentage indicates a weaker NMD effect. A correlation with an R^2 value of 0.8527 is seen, indicating that a shorter distance from the transcript start to the oof ORF correlates positively with less *M1A* mRNA relative to *WT* and therefore stronger NMD.

(C) Correlation between the distance from the end of the newly created oof ORF to the end of the transcript relative to the percent of *M1A / WT* mRNA level from Figure 6G. A correlation with an R^2 value of 0.01081 is seen, indicating esentially no association between the distance from the oof ORF to transcript stop and the strength of NMD.

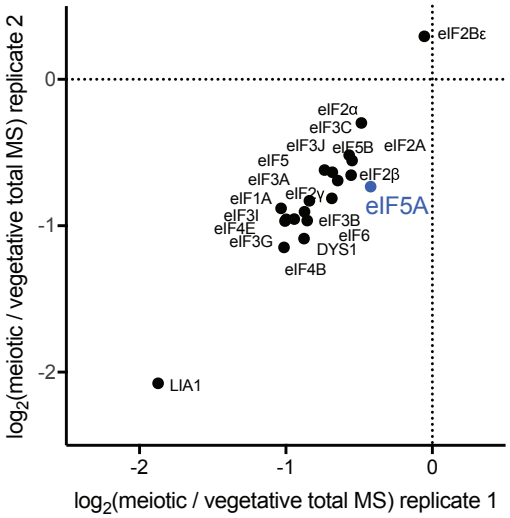

Figure S8: Total protein abundance of initiation and hypusination factors, Related to Figure 7

Enrichment of translation factors (as in Figure 7B) and hypusination factors Lia1 and Dys1 comparing meiotic and vegetative samples for two replicates, determined by quantitative (TMT10) mass spectrometry of whole cell extract from meiotic and vegetative cells.

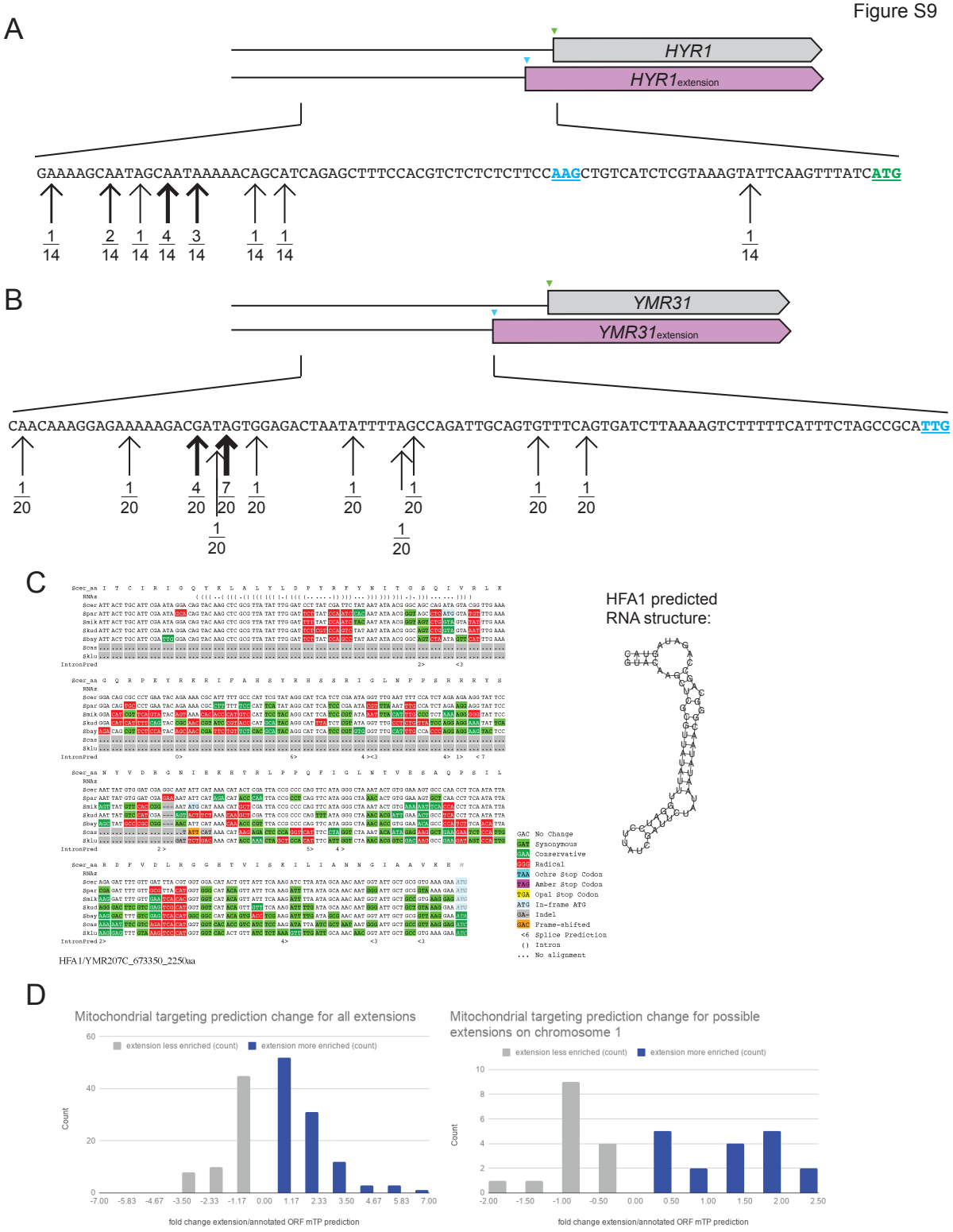

Figure S9: HFA1 RNA structure and mitochondrial targeting sequence prediction, Related to Discussion

(A) 5'RACE analysis of *HYR1.* Locations of transcription start sites are indicated with arrows, with the number of sequencing reads at that site indicated. A total of 14 transcription start sites were sequenced.
(B) 5'RACE analysis of *YMR31*. Locations of transcription start sites are indicated with arrows, with the number of sequencing reads at that site indicated. A total of 20 transcription start sites were sequenced.
(C) Structure prediction for *HFA1*, shown by RNAz depiction in alignment (left), and in predicted structure form (right).
(D) Mitochondrial targeting prediction score changes for extension ORFs relative to the annotated ORF's score (left) and for possible extensions of annotated ORFs on chromosome 1 relative to the annotated ORF's score (right).