Truncated protein isoforms generate diversity of protein localization and function in yeast

Highlights

- 388 truncated isoforms of annotated proteins are translated in budding yeast

- Some derive from full-length transcripts, others from truncated transcripts

- Truncated proteins are regulated and may serve functions that differ from known isoforms

- Differing N termini of truncated isoforms commonly impart distinct cellular localization

Authors

Andrea L. Higdon, Nathan H. Won, Gloria A. Brar

Correspondence

gabrar@berkeley.edu

In brief

Translation initiation at in-frame start codons within annotated ORFs can generate truncated proteins. An algorithm designed to address challenges in the systematic identification of truncated proteins reveals their widespread production in budding yeast and their contributions to the diversification of protein function and localization.

CellPress

Article

Truncated protein isoforms generate diversity of protein localization and function in yeast

Andrea L. Higdon,[1,2] Nathan H. Won,[1] and Gloria A. Brar[1,2,3,*]
[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Center for Computational Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[3]Lead contact
*Correspondence: gabrar@berkeley.edu
https://doi.org/10.1016/j.cels.2024.03.005

SUMMARY

Genome-wide measurement of ribosome occupancy on mRNAs has enabled empirical identification of translated regions, but high-confidence detection of coding regions that overlap annotated coding regions has remained challenging. Here, we report a sensitive and robust algorithm that revealed the translation of 388 N-terminally truncated proteins in budding yeast—more than 30-fold more than previously known. We extensively experimentally validated them and defined two classes. The first class lacks large portions of the annotated protein and tends to be produced from a truncated transcript. We show that two such cases, Yap5^{truncation} and Pus1^{truncation}, have condition-specific regulation and distinct functions from their respective annotated isoforms. The second class of truncated protein isoforms lacks only a small region of the annotated protein and is less likely to be produced from an alternative transcript isoform. Many display different subcellular localizations than their annotated counterpart, representing a common strategy for dual localization of otherwise functionally identical proteins.
A record of this paper's transparent peer review process is included in the supplemental information.

INTRODUCTION

Defining the set of proteins encoded by an organism allows for an understanding of cellular function. This fundamental idea was the motivation for systematic coding sequence prediction immediately following the generation of whole-genome sequences. Initial annotation strategies had known limitations, including difficulty in predicting splice isoforms and a reliance on strict rules to pare down the number of predicted open reading frames (ORFs), such as a minimum length requirement of 100 codons (reviewed in Dinger et al.[1]). Alternative splicing is generally considered to be the major force driving the diversity of protein products from a single locus, but this process is rare in budding yeast (although possibly less rare than historically thought).[2,3] Therefore, a simple model in which one gene encodes one transcript, which is then decoded into one protein product, has been adopted as a general rule in this organism, with only a few isolated exceptions. Genomic techniques, such as transcript isoform sequencing and ribosome profiling, have greatly expanded our understanding of the diversity of transcript and protein products encoded by even very compact genomes like that of budding yeast.[4–8] Ribosome profiling, in particular, has facilitated the identification of a diverse array of translated ORFs, including upstream ORFs (uORFs), intergenic short ORFs (sORFs), and N-terminally extended protein isoforms.

Meiosis in budding yeast is an excellent system for identifying fundamental principles of genome decoding and regulated gene expression. The process of differentiation from a diploid progenitor into haploid gametes requires an intricate and precisely timed series of cellular remodeling events (reviewed in Marston and Amon[9] and van Werven and Amon[10]). Underlying these dramatic changes is a gene expression program that requires dynamic regulation of most of the yeast proteome.[4] Ribosome profiling and sequencing of the 5′ ends of transcripts show that production of non-canonical protein products and alternative transcripts is particularly prevalent during meiosis relative to mitotic growth.[4,5,11] The specific functional relevance of all but a few non-canonical gene products, however, remains unclear. N-terminally truncated proteins, which initiate from downstream in-frame start codons within annotated genes, are a particularly interesting class. Since they are variants of existing proteins, they would seem likely to have molecular function, but their identification has remained difficult by sequencing, conservation, and proteomics-based methods due to specific technical challenges arising from their complete overlap with annotated coding regions. In the case of proteomic approaches, it is difficult to distinguish peptides derived from the truncated and annotated isoforms because the truncated isoform has only one unique tryptic peptide (at the N terminus) relative to the annotated isoform. Conservation analyses, which are valuable for predicting functionally important translation products of many classes, are challenging to apply to the identification of truncated proteins due to both the high background signal expected from conservation of the fully overlapping annotated protein and the lack of synonymous codons for methionine, which leads to high conservation of in-frame AUGs irrespective of their usage as a start codon.

Cell Systems
Article

Table 1. Previously characterized genes with N-terminally truncated protein isoforms

Gene	Function	Called in TIS profiling?	Citation(s)
CCA1 (YER168C)	tRNA CCA addition	yes, no[a]	Wolfe et al.[20]
CRS1 (YNL247W)	cysteinyl-tRNA synthetase	yes	Nishimura et al.[21]
FUM1 (YPL262W)	fumarase	yes	Wu and Tzagoloff[22]
GAT1 (YFL021W)	transcription factor	yes	Rai et al.[23]
GLR1 (YPL091W)	glutathione oxidoreductase	yes	Outten and Culotta[24]
GRX2 (YDR513W)	glutaredoxin	yes	Pedrajas et al.[25]; Porras et al.[26]
KAR4 (YCL055W)	mating factor, mRNA methyltransferase	yes	Gammie et al.[27]
LEU4 (YNL104C)	leucine biosynthesis	yes	Beltzer et al.[28,29]
MOD5 (YOR274W)	tRNA modification	yes	Boguta et al.[30]
SUC2 (YIL162W)	invertase	yes	Carlson and Botstein[31]; Taussig and Carlson[32]
VAS1 (YGR094W)	valyl-tRNA synthetase	no[b]	Chatton et al.[33]
Other relevant genes excluded from validation set			
CCC1 (YLR220W)	vacuolar transporter	not expressed in WT cells	Amaral et al.[34]
HTS1 (YPR033C)	histidine-tRNA synthetase	second isoform is an AUG extension relative to genome annotation	Natsoulis et al.[35]
MRK1 (YDL079C)	glycogen synthase kinase	start is in intron; intronic sequences not considered in our algorithm	Zhou et al.[36]
TRM1 (YDR120C)	tRNA modification	second isoform is an AUG extension relative to genome annotation	Rose et al.[37]

Truncated proteins found by individual previous studies in budding yeast. The top set was used to validate our truncation-calling algorithm. The bottom set was excluded from analysis for the reasons stated.
[a]Two truncated isoforms, one called and one not called but visible by eye in genome browser.
[b]Truncation would have been called but was filtered out because main ORF not called.

Here, we report an algorithm that robustly and sensitively identified hundreds of N-terminally truncated isoforms using data from a modified version of ribosome profiling for identifying translation initiation sites. In addition, we describe two distinct types of regulation responsible for their production and provide experimental evidence for their existence along with insights into their functions. A handful of N-terminally truncated proteins have been validated in several organisms, including humans, and ribosome-profiling-based analyses suggest they might be common in mammalian cells.[7,12–19] Although several individual examples of N-terminally truncated proteins were identified in previous single-gene studies in yeast (Table 1), our genome-wide approach, and subsequent experimental validation, sheds light on their prevalence in this heavily studied model organism. Most of the 388 truncations we identified are dynamically regulated during the meiotic program or under other non-standard laboratory growth conditions. The compactness of the yeast genome and minimal use of splicing have led to the presumption that yeast do not widely use alternative protein isoforms. These very features facilitated our robust global analyses and investigation of their production.

We classified truncations into two broad classes that generally differentiate regulatory and functional characteristics of the isoforms. The first class of truncations lacks a large portion of the annotated protein's N terminus ("distal" truncations) and tends to be encoded by a truncated transcript isoform. We identify and characterize such two examples, produced from the *PUS1*

and *YAP5* loci, whose conserved and well-characterized annotated isoforms encode a pseudouridine synthetase and an activator protein 1 (AP-1) transcription factor, respectively. The truncated isoforms are expressed in a condition-specific manner and lack key domains, resulting in functions that seem distinct from the full-length proteins. The second class of truncations begins closer ("proximal") to the annotated start codon. Proximal truncations are more often encoded by the same transcript as the annotated isoform, likely requiring bypass of the annotated start codon for their translation and allowing simultaneous production of the annotated and truncated isoform from a single transcript. Based on our extensive computational and experimental investigation, we posit that a common role for these truncations is in diversifying the subcellular localization of the encoded protein. We demonstrate that our predictions in this respect are remarkably robust, revealing a case in which two truncations at one locus allow three distinct and simultaneous subcellular localizations. Thus, our study elucidates the potential of truncated protein isoforms to provide proteome diversity and cellular function well beyond what was previously recognized.

RESULTS

Truncated protein isoforms are prevalent in budding yeast

N-terminally truncated protein isoforms initiate at in-frame start codons within canonical genes and therefore share a common
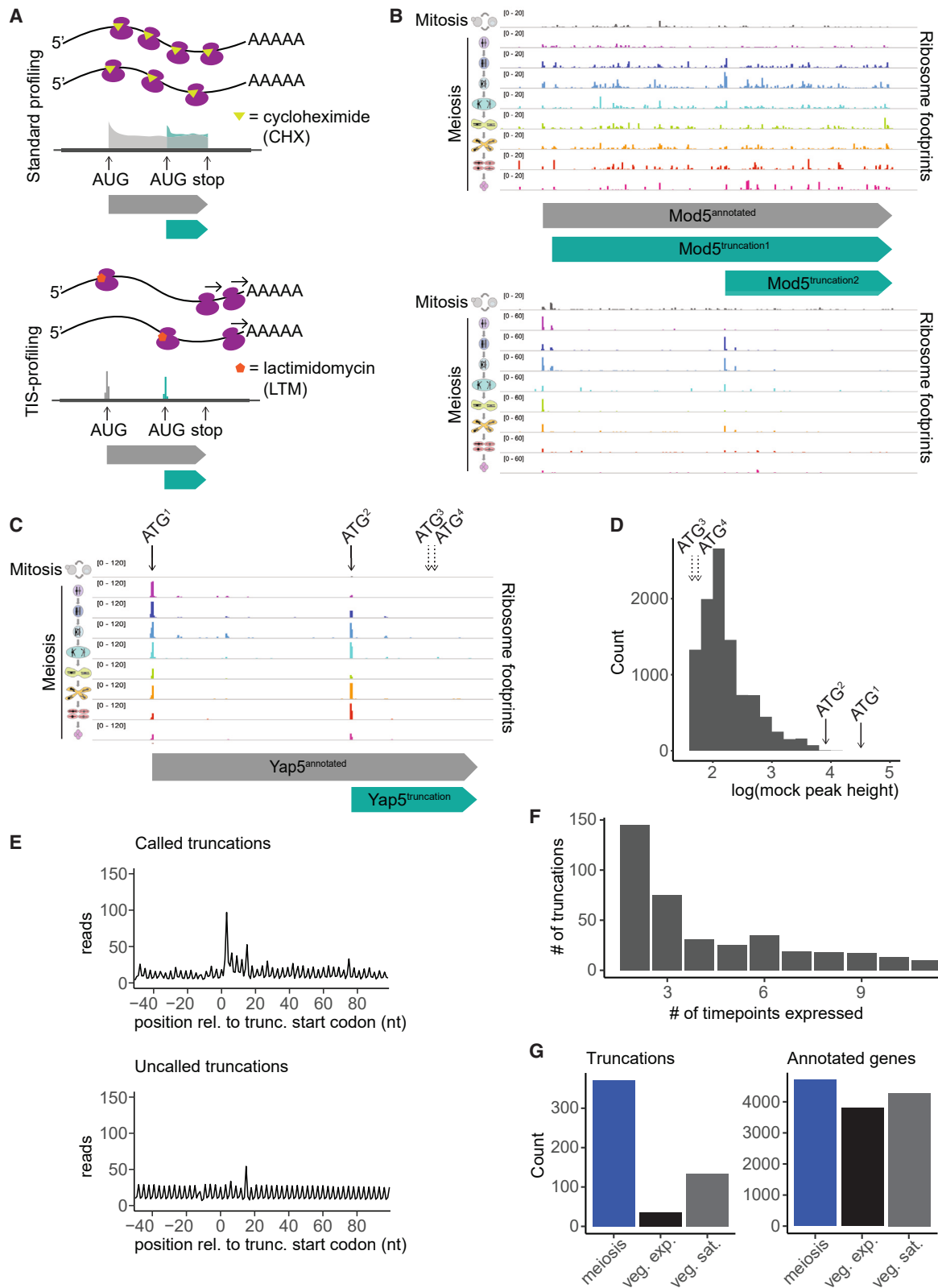
Figure 1. Genome-wide identification of truncated protein isoforms using TIS-profiling data

(A) Schematic comparing standard ribosome profiling (top) and TIS profiling (bottom). In each case, a cartoon of ribosomes translating two mRNAs is shown on top, and sample read density is shown below for a locus containing an annotated and truncated protein isoform.

(legend continued on next page)

C-terminal sequence with their annotated isoform. These truncated isoforms are more challenging to identify than many other types of translated regions. Standard ribosome profiling data, which have enabled global maps of translated regions, report the positions of elongating ribosomes and therefore yield reads mapping across the entire gene, including the entirety of the potential truncated isoform, masking signal for truncated isoforms (Figure 1A). We previously published a translation initiation site (TIS)-profiling dataset with time points spanning stages of meiotic progression in budding yeast cells with the goal of identifying all types of non-canonical translation initiation sites, with a particular focus on those producing N-terminally extended protein isoforms.[6] In contrast to standard ribosome profiling, which typically employs the translation elongation inhibitor cycloheximide (CHX), TIS profiling uses the drug lactimidomycin (LTM) to capture ribosomes immediately after initiation while allowing elongating ribosomes to complete translation (or "run off"; Figure 1A).[7,38] The resulting ribosome footprints are therefore highly enriched at sites of translation initiation, with minimal background reads across the ORF. TIS profiling therefore yields clear initiation-site peaks that are much easier to detect than with standard ribosome profiling. For example, at the *MOD5* locus, we detect both a previously identified truncated isoform (Mod5[truncation1]) and an additional previously unidentified truncated isoform (Mod5[truncation2]) using TIS profiling (Figure 1B). Notably, neither truncated isoform is evident from standard ribosome-profiling data, demonstrating the power of this method for detecting internal translation start sites.

We previously used the program ORF-RATER to call all types of ORFs, including those that are non-canonical, using standard ribosome-profiling and TIS-profiling data collected in vegetative (mitotic exponential and saturated) conditions and 8 time points spanning the major developmental stages of meiosis.[6,16] Although this algorithm performed very well for many types of ORFs, including those that were 5′ extended, it was unable to identify many truncated isoforms that were extremely clear in the TIS-profiling data, such as at the *YAP5* locus (ATG[2], Figure 1C).[16] We hypothesized that this could be due to overweighting of the standard ribosome profiling and underweighting of the TIS-profiling data by the algorithm. It has also been observed that many ribosome profiling algorithms that are trained on annotated ORFs are unable to perform as well with shorter ORFs.[39] To avoid systematically

biasing against sensitive detection of translation initiation sites within annotated ORFs, we developed an algorithm specifically designed for identifying truncated isoforms that relies solely on TIS-profiling data.

Briefly, the goal of this algorithm is to robustly interpret TIS-profiling data by separating true initiation-site signal from the background noise present across translated genes, likely due to a low degree of elongation inhibition by LTM. To model the background signal for each gene at each time point, we randomly sampled three single nucleotide positions from within the annotated ORF and summed their mapped reads to simulate the three nucleotides of a start codon, over which a prominent peak should be present for real translation-initiation sites. Re-sampling 10,000 times provided an empirical null distribution representing the distribution of peak heights that would be expected from the background noise within each gene at each time point (as shown for the *YAP5* locus at 0 h in sporulation media [SPO], Figure 1D). We next determined the observed peak height for all in-frame (potentially truncation-generating) start codons within each annotated gene and used the corresponding empirical distribution to assign a *p* value to each putative truncation initiation site, considering only those with a Benjamini-Hochberg false discovery rate (FDR) of <0.005. In the case of *YAP5*, there were three in-frame start codons within the annotated ORF, two of which fall within the range of peak heights expect from background and are not called (ATG[3] and ATG[4]) and one which was called as significantly above background (ATG[2]; Figures 1C and 1D).

To increase the stringency of our calls, we required that a TIS be called at 2 or more time points, which were typically consecutive, suggesting a high degree of regulation (Table S1). We also required that a truncated isoform begin 5 amino acids (aa) or more from the annotated start codon and be no less than 10 aa in total length. Truncated isoforms for which we could not detect the annotated isoform at any time point were also excluded, as these often represent cases of mis-annotation, where the "truncated" isoform is, in fact, the main isoform. Additional filtering criteria were applied, as described in the methods. It is also important to note that in defining isoform identity, we made the assumption that all truncated isoforms use the same stop codon as the annotated isoform and that there are no differences in propensity for stop codon readthrough between annotated and truncated isoforms.

(B) Standard ribosome-profiling (top) and TIS-profiling (bottom) data for the *MOD5* locus, with the annotated open reading frame in gray and the two truncations identified by our algorithm in turquoise (middle). Cartoons to the left indicate whether the track represents mitotic cells (gray) or meiotic time points (rainbow). Note that the two truncations are not apparent from the standard ribosome-profiling data, but robust peaks at multiple time points are clear in the TIS-profiling data. Scale is linear and represents reads.

(C) TIS-profiling data for the *YAP5* locus. Arrows indicate all in-frame ATGs within the ORF. Cartoons to the left indicate whether the track represents mitotic cells (gray) or meiotic time points (rainbow). Scale is linear and represents reads.

(D) Empirical distribution representing background read density within the *YAP5* annotated ORF at a single TIS-profiling time point (0 h). Arrows indicate approximate read density for called start sites (solid arrows, ATG[1] and ATG[2]) and uncalled in-frame start codons (dashed arrows, ATG[3] and ATG[4]).

(E) Metagene plot of standard ribosome-profiling data for all called truncated isoforms (top) and uncalled controls (bottom) for the region between −50 and +100 nt relative to the truncation start codon. Reads are summed across all time points. For called truncations, downstream (+20 to +70 nt) read density is significantly higher than upstream (−50 to −1 nt) read density ($p < 0.01$, Mann-Whitney U test). For uncalled truncations, there was not a significant difference in upstream and downstream read densities (n.s., Mann-Whitney U test).

(F) Histogram of number of time points at which truncated isoforms are expressed in TIS-profiling data.

(G) Number of truncated isoforms (left) and number of annotated isoforms (right) called in samples from meiotic time points, vegetative exponential growth, and vegetative saturated growth. The variation in the number of called isoforms between conditions is significantly more pronounced for truncated isoforms than for annotated isoforms ($p < 2.2e-16$, Fisher's exact test).

Using this stringent approach, we identified 388 truncated protein isoforms (Table S1). Although the existence of truncated protein isoforms has been known for decades due to observations from single-gene studies, this represents a substantial increase in the number of known cases. To our knowledge, in *Saccharomyces cerevisiae* (*S. cerevisiae*), 12 truncated isoforms (2 of which are present at the same locus, *CCA1*) have been characterized that we would expect to be present in our dataset as well (Table 1). Of these previously known truncations, 10 (83.33%) were called by our algorithm. Of the two that were not called, one was called by the algorithm but filtered out because its annotated isoform was not called (at the *VAS1* locus). The other (the shortest isoform of Cca1) was visible by eye in the TIS profiling data but was not called by our algorithm. We concluded that our approach was sensitive to identifying translation of truncated isoforms and that they are much more prevalent than previously known, with our set of newly identified truncations representing a more than 30-fold increase over the set of characterized truncations.

Since our algorithm used only TIS profiling data to identify truncated isoforms, we were able to evaluate the quality of our calls using matched standard ribosome profiling data. In these data, we would expect each truncated isoform to have a peak in reads at the start codon, followed by slightly elevated read density downstream relative to upstream, since downstream read density should include the contributions of elongating ribosomes associated with both the annotated and truncated isoforms (Figure 1A). We performed metagene analysis for the regions surrounding the predicted start codon for all 388 truncations and indeed saw the expected read density patterns across the gene set (Figure 1E, upper). Read densities downstream of the start codon (+20 to +70 nt; excluding the poorly quantitative region immediately following the start codon) were higher those upstream (-50 to -1 nt), consistent with elevated ribosome footprint density corresponding to the translation of truncated isoforms ($p < 0.01$, Mann-Whitney U test). Notably, this trend was not seen for in-frame start codons that were not called by our algorithm (n.s., Mann-Whitney U test; Figure 1E, lower). Metagene profiles of the TIS-profiling data also showed the expected trends, with a sharp peak present at the start codon for called truncations and virtually no signal at the start codon for uncalled truncations (Figures S1A and S1B). Supporting our assumption that stop codon readthrough does not affect predicted translation products at these sites, metagene analysis of the region downstream of the stop codons for all 388 truncations revealed an absence of reads, paralleling what we observe genome-wide in yeast under mitotic and meiotic conditions (Figure S1C).[4] Together, these data indicate that our truncation-calling approach detects true translation events.

Truncated isoforms are dynamically expressed and enriched in meiosis

Among our set of called truncated isoforms, we observed that many appear to be dynamically expressed during meiosis. At the *MOD5* locus, for example, the smaller truncation (Mod5[trunc.2]) is not translated in exponentially growing mitotic cells but is upregulated specifically during early meiosis (Figure 1B). Over 50% of truncated isoforms in our dataset were called at only two or three

time points, indicating that dynamic expression is very common (Figure 1F). Truncated isoforms were also much more common in meiosis and slightly more common in vegetative saturated growth than in vegetative exponential growth, the most common laboratory growth condition (Figure 1G). Since we analyzed multiple samples collected during meiotic progression but only one each during vegetative exponential growth and vegetative saturated growth, this enrichment could have been due to increased power to detect ORFs during meiosis. However, this pattern was significantly less strong among annotated protein isoforms called by our algorithm, indicating that dynamic meiotic expression of truncated protein isoforms is a true biological phenomenon ($p < 2.2e-16$, Fisher's exact test).

Newly predicted truncated protein isoforms can be detected *in vivo*

To evaluate the quality of our TIS-profiling data and truncation-calling algorithm, we experimentally validated the production of several newly identified truncated proteins. Due to the limitations in resolving similarly sized proteins, we focused on candidates that differed enough in size to be distinguishable from their annotated isoform by western blotting. We integrated a C-terminal epitope tag at the endogenous locus of each protein, such that both the annotated and truncated isoforms should be tagged. We then collected samples at time points throughout meiosis and performed western blotting. Ten predicted truncated isoforms that we examined were clearly detectable by this method, indicating that the truncated proteins are expressed and stable. Some of these truncated protein isoforms display dynamic expression patterns distinct from their annotated isoform, including Yap5[trunc.], Mod5[trunc.], Pex32[trunc.], Pus1[trunc.], Sas4[trunc.], Ssp1[trunc.2], Tpo1[trunc.], and Prp4[trunc.] (Figures 2A–2F, S2A, and S2B). Others, like Ssp1[trunc.1] and Yck1[trunc.1], display expression patterns that mirror the annotated isoform (Figures 2F and 2G).

Although the detected truncated isoforms migrate according to their expected size, it remained possible that the bands could be degradation products of the annotated isoform rather than the product of translation at the predicted alternative in-frame start codon. To test this possibility, we generated strains for two examples, Pus1 and Yap5, for which the ATG start codons for the annotated or predicted truncated isoform were mutated to ATT (isoleucine) to abrogate expression (M[ann.]L and M[trunc.]L, respectively; Figure 2H). Western blot analysis of these strains revealed translation of only the annotated isoform in cells carrying the M[trunc.]L mutation and only the truncated isoform in cells carrying the M[ann.]L mutation, indicating that the truncated isoforms for each gene are indeed the product of translation initiating at the newly predicted start codon internal to the annotated ORF (Figures 2I and 2J).

Our ability to detect most of the predicted truncated isoforms that we tested by western blotting suggests our approach has a low rate of false positives. However, 5 out of 15 predicted truncations that we tested were not visible by western blotting of C-terminally tagged proteins (Yck1[trunc.2], Glk1[trunc.], Ari1[trunc.], Rtt105[trunc.], and Siw14[trunc.]; Figures 2G and S2C–S2F). There are several possible explanations for this: (1) stability of the truncated protein may not be compatible with the epitope tag, (2) they may be challenging to detect for technical reasons, for example, due to low expression or small size, (3) they could be false
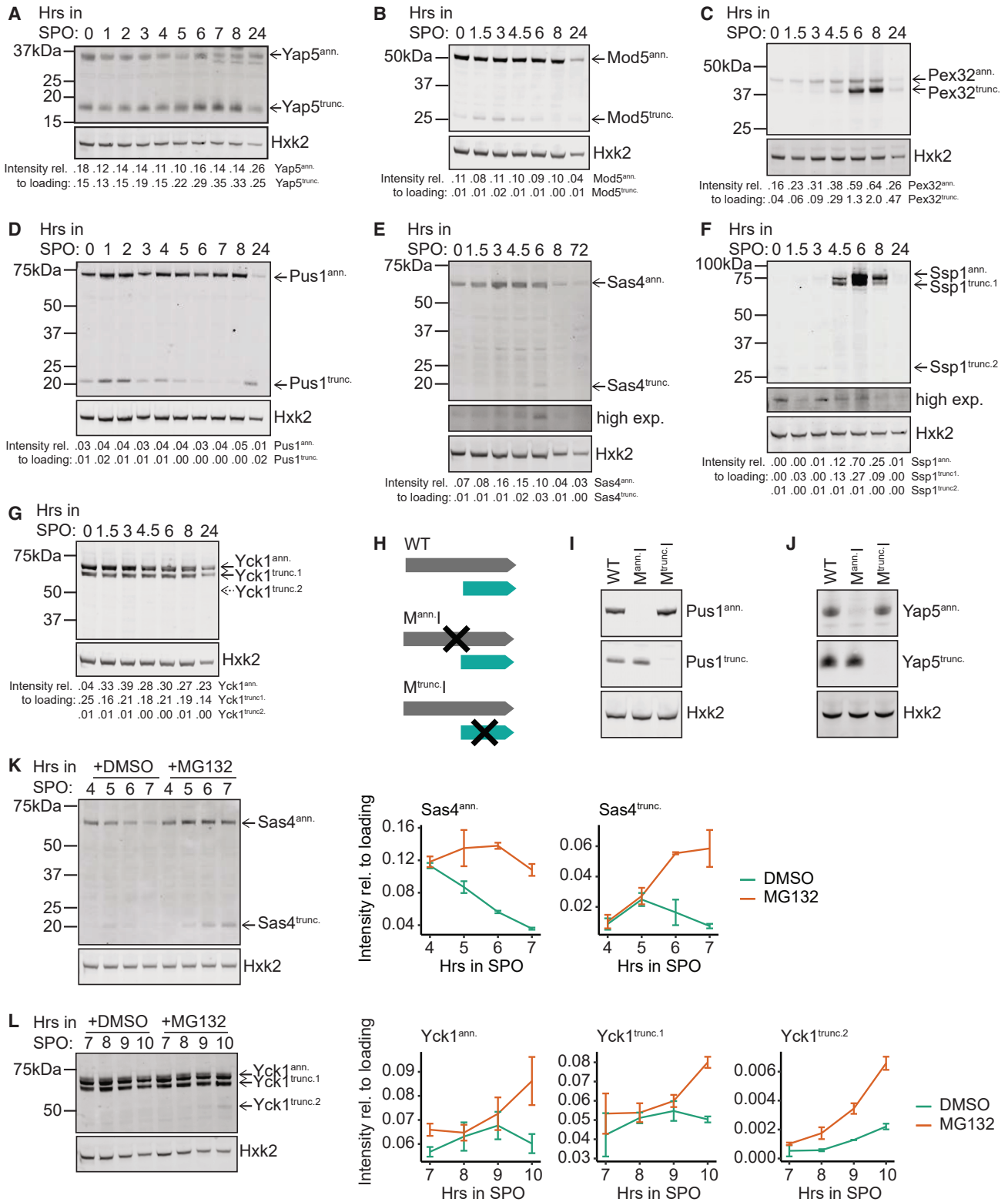
(legend on next page)

positives, or (4) they may be produced but then degraded under normal conditions. To test whether protein degradation was preventing our detection of some truncated protein isoforms, we treated cells carrying C-terminal epitope tags of predicted truncations with the proteasome inhibitor MG132. We timed MG132 treatment and sample collection for western blotting according to the expected timing of expression for each truncation based on the TIS data. We included examples in which the truncation was not detected (Yck1[trunc.2], Ari1[trunc.], Rtt105[trunc.], and Siw14[trunc.]) as well as ones in which the truncation was visible but present at low abundance (Sas4[trunc.], Tpo1[trunc.], Mod5[trunc.], and Prp4[trunc.]). We observed increased abundance, indicating increased stabilization, for previously detectable truncations, including Sas4[trunc.] and Tpo1[trunc.] (Figures 2K and S2G). The previously undetectable second truncation of Yck1[trunc.2] also became visible upon proteasome inhibition, suggesting that this truncation is normally translated but degraded by the proteasome to levels below detection (Figure 2L). Abundance of other truncated proteins, including Mod5[trunc.] and Prp4[trunc.] (Figures S2H and S2I), was minimally affected by proteasome inhibition, suggesting that proteasome-mediated degradation was not the reason for their weak detection by western blotting. Three of the previously undetectable truncations (Ari1[trunc.], Rtt105[trunc.], and Siw14[trunc.]) remained undetectable upon proteasome inhibition (Figures S2J–S2L).

Distal truncations are typically produced from a truncated transcript, while proximal truncations are likely regulated by translational control

Many truncated isoforms are dynamically regulated, some in concert with their annotated isoform and others independently. There are two straightforward models to explain such regulation: (1) translational control, in which the ribosome bypasses one or more in-frame start codons in favor of a downstream one and (2) transcriptional control, in which a truncated transcript encodes the truncated protein and translation initiation occurs at the first in-frame start codon (Figure 3A). To determine the extent of these two possible types of regulation for the hundreds of truncated proteins newly identified by our algorithm, we leveraged a published transcript leader sequencing (TL-seq) dataset collected across meiosis in the same strain background as our TIS-profiling data (SK1).[5] Using these data, we were able to determine whether there is evidence of a 5′ transcript end upstream of our called truncation TISs but downstream of their respective annotated TISs. For *YCK1*, for example, two truncated isoforms (Yck1[truncation1 and 2]) are produced, but there is no evidence of a separate transcript isoform in the TL-seq data, consistent with translational control (Figure 3B). For *YAP5*, by contrast, a clear transcription start site (TSS) is present just upstream of the TIS for the truncated isoform, supporting the transcriptional control model (Figure 3C).

To assess the prevalence of truncated transcripts corresponding to truncated proteins genome-wide, we constructed TL-seq metagene profiles for the regions upstream of translation initiation sites (Figure 3D). For annotated genes, this displays the expected profile of high read density peaking around 50 nt upstream of the TIS, consistent with the average length of a yeast 5′ leader.[40] The metagene for truncated isoforms displays a highly similar profile, indicating that it is common for truncated proteins to be produced from truncated transcripts (Figure 3D). Notably, for this analysis, we excluded truncated isoforms whose annotated isoform would start within the window included in the metagene profile (−200 nt) such that the metagene profile should not include any reads derived from annotated transcript isoforms.

We hypothesized that regulation via truncated transcripts might be more common for truncations that initiate far (distal) from the annotated isoform, since control at the translational level would in most cases require the ribosome to bypass several in-frame start codons, making transcriptional regulation a more

---

Figure 2. Many newly identified truncated protein isoforms can be confirmed by western blotting, and some are stabilized by proteasome inhibition

(A–G) Western blots of samples collected at various time points after transfer of cells to sporulation media (SPO). Hexokinase (Hxk2) is shown as a loading control. For lowly expressed truncated isoforms, a high exposure panel is shown (high exp.). A different strain is used in each case, expressing the indicated C-terminal epitope-tagged protein and enabling the detection of annotated (ann.) and truncated (trunc.) isoforms shown in (A)–(G).

(A) Yap5-FLAG.

(B) Mod5-3V5.

(C) Pex32-3V5.

(D) Pus1-3V5.

(E) Sas4-3V5.

(F) Ssp1-3V5, for which two truncated isoforms were predicted and detected.

(G) Yck1-3V5, for which two truncated isoforms were predicted but only one was successfully detected. Solid arrows indicate isoforms that were predicted and detected. Dashed arrows indicate predicted but undetected isoforms.

(H) Schematic of mutagenesis approach used to validate the production of truncated isoforms from predicted start codons in (I) and (J). Top image indicates the wild-type context (WT), in which both isoforms should be seen. Either the annotated (middle, M[ann.]I) or predicted truncation (bottom, M[trunc.]I) start codon was mutated from encoding methionine (ATG) to isoleucine (ATT) to prevent their translation.

(I) Western blot of start codon mutant strains described in (H). Hexokinase (Hxk2) is shown as a loading control. Images of bands for the annotated and truncation isoforms are from the same blot. Blotting is for Pus1-3V5. Cells were collected at 24 h post-dilution in YEPD.

(J) Western blot of start codon mutant strains described in (H) for Yap5. Hexokinase (Hxk2) is shown as a loading control. Images of bands for the annotated and truncation isoforms are from the same blot. Blotting is for Yap5-FLAG. Cells were collected at 4.5 h in SPO.

(K) Representative western blot (left) and quantification (right) for cells treated with proteasome inhibitor MG132 or vehicle control DMSO, showing stabilization of both the truncated and annotated isoforms of Sas4-3V5 for cells in meiosis. Note that quantification is based on 2 replicates, and error bars represent standard error. Hexokinase (Hxk2) is shown as a loading control.

(L) Representative western blot (left) and quantification (right) for cells treated with proteasome inhibitor MG132 or vehicle control DMSO, showing stabilization of the truncated and annotated isoforms of Yck1-3V5 for cells in meiosis. Note that quantification is based on 2 replicates, and error bars represent standard error. Hexokinase (Hxk2) is shown as a loading control.
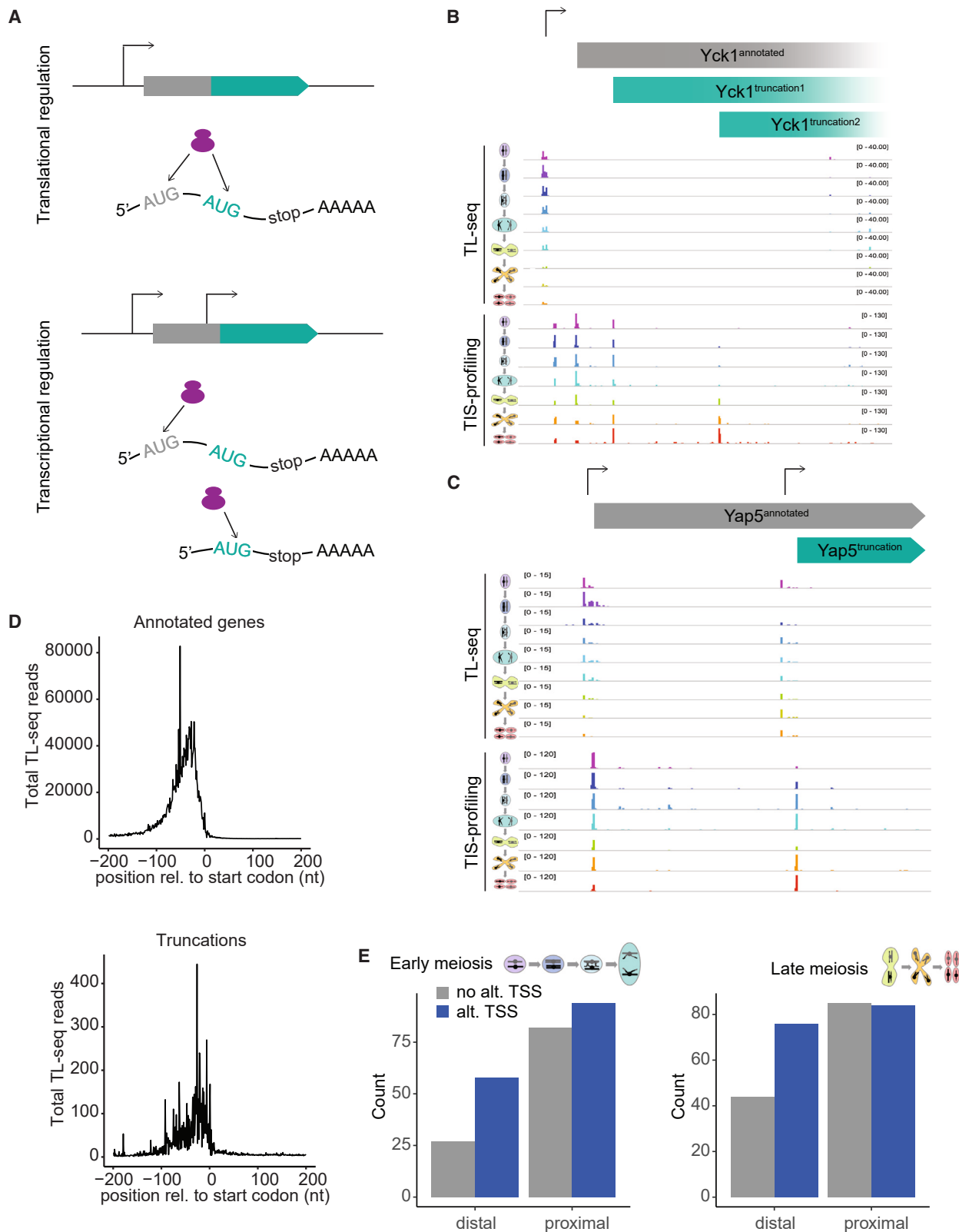
Figure 3. Truncated protein isoforms are often, but not always, produced from truncated transcript isoforms

(A) Schematic of potential regulatory mechanisms for truncated protein isoform production: (1) translational regulation (top) in which a single transcript isoform is produced, and protein isoform production is determined by initiation site selection; (2) transcriptional regulation (bottom) in which an annotated and truncated

parsimonious explanation. Truncated isoforms starting close (proximal) to the annotated start site, however, would be more likely to only require bypassing of the annotated start codon, which could easily occur due to leaky scanning.[41] For purposes of this comparison, we defined proximal truncations as starting within 40 aa of the annotated start codon and distal truncations as starting greater than 40 aa from the annotated start codon (Table S1; Figure S1E). To determine which truncated isoforms had evidence of a corresponding truncated transcript, we used TL-seq data to calculate a "TSS score," which is the ratio of the sum of TL-seq reads 200 nt upstream of the truncation TIS over the sum of reads 200 nt downstream of the truncation TIS. A higher ratio indicates stronger evidence of an alternative transcription 5′ end (likely generated by a downstream transcription start site) that is close to the truncation TIS. To assess statistical significance, we compared each ratio to an empirical distribution of TSS scores created by randomly sampling 200 nt windows from within the same gene. Due to the differences in staging between the two meiotic time courses that generated the TIS profiling and TL-seq data, we were unable to perform high-resolution time point matching between the time courses. However, we were able to achieve a level of temporal resolution that was still compatible with the differences in staging by splitting time points into either early or late meiosis based on correlation coefficients between time points in matched mRNA sequencing (mRNA-seq) data for each time course, as well as expression patterns of key meiotic genes (Figures S3A and S3B). From this analysis, we observed that distal truncations were much more likely to have an additional TSS than proximal truncations, with approximately two-thirds of distal truncations showing evidence of transcriptional regulation via a truncated transcript ($p < 0.05$ for both early and late meiotic groups, Fisher's exact test, Figure 3E; Table S2).

Truncated protein isoforms of Yap5 and Pus1 exhibit condition-specific regulation

To better understand the functional relevance of distal, transcriptionally regulated truncations, we performed in-depth characterization of two examples produced from the *YAP5* and *PUS1* loci. These candidates were selected for further study due to the very well-characterized functions of their annotated isoforms and the tractability of functional hypotheses for their truncated isoform. TIS-profiling data showed a very sharp peak at an in-frame start codon within the *YAP5* locus, indicating translation of a truncated protein isoform (Yap5[truncation]) across all meiotic time points (Fig-

ure 1C). Western blot analysis confirmed that the truncation is expressed across meiosis but is low during vegetative exponential growth (Figures 2A and 4A). Matched TL-seq data show evidence of a transcript isoform with its 5′ end positioned approximately 30 nt upstream of the Yap5[truncation] start codon, suggesting that the truncated protein isoform is produced from an alternative transcript (Figure 3C). Given the dynamic regulation of Yap5[truncation], we wondered whether it had a role related to that of the full-length protein. Annotated Yap5 is an iron-response transcription factor that upregulates its target genes upon exposure to elevated iron in order to mitigate iron toxicity.[42] It constitutively occupies its target promoters and activates target gene transcription upon binding to Fe-S clusters in high iron conditions. The DNA binding domain of Yap5 is in the N-terminal half of the protein and is not present in Yap5[truncation]; the Fe-S cluster binding domain lies in the C-terminal half and remains intact in the truncated isoform.[42] Previous work with artificial truncations of Yap5 showed that a version containing a comparable C-terminal region alone is capable of binding Fe-S clusters, suggesting that the natural truncation is also capable of this function.[43] We therefore hypothesized that Yap5[truncation] may play a role in Fe-S cluster homeostasis.

Fe-S clusters are key cofactors in the electron transport chain and thus important for mitochondrial function. It is therefore notable that Yap5[truncation] is induced during meiosis, a condition requiring respiration, but is lowly expressed in mitotic growth conditions, which favor fermentation. We also observed increased Yap5[truncation] expression in mitotic cells grown to saturation in rich media, a condition in which cells have undergone the diauxic shift from fermentation to respiration upon exhaustion of their fermentable carbon sources (Figure 4A). We wondered whether the Yap5 truncation would also be induced in other respiratory conditions. To test this, we grew cells in dextrose-containing (fermentable) media, and after 4 h of growth, we split the cells into either dextrose- or glycerol-containing (non-fermentable) media. We then assayed Yap5[truncation] production by western blotting and observed upregulation of Yap5[truncation] upon the switch to non-fermentable media (Figure 4B). These experiments suggest that dynamic production of Yap5[truncation] occurs as part of a cellular response to conditions of increased respiratory activity.

Experimental validation, as well as comparison to TL-seq and standard ribosome-profiling data, support the robustness of our approach for identifying hundreds of previously unidentified truncations, a class of non-canonical coding region that has

---

transcript isoform are translated into annotated and truncated protein isoforms, respectively. In each case, a schematic of the locus is shown, with bent arrows representing transcription start sites above and resulting transcript(s) shown below.

(B) TL-seq (above) and TIS-profiling (below) data for the *YCK1* locus showing production of the annotated isoform and two truncated protein isoforms but evidence for only the canonical transcript isoform. Cartoons at left represent meiotic time points (rainbow). Bent arrow represents the predicted transcription start site. The start peak upstream of the annotated start codon corresponds to an upstream open reading frame (uORF). Scale is linear and represents reads.

(C) TL-seq (above) and TIS-profiling (below) data for the *YAP5* locus showing annotated and truncated transcript and protein isoform production. Cartoons at left represent meiotic time points (rainbow). Bent arrows represent predicted transcription start sites. Scale is linear and is based on reads per million (RPM).

(D) Metagene plots of total TL-seq reads for the window −200 to +200 nt surrounding translation-initiation sites for annotated (above) and truncated (below) protein isoforms. Truncations with an annotated start beginning within the −200 nt window were excluded to avoid including reads derived from annotated transcripts.

(E) Bar plot of number of distal (>40 aa from annotated start) and proximal (≤40 aa from annotated start) truncated isoforms, with evidence of an alternative transcription start site ("alt. TSS") or not ("no alt. TSS") based on interpretation of 5′ ends in approximately stage-matched TL-seq data collected in either early (left) or late (right) meiosis. Distal isoforms are significantly more likely to show evidence of an alternative transcript isoform ($p < 0.05$ for both early and late groupings, Fisher's exact test).
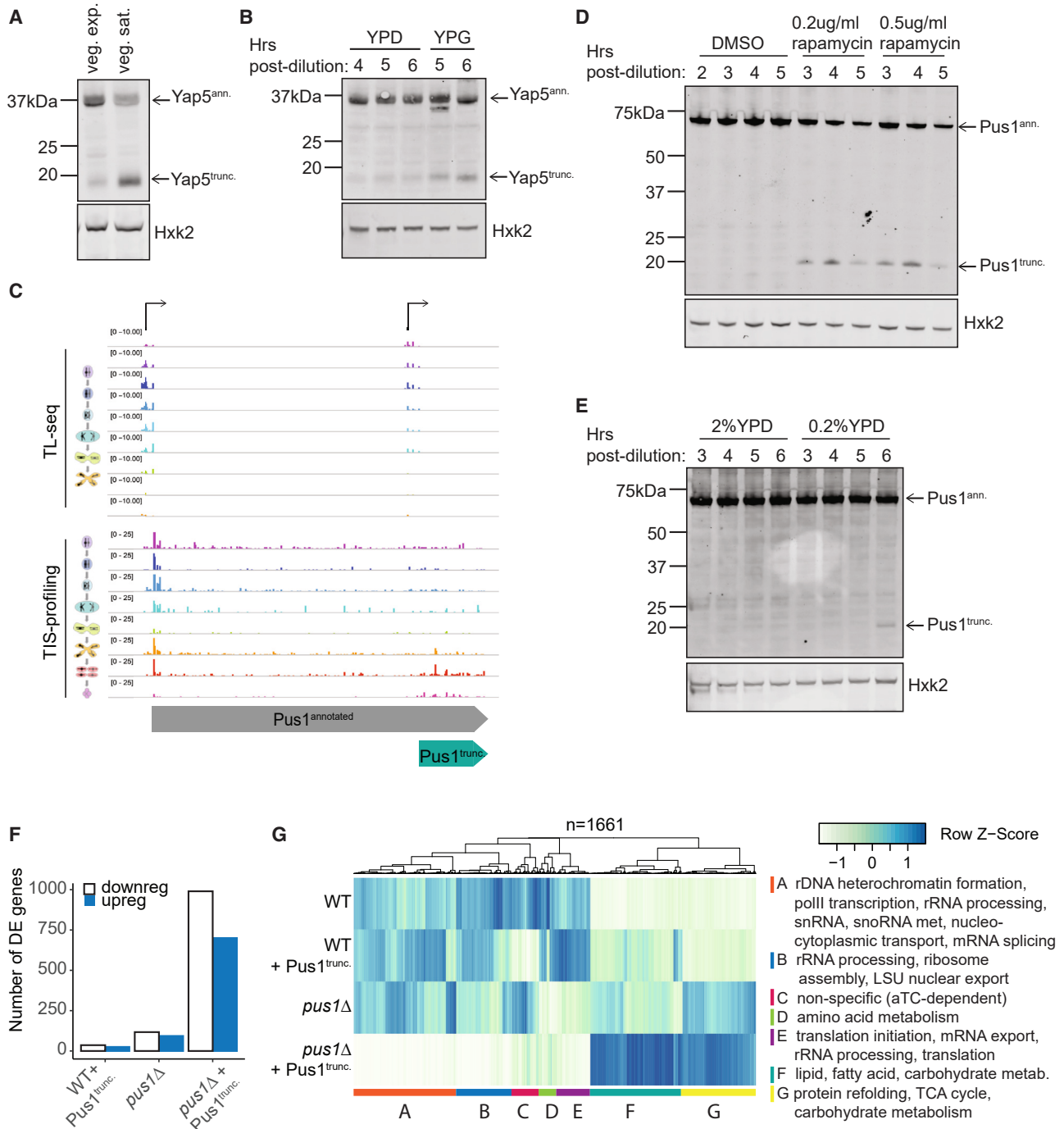
Figure 4. Condition-specific regulation of distal truncated protein isoforms

(A) Western blot for Yap5-FLAG. Samples were collected during vegetative exponential and vegetative saturated growth in rich media; yeast extract, peptone, dextrose (YEPD). Hexokinase (Hxk2) is shown as a loading control.

(B) Western blot for Yap5-FLAG from cultures grown in rich media (YEPD) for 4 h and then transferred to either YEPD (fermentable) or YEPG (non-fermentable) media. Hexokinase (Hxk2) is shown as a loading control.

(C) TL-seq (above) and TIS-profiling (below) data at the *PUS1* locus showing annotated and truncated transcript and protein isoforms. Cartoons at left of each data type indicate meiotic time points (rainbow). Bent arrows represent predicted transcription start sites. Note that very early TL-seq time points are not fully stage-matched to the first TIS-profiling time point, and late TIS-profiling time points do not have matched equivalents in the TL-seq time course, leading to patterns that appear to differ in timing when comparing patterns between the two sets of samples. Scale is linear and represents reads.

(D) Western blot for Pus1-3V5 from samples grown in rich media (YEPD) and treated with either DMSO or two different rapamycin concentrations for the times indicated. Hexokinase (Hxk2) is shown as a loading control.

(legend continued on next page)

been difficult to annotate sensitively and reliably. We chose parameters specifically to minimize false-positive detection, a major issue for this class of protein in our experience.[6] However, this also necessitated excluding some promising predicted truncations that did not pass some of our filters. As an example, TIS-profiling data indicated that the *PUS1* locus also encodes two isoforms, the annotated isoform and a truncated isoform (Pus1[truncation]), but the data were noisier than cases like Yap5[truncation], and this resulted in our algorithm only calling this truncation at one time point, leading to its exclusion from our final list of truncated isoforms (*YAP5*: Figure 3C, *PUS1*: Figure 4C). TL-seq data additionally showed evidence of an alternative transcript isoform beginning about 50–70 nt upstream of the Pus1[truncation] start codon, indicating that the truncated protein is likely to be translated from an alternative transcript (Figure 4C). Moreover, Pus1[truncation] was readily detected by western blotting of cells expressing a C-terminally 3V5-tagged Pus1 and is dynamically expressed across meiosis, with levels peaking early in meiosis and again later in spores (Figure 2D). After confirming expression of Pus1[truncation], we next sought to understand its function. The annotated Pus1 protein is a pseudouridine synthase that is known to play key roles in translation through modification of tRNAs and small nuclear RNAs (snRNAs) (reviewed in Rintala-Dempsey and Kothe[44]). More recently, Pus1 has also been found to modify a subset of mRNAs.[45–50] Pus1[truncation] contains regions involved in RNA binding, present at the C terminus of the full-length protein, but lacks the N-terminal catalytic domain responsible for pseudouridylation in annotated Pus1.[44,51]

Our western blot data indicate that Pus1[truncation] is most abundant in early meiosis and in spores, which are both conditions that display reduced levels of translation, as measured by polysome profiling.[4] To assess whether decreased translation is associated with Pus1[truncation] expression, we analyzed multiple additional conditions in which translation is reduced. First, we investigated the regulation of the *PUS1* locus in a standard ribosome-profiling dataset collected in meiotic cells lacking ribosomal protein Rpl40a, a condition that even more severely reduces translation early in meiosis.[52] Relative to wild type (WT), *rpl40aΔ* cells show reduced translation of the annotated Pus1 isoform and much more prominent translation of Pus1[truncation] (Figure S4A).

To monitor Pus1[truncation] levels in an orthogonal low-translation context, we treated mitotically growing cells with rapamycin, a general inhibitor of growth and ribosome biogenesis. Upon rapamycin treatment, we observed marked induction of Pus1[truncation], compared with undetectable levels in the untreated vehicle control (Figure 4D). Glucose starvation is another

condition that reduces translation. By western blotting, we compared Pus1[truncation] production in cells grown in normal rich media (2% glucose YEPD) to those grown in low-glucose media (0.2% glucose YEPD). After several hours in low-glucose media, we observed slight induction of the truncated isoform, compared with no induction in standard-rich media (Figure 4E). Together, these data are consistent with a role for Pus1[truncation] in conditions in which cellular translation is lowered, perhaps also explaining why this truncation has not been observed previously in the many studies focused on standard nutrient-rich growth conditions.

To further understand the role of Pus1[truncation], we strongly expressed the truncated isoform in vegetative exponential cells, a condition in which the truncation is not normally present, and performed mRNA-seq to assess transcript level changes. We used an anhydrotetracycline (aTC)-inducible allele to conditionally express Pus1[truncation] and collected uninduced and induced samples in both a WT and deletion (*pus1Δ*) background (Figure S4B). Expression of Pus1[truncation] alone had relatively little impact on overall gene expression, as measured by the low number of differentially expressed genes relative to WT (*n* = 61). Deletion of *PUS1* resulted in a slightly increased, but still modest, number of differentially expressed genes (*n* = 210; *p* < 2.2e−16, Fisher's exact test). The combination of *pus1Δ* with Pus1[truncation] expression, by contrast, yielded dramatically higher rates of differential gene expression, suggesting a synthetic phenotype between the two perturbations (*n* = 1,690; *p* < 2.2e−16, Fisher's exact test; Figures 4F and S4C; Table S3).

We performed Gene Ontology (GO) analysis of the differentially expressed genes in *pus1Δ* cells expressing Pus1[truncation] and found that the genes that were downregulated in the *pus1Δ* with Pus1[truncation] expression condition were strongly enriched for GO terms related to translation and RNA modification substrates, including ncRNA production and processing, rRNA processing, and ribosome biogenesis (Figure S4D; Table S4). To better understand the nature of the synthetic phenotype and visualize expression patterns across all our samples, we performed hierarchical clustering of all genes that were differentially expressed between WT and *pus1Δ* with Pus1[truncation] expression (Figure 4G). Cluster C contains genes that are upregulated in both WT and *pus1Δ* cells upon Pus1[truncation] induction. This gene set is generally upregulated upon aTC-induction of entirely unrelated genes and is therefore likely to be drug-dependent rather than Pus1[truncation]-specific (data not shown). Cluster E contains genes that are strongly downregulated in *pus1Δ* cells but are unaffected by Pus1[truncation] expression. This cluster is enriched for genes involved in translational regulation, RNA export, and rRNA

(E) Western blot for Pus1-3V5 from samples grown in either standard (2% YEPD) or reduced-nutrient (0.2% YEPD) media for the time points indicated. Hexokinase (Hxk2) is shown as a loading control.

(F) Bar graph showing the number of genes differentially expressed in the following conditions relative to WT: WT cells carrying a construct to allow anhydrotetracycline-inducible Pus1[trunc.] expression and treated with anhydrotetracycline (aTC; WT + Pus1[trunc.]); *pus1Δ* cells carrying a construct to allow aTC-inducible Pus1[trunc.] expression and treated with vehicle (*pus1Δ*) or aTC (*pus1Δ* + Pus1[trunc.]). In all cases, differential expression is relative to WT cells carrying a construct to allow aTC-inducible Pus1[trunc.] expression and treated with vehicle (WT). Differentially expressed genes were determined using DESeq2.

(G) Clustered heatmap of RNA-seq data for all genes that were differentially expressed between WT cells treated with vehicle (WT) and *pus1Δ* cells treated with aTC (*pus1Δ* + Pus1[trunc.]). Samples include WT cells carrying a construct to allow aTC-inducible Pus1[trunc.] expression and treated with vehicle (WT) or aTC (WT + Pus1[trunc.]), cells deleted for *PUS1* and carrying a construct to allow aTC-inducible Pus1[trunc] expression and treated with vehicle (*pus1Δ*) or aTC (*pus1Δ* + Pus1[trunc.]). Values are the log of the average of two replicates. Seven discrete clusters are indicated to the right with colored bars and letters. Gene Ontology terms enriched in each cluster are indicated to the right.
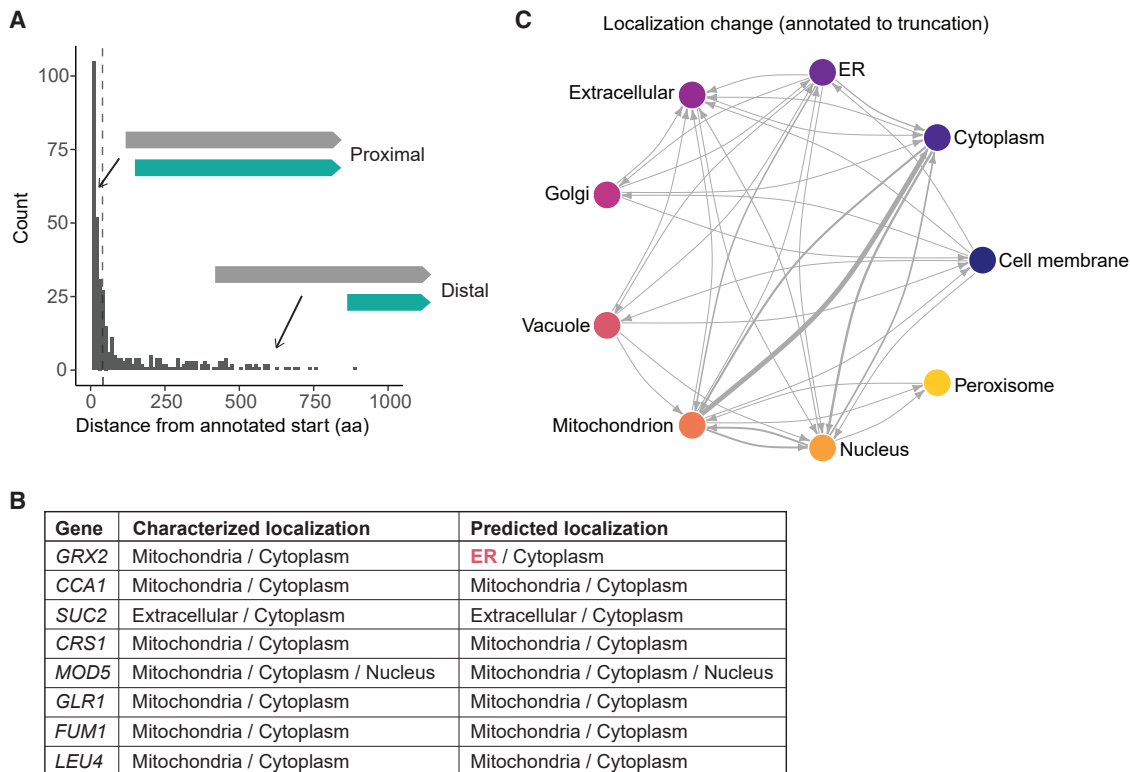
A

C

Localization change (annotated to truncation)

B

Gene	Characterized localization	Predicted localization
GRX2	Mitochondria / Cytoplasm	ER / Cytoplasm
CCA1	Mitochondria / Cytoplasm	Mitochondria / Cytoplasm
SUC2	Extracellular / Cytoplasm	Extracellular / Cytoplasm
CRS1	Mitochondria / Cytoplasm	Mitochondria / Cytoplasm
MOD5	Mitochondria / Cytoplasm / Nucleus	Mitochondria / Cytoplasm / Nucleus
GLR1	Mitochondria / Cytoplasm	Mitochondria / Cytoplasm
FUM1	Mitochondria / Cytoplasm	Mitochondria / Cytoplasm
LEU4	Mitochondria / Cytoplasm	Mitochondria / Cytoplasm

Figure 5. Computational prediction of differentially localized truncated isoforms

(A) Distribution of distances between annotated and truncated isoform translation-initiation sites among 388 truncations called by the algorithm, with cartoons above representing the proximal and distal categorization of truncations. The dotted line represents the 40 aa cutoff between the proximal and distal categories.

(B) Table of previously characterized truncations called by our algorithm, comparing their experimentally characterized localization with their predicted localization using DeepLoc1.0. Discrepancies between predicted and characterized localization are indicated in red. Note that for one gene *CCA1*, our algorithm only called the larger of two known truncated isoforms, so the smaller isoform was not included in the localization predictions.

(C) Schematic of the predicted subcellular localizations of annotated and truncated protein isoforms, with the blunt and pointed end of the arrows representing the annotated and truncated isoforms, respectively. Localization prediction was performed using DeepLoc1.0. Weight of line indicates the number of truncations in each category.

processing, consistent with known roles of Pus1. In light of the observed synthetic effects of expression of Pus1[truncation] in a *pus1Δ* background, two additional clusters are of particular note: (1) cluster B shows very little change upon Pus1[truncation] expression alone, modest downregulation occurring with the *pus1Δ* alone, and much more pronounced downregulation arising in the *pus1Δ* with Pus1[truncation] expression. This cluster is enriched for rRNA processing and ribosome assembly. (2) Cluster A, by contrast, contains genes that are largely unaffected with either of the individual perturbations but are strongly downregulated in the *pus1Δ* with Pus1[truncation] expression. This cluster is strongly enriched for genes involved in rDNA heterochromatin formation, rRNA processing, splicing, and small nucleolar RNA (snoRNA) metabolism. Together, these results indicate that although the loss of Pus1 alone has some impact on processes related to translation and ribosome biogenesis, the addition of Pus1[truncation] expression expands the number of affected genes and increases the severity of the changes to translation-related transcript expression. Thus, Pus1[truncation] appears to have subtle but global effects on translation that manifest primarily in the sensitized hypomodified context of *pus1Δ* cells, arguing that the role of the truncated isoform of

Pus1 is not dependent solely on the function of the annotated full-length Pus1 isoform but may be related to the functions of other pseudouridine synthases that affect translation.

Proximal truncations are a general mechanism for driving multiple localizations of protein products at a single locus

Although in-frame start codons are uniformly distributed throughout coding genes, their usage as translation initiation sites in our dataset is very non-uniform, with a strong bias for more N-terminal (proximal to annotated) start sites (Figure 5A).[53] In fact, nearly 60% of all truncations start within 40 aa of the corresponding annotated start codon. We hypothesized that—unlike in the case of distal truncations, which are likely to encode isolated domains—for proximal truncations, the core functional domains of the protein should typically remain intact in the truncated isoform, but the missing N-terminal sequences could encode localization signals, resulting in differential localization of the truncated isoform. This hypothesis is supported by the fact that among the small set of previously known truncated isoforms in *S. cerevisiae*, nearly all serve the function of differentially localizing a subset of the protein (Table 1; Figure 5B). It has
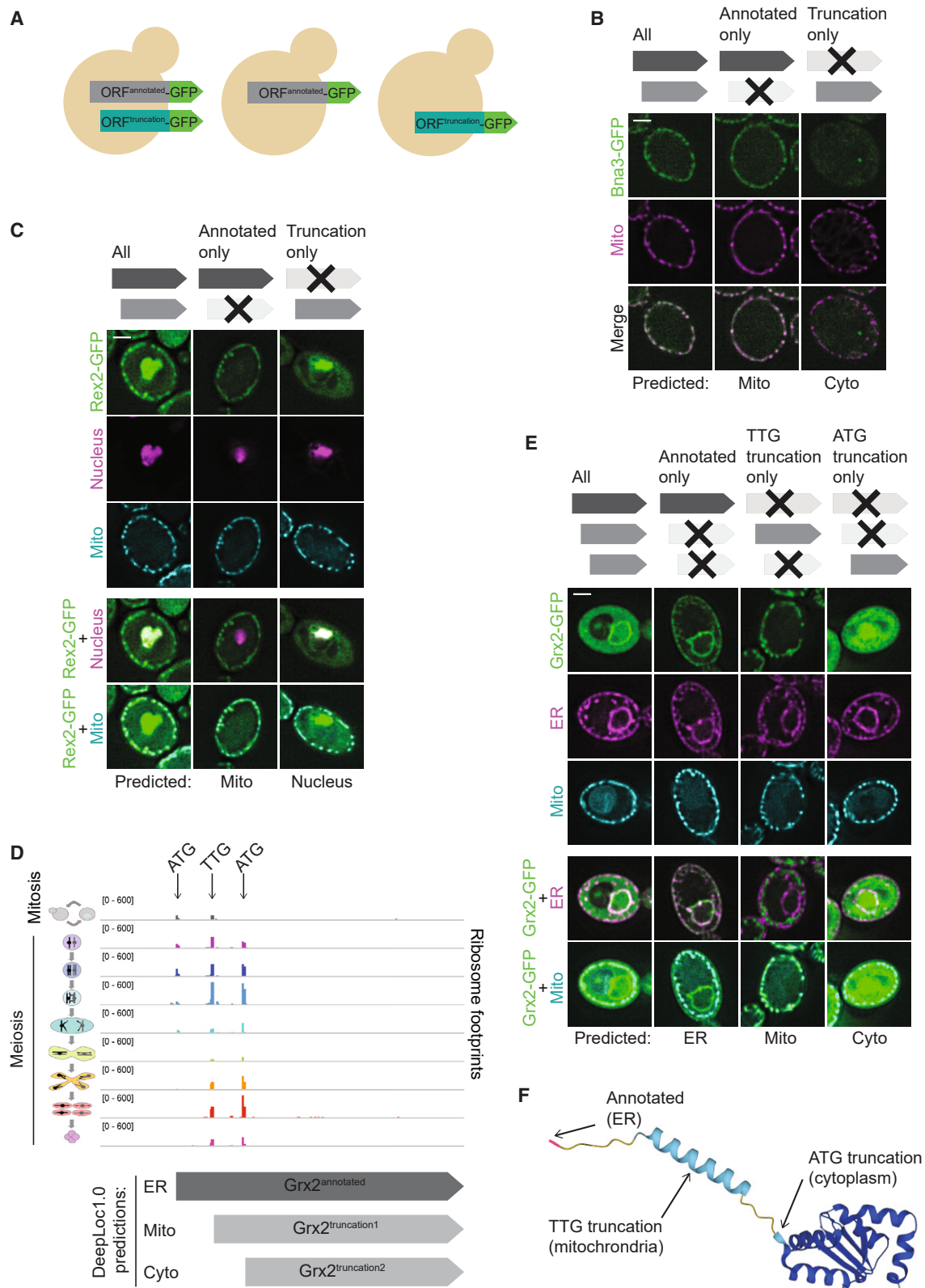
Figure 6. Experimental validation of differentially localized truncated isoforms

(A) Schematic of experimental approach for testing the localization of candidate truncated and annotated isoforms. ORFs of interest were fused to GFP to create a C-terminal tagged version of both annotated and N-terminally truncated isoforms simultaneously. Strains were constructed that contained (1) WT start codons for

similarly been shown that non-AUG extended protein isoforms can be differentially localized relative to their annotated counterparts.[54,55] We reasoned that our dataset would be an excellent opportunity to test whether this is indeed a broad phenomenon and to potentially identify additional differentially localized truncated isoforms.

We used a published algorithm, DeepLoc1.0, to perform localization prediction for our truncated isoforms and their annotated counterparts[56] (Table S5). DeepLoc1.0 is a deep learning algorithm trained on existing protein databases and uses sequence information alone to predict protein localization. It was important to use an algorithm that used sequence rather than homology, as the truncated and annotated isoforms would have very similar homology, given that they share significant amounts of sequence. Since localization signals are frequently encoded at the N terminus, we reasoned that "artificial" truncations could easily generate differentially localized proteins by removing an N-terminal sequence. To determine the expected background rate of differential localization, we generated sets of simulated truncations by randomly sampling in-frame start codons and performing localization prediction on those ORFs compared with the annotated protein. We found that our set of truncations differentially localized at a higher rate than expected by chance, with just over 30% of truncated isoforms localizing to a different compartment than their respective annotated isoforms (Figure S5A).

Among differentially localized truncations, it was most common for the truncations to lose the predicted mitochondrial or nuclear localization of their annotated counterpart, while cytoplasmic or extracellular localizations were the most likely localizations to be gained (Figures 5C and S5B). For initial validation of this prediction method, we compared the predicted and experimentally determined localizations for previously characterized truncations. For all but one of these genes, *GRX2*, the localization predictions match the characterized localization of the known isoforms, suggesting that the localization predictions were robust (Figure 5B).

We selected six predicted dual-localized truncated proteins and sought to experimentally validate their localization using fluorescence microscopy (Figure S6A). Candidate proteins were primarily selected based on the quality of TIS profiling data at the relevant locus and their ability to be tagged and imaged. Five candidates had no known alternative isoforms in existing literature. The sixth, Grx2, had a characterized truncated isoform for which the localization predictions did not match the characterized localizations (Figure 5B). For each locus of interest, we inserted a fluorescently tagged transgene under the control of the gene's endogenous promoter. To determine the localization of each isoform separately, we generated three strains for each gene: (1) all isoforms, with unmutated start codons; (2) annotated only, with the truncation start codon mutated to isoleucine; and (3) truncation only, with the annotated start codon mutated to isoleucine (Figures 6A and S6B–S6D). Two genes, *IGO2* and *APE4*, were not possible to validate—the truncated isoform of Igo2 was not detectable by western blotting or microscopy, and although both isoforms of Ape4 could be detected by western blotting, they were too low-abundance to detect by microscopy. For the remaining candidates, we performed fluorescence microscopy, collecting images at representative meiotic time points at which the respective truncations are expressed.

Bna3 is a kynurenine aminotransferase known to be dual localized to the mitochondria and cytoplasm, but with only one characterized isoform and no known mechanism for its dual localization.[57,58] Since the annotated isoform was predicted to be mitochondrial and the truncated isoform was predicted to be cytoplasmic, we hypothesized that the mechanism of the previously observed dual localization could be via production of two isoforms. This prediction was indeed validated by microscopy, as the "annotated only" strain showed only mitochondrial localization and the "truncation only" strain showed cytoplasmic localization (Figure 6B).

Rex2 is an RNA exonuclease involved in snRNA, rRNA, and ncRNA processing, with additional roles in mitochondrial DNA escape.[59,60] It is characterized as having mitochondrial localization, but its role in processing nuclear RNAs strongly suggested nuclear localization as well. This information aligned well with the DeepLoc1.0 prediction that the annotated isoform is mitochondrial, and the truncated isoform is nuclear. Upon imaging by microscopy, the strain containing only the annotated isoform showed mitochondrial localization, and the strain containing only the truncated isoform showed nuclear localization, in support of the predicted localizations for both isoforms (Figure 6C).

both annotated and truncated isoforms, (2) the truncation start codon mutated to encode isoleucine, or (3) the annotated start codon mutated to encode isoleucine.

(B) Fluorescence microscopy of C-terminally GFP-tagged Bna3, collected at 6 h in SPO, using the approach outlined in (A). Mitochondrial localization is indicated by Cit1-mCardinal. DeepLoc1.0-based predictions are shown below, and constructs analyzed in each column are illustrated above. Scale bar is 2 μm. Images are representative of >200 individual cells.

(C) Fluorescence microscopy of C-terminally GFP-tagged Rex2, collected at 1 h in SPO, using the approach outlined in (A). Mitochondrial localization is indicated by Su9MTS-BFP; nuclear localization is indicated by Htb1-mCherry. Separate merged images for nuclear and mitochondrial signal are shown at the bottom. DeepLoc1.0-based predictions are shown below, and constructs analyzed in each column are illustrated above. Scale bar is 2 μm. Images are representative of >200 individual cells.

(D) TIS-profiling data at the *GRX2* locus across all meiotic time points. Cartoons to the left indicate whether track represents mitotic cells (gray) or meiotic time points (rainbow). Arrows indicate start codons for the annotated isoform, and cartoons of the annotated and two predicted truncated isoforms are illustrated below, and predicted localizations for each isoform are to the left. Scale is linear and represents reads.

(E) Fluorescence microscopy of C-terminally GFP-tagged Grx2, collected at 0 h in SPO, using an expanded version of the approach outlined in (A). Strains were designed to express all isoforms, annotated isoform only, TTG-truncated isoform only, or ATG-truncated isoform only. Mitochondrial localization is indicated by Su9MTS-BFP; ER localization is indicated by mCherry-HDEL. Separate merged images for ER and mitochondrial signal are shown at the bottom. DeepLoc1.0-based predictions are shown below. Constructs analyzed in each column are illustrated above. Scale bar is 2 μm. Images representative of >200 individual cells.

(F) AlphaFold2 structural prediction for full-length Grx2, with arrows indicating the residues corresponding to the start codons of the annotated (ER), TTG truncation (mitochondrial), and ATG truncation (cytoplasmic) isoforms.

Ath1 is an acid trehalase involved in extracellular trehalose degradation. Its localization has been a topic of debate, with one study showing it to be vacuolar and another showing it to be periplasmic.[61,62] According to our predictions, the annotated isoform is Golgi-localized, and the truncated isoform is extracellular. We hypothesized that the extracellular localization of the truncation could correspond to the characterized periplasmic localization that had previously been attributed to the annotated isoform. This was supported by the fact that the periplasmic localization had been observed using an endogenous C-terminal tag, which would result in tagging of both isoforms.[61] The vacuolar localization, however, was observed with an N-terminal tag under a strong promoter, which would only tag the annotated isoform and could result in vacuolar signal simply due to protein degradation.[62] Microscopy revealed that indeed, the truncated isoform in isolation localizes to the periplasmic space (Figure S6E). Its levels also increase in spores, which aligns with its known function in breaking down trehalose, a storage carbohydrate that needs to be mobilized in spores. The annotated isoform, however, was not possible to visualize. This suggests that the truncated isoform is likely the main functional isoform, at least for the characterized function, and that the annotated isoform is either not produced or is at very low abundance under the conditions we assayed. This is consistent with published work showing that artificially periplasm-targeted Ath1 is sufficient for growth on trehalose, while vacuole-targeted Ath1 is not.[61]

Finally, we investigated the localization of Grx2, a glutaredoxin with two known in-frame start codons, one producing the annotated mitochondrial isoform and the second producing a truncated cytoplasmic isoform.[25] It was the only characterized gene with a truncated isoform for which our localization predictions did not match the characterized localizations, with the annotated isoform predicted to be endoplasmic reticulum (ER)-localized rather than the characterized mitochondrial localization (Figure 5B). To our surprise, when we imaged the annotated isoform alone, we saw clear ER localization, in line with our prediction but contradicting the established localization (Figure 5B). To reconcile this with the published data, we returned to the TIS-profiling data and noticed an additional in-frame start codon with clear signal, this time at a TTG codon, one of the most efficient of nine known near-cognate start codons (differing from ATG by one nucleotide; Figure 6D).[63,64] We hypothesized that this could be the source of the missing mitochondrial localization, and indeed, the DeepLoc1.0 algorithm predicted mitochondrial localization for this isoform. To validate this prediction by microscopy, we generated strains that expressed each isoform alone by mutating the other two start codons, either to isoleucine (ATT) for the ATG starts or to a non-near-cognate leucine (TTA) for TTG. The three isoforms localized as predicted, with the annotated isoform in the ER, the TTG truncation in the mitochondria, and the ATG truncation in the cytoplasm (Figure 6E). Therefore, our truncation-calling algorithm, in combination with DeepLoc1.0, correctly predicted the existence and localization of several previously characterized truncations and provided insight into the regulation of Grx2, a glutaredoxin that is localized to three different subcellular compartments by expression of three protein isoforms with slightly different N termini. These localization analyses revealed a previously undetected set of dual-localized isoforms and suggest that production of proximal truncations may be a widespread strategy for targeting one protein to two or more subcellular localizations.

DISCUSSION

To date, only a handful of truncated isoforms have been characterized, and most were identified in single-gene studies of very well-studied biological pathways (Table 1). Global approaches to systematically identify these isoforms by ribosome profiling and mass spectrometry have been hampered by limitations imposed by their overlap with known proteins. Here we develop an algorithm for identifying truncated protein isoforms and report the translation of 388 truncated proteins in budding yeast, an organism generally considered to have little protein isoform diversity. Our analysis of this dataset represents a dramatic increase in the number of known truncated proteins in yeast and provides an unbiased picture of the prevalence and contexts in which these isoforms are present. Although functions for most remain to be investigated, we show evidence of functional activity for several truncations, with examples of proximal truncations serving to differentially localize proteins and distal truncations acting in condition-specific roles that may differ from the function of their annotated isoform. In addition to the rich global information it contains, we believe that this dataset will be a valuable resource for single-gene studies, as it will provide more complete information about the gene products present at a given locus and may be helpful for generating functional hypotheses.

We orthogonally validated the production of numerous previously undetected truncated isoforms by performing western blotting for C-terminally epitope-tagged proteins (Figures 2A–2G and S2A–S2F). Of the candidates we chose, 10 were detectable by western blotting under WT conditions (with active proteasomes) and generally at time points consistent with the TIS-profiling data. The remaining five truncations that we tested, however, were not readily detectable by western blotting under these conditions, despite the strong and specific signal in the TIS-profiling data. We showed that several truncated proteins are subject to proteasome-mediated degradation, which in some cases made their detection in untreated cells more challenging (Figures 2K, 2L, and S2G–S2I). The observed degradation of this set of truncated proteins could suggest that they are turned over because they are non-functional or deleterious. It is also possible, however, that instability is an important functional feature of the protein in some cases. For example, one previously characterized truncation, produced from the *KAR4* locus, is much less stable than the annotated isoform, which is constitutively expressed. Truncated Kar2 is expressed in response to mating pheromone and serves to provide a large boost in protein expression during mating, after which it is rapidly degraded to minimize toxicity of Kar4 to the cell.[27] Such regulation could exist at other loci with truncated isoforms, providing stable and unstable pools of protein that are either functionally identical (in the case of proximal truncations) or functionally distinct (in the case of distal truncations). In general, we observed that truncated proteins beginning closer to the annotated start codon are more robustly and stably expressed than more distal truncations, which tended to be much more difficult to detect despite

often having strong signal in the TIS-profiling data. We suspect that more proximal truncations are more likely to be stable because they are quite similar in size and composition to the annotated protein. Given the outsized importance of the N terminus of proteins for protein stability, however, even a slight difference between similarly sized isoforms could confer differences in stability.

We observed dynamic regulation for most truncated isoforms in the TIS-profiling data (Figure 1F), providing support for specific cellular functions for these newly identified proteins. Although TIS profiling can give an approximate picture of regulatory patterns, it is not a robustly quantitative measure, and thus specific peak heights should not be overinterpreted.[6] We did, however, observe consistent regulatory patterns by western blotting for several examples, indicating that the TIS profiling can be informative for regulatory trends. In many cases, this dynamic regulation is likely facilitated by the presence of a truncated transcript rather than through translational regulation, although it is important to note that the presence of a truncated transcript does not preclude the possibility that some truncated protein is produced from the longer, annotated transcript. We hypothesize that truncated proteins without a detectable truncated transcript are typically produced via start codon readthrough of the annotated start codon, although this remains untested in the cases presented here (Figure 3E). It remains possible, of course, that a subset of the truncated proteins apparently lacking a truncated transcript in fact arise from false positives in the TIS profiling or false negatives in the TL-seq data; however, we expect that this would be a minor contribution.

A large subset of the proximal truncations that we identified seem to lead to otherwise functionally identical proteins being targeted to different subcellular locations, which provides a mechanism for how in so many cases, similar cellular functions are performed in multiple cellular compartments. For example, DNA replication, transcription, and translation all occur in both the mitochondria and either the nucleus or cytoplasm. These types of related but spatially separated functions can be encoded either by two separate genes (through gene duplication or functional convergence) or by a single gene.[65] Previous single-gene studies made it clear that a single locus can encode multiple differentially localized protein isoforms, but the full extent of the phenomenon was unknown, and the known examples were biased toward very well-studied biological pathways (Table 1). For example, several previously characterized truncated isoforms were aa tRNA synthetases, known to act on both cytoplasmic and mitochondrial tRNAs, and were therefore clear candidates for this type of regulation. Here, genome-wide data allowed us to see the extent of this type of regulation in yeast in a less biased way. Candidates that we validated for their predicted localization differences spanned a diverse range of functions and revealed a variety of ways in which knowledge of this cellular strategy can enhance our understanding of gene function and regulation.

In the case of Bna3, dual protein localization was already established, but its basis was unknown. The two isoforms at this locus follow a common trend seen among previously characterized truncations, in which the longer form has a mitochondrial signal sequence that is lost in the truncated isoform, causing it to default to cytoplasmic localization (Figure 6B). *REX2* presented a slightly different scenario, in which dual localization had not been established but was very likely given known functional information. Rex2 had been characterized as a mitochondrial protein by mitochondrial fractionation of overexpression strains.[59] Despite not being explicitly characterized as nuclear by microscopy, it was also shown to be involved in rRNA and snRNA processing, strongly suggesting a nuclear localization.[60] Our predictions and validation support both localizations and suggest that the mechanism of dual localization is through production of two differentially targeted protein isoforms (Figure 6C). It is also an interesting example of one signal sequence being removed (mitochondrial) and another being unmasked (nuclear) in the truncated isoform. Interestingly, we did not find dual localization for *ATH1* and instead showed that the characterized function is likely carried out by the truncated isoform and that the annotated isoform is likely not expressed at appreciable levels (Figure S6E). Identification of the truncated isoform and subsequent localization predictions were valuable for explaining inconsistencies in the existing literature.

The *GRX2* locus provides an additional demonstration of how TIS profiling can help reconcile confusing information about a gene's regulation and function. Based on previous characterization, *GRX2* was thought to produce two protein isoforms, one mitochondrial and one cytoplasmic.[26] In that study, however, three bands were observed by western blotting, which likely correspond to the three isoforms that we identified. The long isoform, however, was hypothesized to be mitochondrial, and the intermediate isoform was attributed to processing of the mitochondrial targeting sequence from the longer isoform. A small amount of protein was also detected in the ER, but this was attributed to slow import kinetics into the mitochondria. With the additional insight provided by the TIS profiling paired with localization prediction, we were able to identify a third isoform of the redox-regulator Grx2. Visualization of the protein structure using AlphaFold2 shows that all three isoforms, which localize to three different cellular compartments, still retain the structured, functional core of the protein (Figures 6D–6F).[66]

The identification of a TTG-initiated truncation at the *GRX2* locus raises the question of whether near-cognate start codons should have been included in our truncation-calling algorithm. We chose to exclude them because visual analysis indicated that most near-cognate-driven truncations were false positives caused by background noise within genes and would likely require different treatment and calling thresholds than AUG initiation sites. Although the *GRX2* locus shows that near-cognate-initiated truncations can be made, we still believe that they are very rare. Translation at near-cognate start codons alone is often not sufficient to stabilize a transcript, as shown in past work in which mutation of the annotated start codon of transcripts encoding a near-cognate-initiated extended protein isoform led to nonsense-mediated decay of the transcript; this was caused by efficient translation initiation at an out-of-frame AUG codon downstream following inefficient initiation at the in-frame near-cognate codon.[6] Therefore, we suspect that most cases of near-cognate truncations would need to arise in a context similar to *GRX2*, in which a truncated transcript bears an ATG truncation, whose translation ensures the stability of the transcript, paired with an upstream near-cognate-initiated isoform. In these

cases, the near-cognate isoform is essentially behaving like an N-terminal extension within the context of the truncated transcript.

We observed dynamic and condition-specific regulation for two distal truncations that we investigated experimentally, Yap5^{truncation} and Pus1^{truncation}, which suggests functional relevance. Yap5^{truncation} contains the Fe-S cluster binding domain located in the C-terminal half of the annotated protein and is markedly similar in size to an artificial truncation of Yap5 that was shown to effectively bind Fe-S clusters.[43] We show that Yap5^{truncation} is specifically induced under multiple respiratory conditions: meiosis, saturated growth, and growth in non-fermentable media (Figures 2A, 4A, and 4B). This suggests that it may be involved in responding to elevated respiratory activity, a role that could be related to its ability to bind Fe-S clusters, important cofactors in the electron transport chain. Further work will be necessary to elucidate its specific functional role.

Pus1^{truncation} is produced throughout meiosis and contains the positively charged residues involved in RNA binding of the full-length protein.[51] We show that this truncated isoform is likely nutrient-regulated since it is expressed in the low nutrient media that induces meiosis, as well as upon glucose starvation and rapamycin treatment (Figures 2D, 4D, and 4E). This nutrient regulation is intriguing given that a number of Pus1-dependent modifications in mRNA are dynamically regulated during nutrient deprivation.[46,50] The induction of Pus1^{trunction} upon rapamycin treatment is also interesting, and the formation of some pseudouridines by other Pus proteins is known to be dynamically regulated by the TOR pathway, and a pseudouridine in the U6 snRNA is introduced by Pus1 during filamentous growth, also regulated by the TOR pathway.[45,67,68]

mRNA-seq of cells expressing Pus1^{truncation} in WT and pus1Δ backgrounds under rich growth conditions revealed mild effects of either Pus1^{truncation} expression or PUS1 deletion alone and a much more dramatic effect when deletion of PUS1 is combined with Pus1^{truncation} expression (Figures 4F and 4G). This more severe synthetic gene expression phenotype suggests that Pus1^{truncation} has effects that are not directly related to full-length Pus1 function, potentially also affecting targets of other pseudouridine synthases as well. This is perhaps unsurprising given that the specificity of the enzyme is primarily conferred by the catalytic domain, and there is little reason to think the RNA binding domain alone would be specific to Pus1 targets.

The precise reason for the strong downregulation of genes involved in ribosome biogenesis, rRNA processing, and ncRNA processing in cells expressing Pus1^{truncation} and deleted for PUS1 is not immediately obvious (Figure S4D). Pseudouridine synthases as a group perform extensive pseudouridylation of rRNA, tRNA, snRNA, snoRNA, and mRNA targets, any of which could have important impacts on translation and RNA processing (reviewed in Rintala-Dempsey and Kothe[44]). Pus1 itself has been shown to modify ribosomal mRNAs, including 5 subunits of the ribosomal large subunit, as well as RNase MRP, which is involved in the maturation of rRNA.[46,50] Given that Pus1^{truncation} contains regions involved in RNA-binding, we hypothesize that it could occlude target binding by Pus1 and potentially other pseudouridine synthases as well.[51] The presence of a synthetic

effect with deletion of PUS1 is suggestive of a role in modulating pseudouridylation. While deletion of PUS1 alone results in viable cells and only mild phenotypic effects, much more dramatic synthetic phenotypes have been observed when PUS1 deletion is combined with deletion of other pseudouridine synthases or with mutations that compromise tRNA stability.[69–71] If Pus1^{truncation} has a role related to pseudouridylation, a synthetic phenotype with deletion of PUS1 would be expected (and this is what we observe). Further study will be necessary to understand the specific mechanistic role of Pus1^{truncation}.

Our hypotheses for Yap5^{truncation} and Pus1^{truncation} function were notably tied to the known functional characteristics of their annotated isoforms. Whether this is a valid approach is unclear, as many truncations—particularly distal truncations—lack key functional sequences of the annotated protein. The degree to which the annotated and truncated isoform differs in function likely varies depending on the extent of the truncation, with the proximal truncations being much more likely to share functional characteristics with the annotated isoform and only varying in typical N-terminally encoded characteristics such as stability and localization. Distal truncations, on the other hand, are more likely to be missing key functional domains and may not even contain any intact domains, making it much more difficult to generate a rational prediction for their functions.

Beyond the additional regulation we uncovered for Pus1 as a result of identifying its truncated isoform, this case highlights a key point: our truncation-calling algorithm is stringent, likely excluding a number of real truncated isoforms in order to minimize false-positive calls, which previously plagued identification of N-terminal truncations. Systematic identification of this type of non-canonical protein is fundamentally distinct from other classes, including ORFs in upstream regions (uORFs), those that are short and intergenic (sORFs), those downstream of annotated ORFs (dORFs), and N-terminally extended ORFs. In all other cases, some or all of the newly identified ORF is non-overlapping with an annotated ORF. Thus, approaches based on standard ribosome-profiling data that leverage initiation codon peaks resulting from CHX pre-treatment, periodicity resulting from elongation, or simply ribosome density are not effective for stringently calling N-terminal truncations. Even approaches independent of standard ribosome profiling, like those that use evolutionary conservation to identify coding regions, are problematic for this class in particular; and those that rely on machine learning-based analysis of TIS mapping in conjunction with standard ribosome-profiling analysis generate an exceedingly high level of false positives and negatives, likely due to multiple of the factors noted above.[6] Our systematic identification of truncated proteins, by contrast, relied entirely on TIS-profiling data. Relative to standard ribosome profiling, the TIS-profiling data are much simpler to interpret since the reads are highly enriched at sites of translation initiation, and signal is not obscured by elongating ribosomes from the overlapping annotated ORF, allowing us to very robustly identify truncated protein initiation sites (Figures 1A and 1B). Analysis of parallel standard ribosome-profiling data provided clear validation of our calls, suggesting that future studies can rely on TIS profiling for protein isoform identification (Figure 1E). Furthermore, multiple lines of experimental testing revealed that these isoforms can have

localization and function distinct from their corresponding annotated ORF.

Non-canonical translation has previously been shown to be higher during meiosis and other stress conditions, and from our observations in this dataset, truncated isoforms are no exception.[4,6,72] This could be evolutionarily beneficial, allowing cells to sample a greater proteomic diversity to adapt to new environments. Some truncated isoforms may not currently be "useful" to cells but may eventually over time become functional. Although the use of different isoforms bears some similarity to gene duplication, it is markedly different in that the shared sequences between the two isoforms are unable to evolve separately. Only the region missing in the truncation can change independently between the two isoforms. N-terminal sequences, however, are often particularly important for gene function, as we have shown for localization and stability in this study. Therefore, having a mechanism to test out different N-terminal sequences while still retaining protein production from the annotated start site could be beneficial. Future work on the prevalence and conservation of truncated isoforms across different stress conditions and other organisms will further elucidate both the functional relevance and evolutionary processes giving rise to truncated proteins.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Yeast strain construction
  - Yeast growth and sporulation
- METHOD DETAILS
  - Protein extraction and western blotting
  - Proteasome inhibition
  - Growth in non-fermentable media
  - Rapamycin treatment
  - Low glucose growth
  - Pus1^{truncation} overexpression
  - RNA extraction
  - Poly-A selection and RNA-seq
  - Live imaging
  - Ribosome profiling metagene analysis
  - TL-seq metagene analysis
  - Staging comparison between TIS-profiling and TL-seq time courses
  - Localization prediction
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sequence alignment, quantification, and differential expression analysis
  - Gene ontology enrichment analysis
  - Truncation calling algorithm
  - TL-seq peak calling

AUTHOR CONTRIBUTIONS

A.L.H. and G.A.B. conceived most aspects of this study. Western blot validation of truncations and proteasome inhibition experiments were performed by A.L.H. and N.H.W. All other experiments were performed by A.L.H. Algorithm development and data analysis were performed by A.L.H. The manuscript was written and edited by A.L.H. and G.A.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. PLoS Comput. Biol. 4, e1000176. https://doi.org/10.1371/journal.pcbi.1000176.

2. Ares, M., Grate, L., and Pauling, M.H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA 5, 1138–1139. https://doi.org/10.1017/S1355838299991379.

3. Douglass, S.M., Leung, C.S., and Johnson, T.L. (2019). Extensive splicing across the Saccharomyces cerevisiae genome. Preprint at bioRxiv. https://doi.org/10.1101/515163.

4. Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. Science 335, 552–557. https://doi.org/10.1126/science.1215110.

5. Chia, M., Li, C., Marques, S., Pelechano, V., Luscombe, N.M., and van Werven, F.J. (2021). High-resolution analysis of cell-state transitions in yeast suggests widespread transcriptional tuning by alternative starts. Genome Biol. 22, 34. https://doi.org/10.1186/s13059-020-02245-3.

6. Eisenberg, A.R., Higdon, A.L., Hollerer, I., Fields, A.P., Jungreis, I., Diamond, P.D., Kellis, M., Jovanovic, M., and Brar, G.A. (2020). Translation Initiation Site Profiling Reveals Widespread Synthesis of Non-AUG-Initiated Protein Isoforms in Yeast. Cell Syst. 11, 145–160.e5. https://doi.org/10.1016/j.cels.2020.06.011.

7. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802. https://doi.org/10.1016/j.cell.2011.10.002.

8. Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature 497, 127–131. https://doi.org/10.1038/nature12121.

9. Marston, A.L., and Amon, A. (2004). Meiosis: cell-cycle controls shuffle and deal. Nat. Rev. Mol. Cell Biol. 5, 983–997. https://doi.org/10.1038/nrm1526.

10. van Werven, F.J., and Amon, A. (2011). Regulation of entry into gameto-genesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 3521–3531. https://doi.org/10.1098/rstb.2011.0081.

11. Sing, T.L., Conlon, K., Lu, S.H., Madrazo, N., Morse, K., Barker, J.C., Hollerer, I., Brar, G.A., Sudmant, P.H., and Ünal, E. (2022). Meiotic cDNA libraries reveal gene truncations and mitochondrial proteins important for competitive fitness in Saccharomyces cerevisiae. Genetics *221*, iyac066. https://doi.org/10.1093/genetics/iyac066.

12. Bigler, J., Hokanson, W., and Eisenman, R.N. (1992). Thyroid Hormone Receptor Transcriptional Activity Is Potentially Autoregulated by Truncated Forms of the Receptor. Mol. Cell. Biol. *12*, 2406–2417. https://doi.org/10.1128/mcb.12.5.2406-2417.1992.

13. Brubaker, S.W., Gauthier, A.E., Mills, E.W., Ingolia, N.T., and Kagan, J.C. (2014). A Bicistronic MAVS Transcript Highlights a Class of Truncated Variants in Antiviral Immunity. Cell *156*, 800–811. https://doi.org/10.1016/j.cell.2014.01.021.

14. Delmas, V., Laoide, B.M., Masquilier, D., De Groot, R.P., Foulkes, N.S., and Sassone-Corsi, P. (1992). Alternative usage of initiation codons in mRNA encoding the cAMP-responsive-element modulator generates regulators with opposite functions. Proc. Natl. Acad. Sci. USA *89*, 4226–4230.

15. Falvey, E., Fleury-Olela, F., and Schibler, U. (1995). The rat hepatic leukemia factor (HLF) gene encodes two transcriptional activators with distinct circadian rhythms, tissue distributions and target preferences. EMBO J. *14*, 4307–4317. https://doi.org/10.1002/j.1460-2075.1995.tb00105.x.

16. Fields, A.P., Rodriguez, E.H., Jovanovic, M., Stern-Ginossar, N., Haas, B.J., Mertins, P., Raychowdhury, R., Hacohen, N., Carr, S.A., Ingolia, N.T., et al. (2015). A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. Mol. Cell *60*, 816–827. https://doi.org/10.1016/j.molcel.2015.11.013.

17. Iliopoulos, O., Ohh, M., and Kaelin, W.G. (1998). pVHL19 is a biologically active product of the von Hippel–Lindau gene arising from internal translation initiation. Proc. Natl. Acad. Sci. USA *95*, 11661–11666. https://doi.org/10.1073/pnas.95.20.11661.

18. Sasaki, A., Inagaki-Ohara, K., Yoshida, T., Yamanaka, A., Sasaki, M., Yasukawa, H., Koromilas, A.E., and Yoshimura, A. (2003). The N-terminal Truncated Isoform of SOCS3 Translated from an Alternative Initiation AUG Codon under Stress Conditions Is Stable Due to the Lack of a Major Ubiquitination Site, Lys-6. J. Biol. Chem. *278*, 2432–2436. https://doi.org/10.1074/jbc.C200608200.

19. Smyth, J.W., and Shaw, R.M. (2013). Autoregulation of Connexin43 Gap Junction Formation by Internally Translated Isoforms. Cell Rep. *5*, 611–618. https://doi.org/10.1016/j.celrep.2013.10.009.

20. Wolfe, C.L., Lou, Y.C., Hopper, A.K., and Martin, N.C. (1994). Interplay of heterogeneous transcriptional start sites and translational selection of AUGs dictate the production of mitochondrial and cytosolic/nuclear tRNA nucleotidyltransferase from the same gene in yeast. J. Biol. Chem. *269*, 13361–13366. https://doi.org/10.1016/S0021-9258(17)36841-2.

21. Nishimura, A., Nasuno, R., Yoshikawa, Y., Jung, M., Ida, T., Matsunaga, T., Morita, M., Takagi, H., Motohashi, H., and Akaike, T. (2019). Mitochondrial cysteinyl-tRNA synthetase is expressed via alternative transcriptional initiation regulated by energy metabolism in yeast cells. J. Biol. Chem. *294*, 13781–13788. https://doi.org/10.1074/jbc.RA119.009203.

22. Wu, M., and Tzagoloff, A. (1987). Mitochondrial and cytoplasmic fumarases in Saccharomyces cerevisiae are encoded by a single nuclear gene FUM1. J. Biol. Chem. *262*, 12275–12282. https://doi.org/10.1016/S0021-9258(18)45347-1.

23. Rai, R., Tate, J.J., Georis, I., Dubois, E., and Cooper, T.G. (2014). Constitutive and Nitrogen Catabolite Repression-sensitive Production of Gat1 Isoforms. J. Biol. Chem. *289*, 2918–2933. https://doi.org/10.1074/jbc.M113.516740.

24. Outten, C.E., and Culotta, V.C. (2004). Alternative start sites in the Saccharomyces cerevisiae GLR1 gene are responsible for mitochondrial and cytosolic isoforms of glutathione reductase. J. Biol. Chem. *279*, 7785–7791. https://doi.org/10.1074/jbc.M312421200.

25. Pedrajas, J.R., Porras, P., Martínez-Galisteo, E., Padilla, C.A., Miranda-Vizuete, A., and Bárcena, J.A. (2002). Two isoforms of Saccharomyces cerevisiae glutaredoxin 2 are expressed in vivo and localize to different subcellular compartments. Biochem. J. *364*, 617–623. https://doi.org/10.1042/BJ20020570.

26. Porras, P., Padilla, C.A., Krayl, M., Voos, W., and Bárcena, J.A. (2006). One single in-frame AUG codon is responsible for a diversity of subcellular localizations of glutaredoxin 2 in Saccharomyces cerevisiae. J. Biol. Chem. *281*, 16551–16562. https://doi.org/10.1074/jbc.M600790200.

27. Gammie, A.E., Stewart, B.G., Scott, C.F., and Rose, M.D. (1999). The Two Forms of Karyogamy Transcription Factor Kar4p Are Regulated by Differential Initiation of Transcription, Translation, and Protein Turnover. Mol. Cell. Biol. *19*, 817–825.

28. Beltzer, J.P., Chang, L.F., Hinkkanen, A.E., and Kohlhaw, G.B. (1986). Structure of yeast LEU4. The 5' flanking region contains features that predict two modes of control and two productive translation starts. J. Biol. Chem. *261*, 5160–5167. https://doi.org/10.1016/S0021-9258(19)89228-1.

29. Beltzer, J.P., Morris, S.R., and Kohlhaw, G.B. (1988). Yeast LEU4 encodes mitochondrial and nonmitochondrial forms of alpha-isopropylmalate synthase. J. Biol. Chem. *263*, 368–374. https://doi.org/10.1016/S0021-9258(19)57402-6.

30. Boguta, M., Hunter, L.A., Shen, W.C., Gillman, E.C., Martin, N.C., and Hopper, A.K. (1994). Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol. Mol. Cell. Biol. *14*, 2298–2306. https://doi.org/10.1128/mcb.14.4.2298-2306.1994.

31. Carlson, M., and Botstein, D. (1982). Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. Cell *28*, 145–154. https://doi.org/10.1016/0092-8674(82)90384-1.

32. Taussig, R., and Carlson, M. (1983). Nucleotide sequence of the yeast SUC2 gene for invertase. Nucleic Acids Res *11*, 1943–1954. https://doi.org/10.1093/nar/11.6.1943.

33. Chatton, B., Walter, P., Ebel, J.P., Lacroute, F., and Fasiolo, F. (1988). The yeast VAS1 gene encodes both mitochondrial and cytoplasmic valyl-tRNA synthetases. J. Biol. Chem. *263*, 52–57. https://doi.org/10.1016/S0021-9258(19)57354-9.

34. Amaral, C., Vicente, C.T., Caetano, S.M., Gaspar-Cordeiro, A., Yang, Y., Cloetens, P., Romão, C.V., Rodrigues-Pousada, C., and Pimentel, C. (2021). An Internal Promoter Drives the Expression of a Truncated Form of CCC1 Capable of Protecting Yeast from Iron Toxicity. Microorganisms *9*, 1337. https://doi.org/10.3390/microorganisms9061337.

35. Natsoulis, G., Hilger, F., and Fink, G.R. (1986). The HTS1 gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of S. cerevisiae. Cell *46*, 235–243. https://doi.org/10.1016/0092-8674(86)90740-3.

36. Zhou, S., Sternglanz, R., and Neiman, A.M. (2017). Developmentally regulated internal transcription initiation during meiosis in budding yeast. PLOS ONE *12*, e0188001. https://doi.org/10.1371/journal.pone.0188001.

37. Rose, A.M., Joyce, P.B., Hopper, A.K., and Martin, N.C. (1992). Separate information required for nuclear and subnuclear localization: additional complexity in localizing an enzyme shared by mitochondria and nuclei. Mol. Cell. Biol. *12*, 5652–5658. https://doi.org/10.1128/mcb.12.12.5652-5658.1992.

38. 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.

39. Spealman, P., Naik, A., and McManus, J. (2021). uORF-seqr: A Machine Learning-Based Approach to the Identification of Upstream Open Reading Frames in Yeast. In Ribosome Profiling Methods in Molecular Biology, V.M. Labunskyy, ed. (Springer), pp. 313–329. https://doi.org/10.1007/978-1-0716-1150-0_15.

40. Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. Science *320*, 1344–1349. https://doi.org/10.1126/science.1158441.

41. Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13–37. https://doi.org/10.1016/j.gene.2005.06.037.

42. Li, L., Bagley, D., Ward, D.M., and Kaplan, J. (2008). Yap5 Is an Iron-Responsive Transcriptional Activator That Regulates Vacuolar Iron Storage in Yeast. Mol. Cell. Biol. *28*, 1326–1337. https://doi.org/10.1128/MCB.01219-07.

43. Rietzschel, N., Pierik, A.J., Bill, E., Lill, R., and Mühlenhoff, U. (2015). The Basic Leucine Zipper Stress Response Regulator Yap5 Senses High-Iron Conditions by Coordination of [2Fe-2S] Clusters. Mol. Cell. Biol. *35*, 370–378. https://doi.org/10.1128/MCB.01033-14.

44. Rintala-Dempsey, A.C., and Kothe, U. (2017). Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? RNA Biol. *14*, 1185–1196. https://doi.org/10.1080/15476286.2016.1276150.

45. Basak, A., and Query, C.C. (2014). A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. Cell Rep. *8*, 966–973. https://doi.org/10.1016/j.celrep.2014.07.004.

46. Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W.V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143–146. https://doi.org/10.1038/nature13802.

47. Lovejoy, A.F., Riordan, D.P., and Brown, P.O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One *9*, e110799. https://doi.org/10.1371/journal.pone.0110799.

48. Massenet, S., Motorin, Y., Lafontaine, D.L.J., Hurt, E.C., Grosjean, H., and Branlant, C. (1999). Pseudouridine Mapping in the Saccharomyces cerevisiae Spliceosomal U Small Nuclear RNAs (snRNAs) Reveals that Pseudouridine Synthase Pus1p Exhibits a Dual Substrate Specificity for U2 snRNA and tRNA. Mol. Cell. Biol. *19*, 2142–2154. https://doi.org/10.1128/MCB.19.3.2142.

49. Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E., and Grosjean, H. (1998). The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. RNA *4*, 856–869. https://doi.org/10.1017/s1355838298980396.

50. Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., León-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., et al. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148–162. https://doi.org/10.1016/j.cell.2014.08.028.

51. Czudnochowski, N., Wang, A.L., Finer-Moore, J., and Stroud, R.M. (2013). In human pseudouridine Synthase 1 (hPus1), a C-Terminal Helical Insert Blocks tRNA from Binding in the Same Orientation as in the Pus1 Bacterial Homologue TruA, Consistent with Their Different Target Selectivities. J. Mol. Biol. *425*, 3875–3887. https://doi.org/10.1016/j.jmb.2013.05.014.

52. Cheng, Z., Mugler, C.F., Keskin, A., Hodapp, S., Chan, L.Y.-L., Weis, K., Mertins, P., Regev, A., Jovanovic, M., and Brar, G.A. (2019). Small and Large Ribosomal Subunit Deficiencies Lead to Distinct Gene Expression Signatures that Reflect Cellular Growth Rate. Mol. Cell *73*, 36–47.e10. https://doi.org/10.1016/j.molcel.2018.10.032.

53. Bazykin, G.A., and Kochetov, A.V. (2011). Alternative translation start sites are conserved in eukaryotic genomes. Nucleic Acids Res. *39*, 567–577. https://doi.org/10.1093/nar/gkq806.

54. Monteuuis, G., Miścicka, A., Świrski, M., Zenad, L., Niemitalo, O., Wrobel, L., Alam, J., Chacinska, A., Kastaniotis, A.J., and Kufel, J. (2019). Non-canonical translation initiation in yeast generates a cryptic pool of mitochondrial proteins. Nucleic Acids Res. *47*, 5777–5791. https://doi.org/10.1093/nar/gkz301.

55. Tang, H.L., Yeh, L.S., Chen, N.K., Ripmaster, T., Schimmel, P., and Wang, C.C. (2004). Translation of a Yeast Mitochondrial tRNA Synthetase Initiated at Redundant non-AUG Codons. J. Biol. Chem. *279*, 49656–49663. https://doi.org/10.1074/jbc.M408081200.

56. Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics *33*, 3387–3395. https://doi.org/10.1093/bioinformatics/btx431.

57. Karniely, S., Rayzner, A., Sass, E., and Pines, O. (2006). α-Complementation as a probe for dual localization of mitochondrial proteins. Exp. Cell Res. *312*, 3835–3846. https://doi.org/10.1016/j.yexcr.2006.08.021.

58. Wogulis, M., Chew, E.R., Donohoue, P.D., and Wilson, D.K. (2008). Identification of Formyl Kynurenine Formamidase and Kynurenine Aminotransferase from Saccharomyces cerevisiae Using Crystallographic, Bioinformatic and Biochemical Evidence. Biochemistry *47*, 1608–1621. https://doi.org/10.1021/bi701172v.

59. Hanekamp, T., and Thorsness, P.E. (1999). YNT20, a bypass suppressor of yme1 yme2, encodes a putative 3′-5′ exonuclease localized in mitochondria of Saccharomyces cerevisiae. Curr. Genet. *34*, 438–448. https://doi.org/10.1007/s002940050418.

60. van Hoof, A., Lennertz, P., and Parker, R. (2000). Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. EMBO J. *19*, 1357–1365. https://doi.org/10.1093/emboj/19.6.1357.

61. He, S., Bystricky, K., Leon, S., François, J.M., and Parrou, J.L. (2009). The Saccharomyces cerevisiae vacuolar acid trehalase is targeted at the cell surface for its physiological function. FEBS Journal *276*, 5432–5446. https://doi.org/10.1111/j.1742-4658.2009.07227.x.

62. Huang, J., Reggiori, F., and Klionsky, D.J. (2007). The transmembrane domain of acid trehalase mediates ubiquitin-independent multivesicular body pathway sorting. Mol. Biol. Cell *18*, 2511–2524. https://doi.org/10.1091/mbc.e06-11-0995.

63. Kearse, M.G., and Wilusz, J.E. (2017). Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. *31*, 1717–1731. https://doi.org/10.1101/gad.305250.117.

64. Kolitz, S.E., Takacs, J.E., and Lorsch, J.R. (2009). Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. RNA *15*, 138–152. https://doi.org/10.1261/rna.1318509.

65. Danpure, C.J. (1995). How can the products of a single gene be localized to more than one intracellular compartment? Trends Cell Biol. *5*, 230–238. https://doi.org/10.1016/S0962-8924(00)89016-9.

66. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

67. Wu, G., Xiao, M., Yang, C., and Yu, Y.T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J. *30*, 79–89. https://doi.org/10.1038/emboj.2010.316.

68. Wu, G., Radwan, M.K., Xiao, M., Adachi, H., Fan, J., and Yu, Y.T. (2016). The TOR signaling pathway regulates starvation-induced pseudouridylation of yeast U2 snRNA. RNA *22*, 1146–1152. https://doi.org/10.1261/rna.056796.116.

69. Grosshans, H., Lecointe, F., Grosjean, H., Hurt, E., and Simos, G. (2001). Pus1p-dependent tRNA Pseudouridinylation Becomes Essential When tRNA Biogenesis Is Compromised in Yeast. J. Biol. Chem. *276*, 46333–46339. https://doi.org/10.1074/jbc.M107141200.

70. Khonsari, B., and Klassen, R. (2020). Impact of Pus1 Pseudouridine Synthase on Specific Decoding Events in Saccharomyces cerevisiae. Biomolecules *10*, 729. https://doi.org/10.3390/biom10050729.

71. Wu, G., Adachi, H., Ge, J., Stephenson, D., Query, C.C., and Yu, Y.T. (2016). Pseudouridines in U2 snRNA stimulate the ATPase activity of

Prp5 during spliceosome assembly. EMBO J. *35*, 654–667. https://doi.org/10.15252/embj.201593113.

72. Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertins, P., Carr, S.A., Jovanovic, M., and Brar, G.A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. Cell *172*, 910–923.e16. https://doi.org/10.1016/j.cell.2018.01.035.

73. 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.

74. Longtine, M.S., Mckenzie, A., III, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961. https://doi.org/10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U.

75. Azizoglu, A., Brent, R., and Rudolf, F. (2021). A precisely adjustable, variation-suppressed eukaryotic transcriptional controller to enable genetic discovery. eLife *10*, e69549. https://doi.org/10.7554/eLife.69549.

76. van Overbeek, M., Sasaki, M., Tischfield, S.E., Sampson, J.-A., Globus, S.T., Socci, N.D., and Keeney, S. (2012). Revised genome assembly for the S. cerevisiae SK1 strain. https://www.yumpu.com/en/document/view/6419073/revised-genome-assembly-for-the-s-cerevisiae-sk1-strain-june-1-.

77. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566. https://doi.org/10.1038/nprot.2013.092.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-GFP	Clontech	Cat#632381, RRID:AB_2313808
Rabbit anti-hexokinase antibody	Rockland	Cat#100-4159, RRID:AB_21991
Anti-mouse 800 secondary	LI-COR	Cat#926-32210, RRID:AB_621842
Anti-rabbit 680 secondary	LI-COR	Cat#926-68071, RRID:AB_10956166
Chemicals, peptides, and recombinant proteins		
PMSF	Sigma	Cat#78830
cOmplete mini EDTA-free protease inhibitor cocktail	Roche	Cat#29384100
Acid-washed glass beads	Sigma	Cat#G8772
PBS Odyssey Blocking Buffer	LI-COR	Cat#927-40100
Rapamycin	Millipore	Cat#553211
Anhydrotetracycline	Millipore	Cat#37919
Acid phenol, pH4.3	Ambion	Cat#AM9722
Critical commercial assays		
Qubit RNA BR Assay	ThermoFisher	Cat#Q10211
NEXTFLEX Poly(A) Beads 2.0	Illumina	Cat#NOVA-512992
NEXTFLEX Rapid Directional RNA-Seq Kit 2.0	Illumina	Cat#NOVA-5198-02
Deposited data		
Raw sequencing data for Pus1^{truncation} mRNAseq	This paper	GSE235852
TIS-profiling data	Eisenberg et al.[6]	GSE150375
Standard Ribosome profiling data	Cheng et al.[72]	GSE108778
Experimental models: Organisms/strains		
All yeast strains are listed in Table S6		N/A
Recombinant DNA		
All plasmids used are listed in Table S6		N/A
Software and algorithms		
Truncation calling algorithm	This paper	https://github.com/alhigdon92/Truncated-protein-isoforms
STAR	Dobin et al.[73]	https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf
ImageStudio Lite Software	LI-COR	https://www.licor.com/bio/image-studio-lite/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gloria Brar (gabrar@berkeley.edu).

Materials availability

All yeast strains used in this study are available upon request from the corresponding author.

Data and code availability

- Sequencing data have been deposited at NCBI GEO and are publicly available as of the date of publication. The accession number is listed in the key resources table.
- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Yeast strain construction

Strains were constructed in the SK1 background of *Saccharomyces cerevisiae*. Strains and plasmids used for this study are listed in Tables S6 and S7. Deletion strains were created using pÜB81, and C-terminal 3V5 or FLAG-tagged strains were generated via Pringle tagging at the endogenous locus using pÜB81 or pÜB166.[74] GFP-tagged strains for microscopy were generated using Pmel-digested single-integration plasmids constructed via Gibson assembly of PCR-amplified fragments containing the ORF of interest along with its own 5' leader region amplified from genomic DNA and backbone fragments containing either a GFP tag, *ADH1* terminator, and a *TRP1* selection marker (pÜB629) or an mCherry tag, *ADH1* terminator, and a *HIS3* selection marker (pÜB1736). Start codon mutants were generated from single-integration plasmids described above by PCR amplifying fragments for Gibson assembly using primers containing the desired point mutation. Pus1^{truncation} overexpression strains were generated following the WTC$_{846}$ system.[75] The truncated open reading frame sequence was inserted into a single-integration plasmid downstream of pTetO7.1 by Gibson assembly of PCR fragments from genomic DNA and backbone fragments from pUB2344. Transformants were crossed into strains containing pRNR2-TetR-Tup1 and pTetO7.1-TetR.

Yeast growth and sporulation

For vegetative experiments, strains were grown in YEPD at 30°C. Strains were inoculated and grown overnight to reach saturation (OD600 > 10), then back-diluted to an OD600 of 0.2 and grown to desired OD600. For meiotic time courses, strains were inoculated into YEPD supplemented with uracil and tryptophan (1% yeast extract, 2% peptone, 2% glucose, 22.4 mg/L uracil, and 80 mg/L tryptophan) and grown for 24h at RT to an OD600 ≥ 10, then diluted to an OD600 of 0.25 in buffered YTA (1% yeast extract, 2% bacto tryptone, 1% potassium acetate, and 50 mM potassium phthalate) and grown for 16h at 30°C to an OD$_{600}$ ≥ 5. Cells were spun down and washed once with sterile MilliQ water before resuspension in sporulation media (SPO; 2% potassium acetate supplemented with amino acids (40 mg/L adenine, 40 mg/L uracil, 10 mg/L histidine, 10 mg/L leucine and 10 mg/L tryptophan)) at OD600 = 1.85 and shaken at 30°C, with timepoints collected at times indicated in figures.

METHOD DETAILS

Protein extraction and western blotting

Strains were grown in specified media and 2 or 3.3 OD600 equivalents of cells were collected for vegetative and meiotic cultures, respectively. Samples were incubated in 5% TCA for ≥10mins at 4°C then spun down, washed once with TE, once with acetone, then dried overnight. Pellets were resuspended in 150ul of lysis buffer (50mM Tris-HCl, 1mM EDTA, 3mM DTT, 1.1mM PMSF (Sigma), and 1X cOmplete mini EDTA-free protease inhibitor cocktail (Roche)) and cells were lysed by bead-beating for 5min at RT. SDS loading buffer was added to 1X and samples were incubated at 50°C for 10min and beads were pelleted by centrifugation. Samples were run on a 4-12% Bis-Tris gel at 160V for 5min followed by 200V for 25min. Transfer to nitrocellulose membrane was performed using a semi-dry transfer system (Trans-Blot Turbo, BioRad) with a standard 30 min transfer. The membrane was blocked in 5% milk PBS-T for 1 hour at RT and incubated in primary antibody overnight at 4°C. Primary antibodies were diluted in 5% milk in PBS-T + 0.01% sodium azide (1:2,000 for mouse anti-GFP (Clontech) and mouse anti-3V5 (Invitrogen), 1:1000 for mouse anti-FLAG (Sigma), and 1:10,000 for rabbit anti-hexokinase (Rockland). Membrane was washed 3X in PBS-T then incubated in secondary antibody (1:15,000 anti-mouse 800 and anti-rabbit 680 in LI-COR PBS blocking buffer) for 1 hour at RT, then washed 3X in PBS-T before imaging on the LI-COR Odyssey Imager. Analysis and quantification was performed using ImageStudio Lite software.

Proteasome inhibition

Strains were constructed in a *pdr5Δ* background to confer drug sensitivity. Standard meiosis conditions were used as described above. At the designated time point, cultures were split into two cultures and 100uM MG132 or DMSO (vehicle) was added. Cells were collected for protein extraction and western blot as described above at time points indicated in figures.

Growth in non-fermentable media

Cells were grown to saturation overnight in YEPD then back-diluted to an OD600 of 0.2 in YEPD. At 4h post-dilution, cells were spun down and resuspended in either YEPD (fermentable) or YEPG (non-fermentable). Cells were collected for protein extraction and western blot as described above at time points indicated in figures.

Rapamycin treatment

Cells were grown to saturation overnight in YEPD then back-diluted to an OD600 of 0.2 in YEPD. At 2 hours, cultures were split and treated with either rapamycin (0.2ug/ml or 0.5 ug/ml) or DMSO (vehicle). Cells were collected for protein extraction and western blot as described above at time points indicated in figures.

Low glucose growth

Cells were grown to saturation overnight in YEPD then back-diluted to an OD600 of 0.2 in either YEPD with 2% dextrose (normal) or 0.2% dextrose (low). Cells were collected for protein extraction and western blot as described above at time points indicated in figures.

Pus1^truncation overexpression

Cultures were grown to saturation overnight in YEPD then back-diluted to an OD600 of 0.2 in YEPD and treated immediately with either 1ug/ml anhydrotetracycline (aTC) or DMSO (vehicle). Samples were collected 3h post-dilution.

RNA extraction

5ODs of cells were pelleted by centrifugation and flash frozen in liquid nitrogen. Cells were thawed on ice and resuspended in TES buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5% SDS). An equal volume of acid phenol (pH4.3, Sigma-Aldrich) was added. Samples were shaken at 1400rpm for 30min at 65°C, then spun down at 4°C. The aqueous phase was transferred to a new tube containing 350ul chloroform. Samples were spun down and the aqueous layer was transferred to a new tube containing 100% isopropanol and with 350mM sodium acetate (pH5.2). Samples were precipitated overnight at -20°C. RNA was pelleted by centrifugation and pellets were washed with 80% ethanol, dried, resuspended in DEPC water for 10min at 37°C. Total RNA was quantified using the Qubit RNA BR Assay Kit (ThermoFisher).

Poly-A selection and RNA-seq

Poly-A selection was performed using the NEXTFLEX Poly(A) Beads 2.0 kit with 5ug total RNA (NOVA-512992). RNA-seq libraries were prepared from the resulting poly-A selected RNA using the NEXTFLEX Rapid Directional RNA-Seq Kit 2.0 (NOVA-5198-02). Libraries were quantified and quality checked using the Agilent 4200 TapeStation (Agilent Biotechnologies Inc). Samples were sequenced on the NovaSeqX sequencer.

Live imaging

At designated time points 2ul of meiotic culture was placed on a glass slide and imaged immediately. Images were acquired using a DeltaVision Elite wide-field fluorescence microscope (GE Healthcare), a 100X/1.40 NA oil-immersion objective (DeltaVision, GE Healthcare, Sunnyvale, CA), and the following filters: FITC, mCherry, DAPI. 30 z-stacks were collected with 0.2uM spacing. Images were deconvolved using softWoRx imaging software (GE Healthcare).

Ribosome profiling metagene analysis

Reads were averaged across all truncations at all positions between -50nt and +100nt surrounding truncated isoform TISs. We excluded truncations that begin within 50nt of the annotated isoform to avoid including reads associated with annotated start peaks that would confound the 5' signal. To prevent the profile from being overpowered by single highly expressed genes, we excluded genes with a Z-score greater than 10 at any position.

TL-seq metagene analysis

Reads were summed across all truncations at all positions between -200nt and +200nt surrounding annotated or truncated isoform TISs. We excluded truncations that begin within 200nt of the annotated isoform to avoid including reads associated with annotated TSSs. To prevent the profile from being overpowered by single highly expressed genes, we excluded genes with a Z-score greater than 10 at any position.

Staging comparison between TIS-profiling and TL-seq time courses

Stage matching between time courses was performed using an mRNA-seq time course collected in parallel with the TL-seq time course and an mRNA-seq time course collected under matched strain and growth conditions as the TIS-profiling time course. Note that staging was performed differently for the two time courses – the TIS-profiling was synchronized naturally via starvation conditions, whereas the TL-seq time course was synchronized via inducible expression of meiotic master regulator transcription factors *IME1* and *NDT80*. See Chia et al.[5] and Cheng et al.[72] for details. Timepoints for each time course were split into either early-meiotic (TIS-profiling: 0h, 1.5h, 3h, 4.5h; TL-seq: 0h, 2h, 3h, 4h, 5h, 6h) or late-meiotic (TIS-profiling: 4.5h, 6h, 8h, 10h; TL-seq: 5h, 6h, 7h, 8h, 9h) based on Pearson correlation of $\log_2$-transformed RPKMs (reads per kilobase million) and expression patterns of key meiotic genes in the mRNA-seq time courses (Figures S3A and S3B).

Localization prediction

DeepLoc1.0 was run on the amino acid sequence of all called truncated isoforms as well as their corresponding annotated isoform.[56] As a control, we generated sets of simulated truncations by randomly sampling in-frame ATGs within annotated genes that are not

called as real start sites. To ensure that the length distribution of the control set approximately matched the set of real truncations, for each real truncation we randomly sampled an in-frame start with the distance from annotated isoform within +/- 5 amino acids of that of the real truncation.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequence alignment, quantification, and differential expression analysis

Sequencing data were aligned to the SK1 genome using STAR.[76] A-site mapping for standard ribosome profiling and TIS-profiling data was performed as previously described.[6] Differential expression analysis for mRNA-seq data was performed using DESeq2. Hierarchical clustering was performed using complete-linkage clustering on the Pearson correlation of the $\log_2$-transformed average of 2 replicates. Genome browser visualization was performed using IGV.

Gene ontology enrichment analysis

GO analysis was performed using the PANTHER classification system.[77]

Truncation calling algorithm

Analysis was performed on TIS-profiling data collected at 0h, 1.5h, 3h, 4.5h, 6h, 8h, 10h, and 22h after addition to sporulation media (SPO), as well as in vegetative exponential, vegetative saturated growth and a MATa/a non-meiotic starvation control collected at 4.5h in SPO, as described in Eisenberg et al.[6] To generate a list of putative truncation-generating start codons, we first found all in-frame start codons within annotated exons. For each potential start codon (ATG) at each timepoint, a "peak sum" was calculated by summing the reads at the three nucleotides corresponding to the start codon. In-frame AUGs with a peak sum of less than or equal to 5 were excluded from further analysis. To model the background reads for each gene at each time point, we generated an empirical distribution of peaks sums from sets of three independent nucleotides that were randomly sampled with replacement (10,000x). The empirical p-value for each putative start codon, including annotated start codons, was determined by comparing the peak sum for the codon of interest to the empirical distribution. Annotated and putative truncation start codons were then filtered with the following criteria: p-value ≤ 0.0015 and >11 reads for at least one nucleotide in the start codon. This p-value cutoff corresponds to a Benjamini-Hochberg FDR of <0.005. To be considered in the final set, each truncation was required to be called at 2 or more timepoints. Putative truncations were additionally required to start ≥ 5aa from the annotated start codon and have an ORF length >10aa. Cases of likely mis-annotation, where the "truncated" isoform is likely the dominant isoform, were also removed; this gene set was generating through computational filtering to identify genes where the annotated isoform was not called followed by manual curation through visualization in a genome browser. To define the full truncated isoform sequence, we assumed that the truncated isoforms have the same stop codon as their corresponding annotated isoform. As we previously noted, peak heights are not strictly quantitative but we did notice the aggregate trend that annotated isoforms tended to have higher peaks than truncated isoforms, which may reflect a general tendency for lower expression of truncated protein isoforms (Figure S1D).[6]

TL-seq peak calling

Counts per site were extracted from published bigwig files using custom scripts.[5] To call protein isoforms with 5' transcript ends upstream of their start codons, for each gene an upstream-to-downstream ratio was calculated, such that { ratio = sum(reads 200nt upstream)/sum(reads 200nt downstream) }. Each gene's upstream-to-downstream ratio was compared to an empirical distribution of 10,000 random ratios obtained by taking the ratio of the sums of two randomly sampled groups of 200 sites within the gene, sampled with replacement. Reads upstream of the annotated TIS were masked to avoid including reads derived from 5' ends of annotated transcript isoforms. A p-value was calculated using the empirical cumulative distribution function of these ratios, and a p-value cutoff of 0.1 was used. To exclude genes with very sparse or no coverage we additionally required a variance greater than 0.05 for the distribution of sample ratios.