Published in final edited form as:
Cell Rep. 2014 June 26; 7(6): 1858–1866. doi:10.1016/j.celrep.2014.05.023.

Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*

Jenna E. Smith[1], **Juan R. Alvarez-Dominguez**[2], **Nicholas Kline**[1], **Nathan J. Huynh**[1], **Sarah Geisler**[1,3], **Wenqian Hu**[2], **Jeff Coller**[1], and **Kristian E. Baker**[1,*]

[1]Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH, 44106 USA

[2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142 USA

SUMMARY

High-throughput gene expression analysis has revealed a plethora of previously undetected transcripts in eukaryotic cells. In this study we investigate >1100 unannotated transcripts in yeast predicted to lack protein-coding capacity. We show that a majority of these RNAs are enriched on polyribosomes akin to mRNAs. Ribosome profiling demonstrates that many bind translocating ribosomes within predicted open-reading frames 10–96 codons in size. We validate expression of peptides encoded within a subset of these RNAs and provide evidence for conservation among yeast species. Consistent with their translation, many of these transcripts are targeted for degradation by the translation-dependent, nonsense-mediated RNA decay (NMD) pathway. We identify lncRNAs also sensitive to NMD, indicating translation of non-coding transcripts also occurs in mammals. These data demonstrate transcripts considered to lack coding potential are *bona fide* protein-coding, and expand the proteome of yeast and possibly other eukaryotes.

INTRODUCTION

The recent advent of high-throughput DNA sequencing technologies has led to the detection of a plethora of novel RNA transcripts and the revelation that vast regions of the genome once thought to be transcriptionally silent are, in fact, actively engaged by RNA polymerases (ENCODE Project Consortium, 2011). While some of these RNA products arguably represent transcriptional noise, a growing body of evidence suggests that many may have *bona fide* function in the cell. In particular, long non-coding RNAs (lncRNAs) have emerged as important regulators of gene expression, with established roles in

*Correspondence: K. Baker - keb22@case.edu, 216-368-0277, 216-368-2010.
[3]Present address: Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, 4058 Basel, Switzerland

ACCESSION NUMBERS

NCBI BioProject accession number for presented data is PRJNA245106.

epigenetic modification of chromatin, transcriptional control, and mRNA regulation post-transcriptionally (Geisler and Coller, 2013).

lncRNAs are classified based on transcript size (>200 nucleotides [nt] in length) and as lacking computationally-predicted protein coding regions of significant size and/or conservation (Derrien et al., 2012). The general assumption that lncRNAs are not translated is, however, at odds with their striking similarity to protein-coding mRNAs. Specifically, most lncRNAs are products of RNA polymerase II and harbor 5′ methyl-guanosine caps and 3′ termini of polyadenosine residues (Guttman et al., 2009) - key features promoting the efficient translation of mRNA. Indeed, investigation into a role for lncRNAs as templates for protein synthesis has suggested that these transcripts may associate with the cellular translation machinery. Polyribosome purification and genome-wide ribosome profiling have shown that lncRNAs co-fractionate with and/or bind ribosomes (Ingolia et al., 2011; Chew et al., 2012; Brar et al., 2012; van Heesch et al., 2014). The predictive value of ribosome profiling to define protein-coding potential has, however, been recently challenged (Guttman et al., 2013), and the overall contribution to the proteome of peptides generated from translation of lncRNA is suggested to be low (Bánfai et al., 2012). Therefore, it remains unclear how widespread the translation of predicted non-coding RNAs may be and what percentage of lncRNAs function strictly as regulatory RNA.

Similar to metazoa, budding yeast *Saccharomyces cerevisiae* has been shown to express an extensive repertoire of novel transcripts (David et al., 2006; Nagalakshmi et al., 2008). Study of a limited number of RNAs in this class has implicated them in controlling gene expression generally through transcriptional regulation or interference (Geisler and Coller, 2013), however, like lncRNAs, the function of most unannotated transcripts in yeast and the extent of their biological role in the cell remains unknown. In this study we investigate hundreds of previously unannotated transcripts in yeast and provide strong evidence that many of these RNAs possess protein-coding capacity. Specifically, we find unannotated RNAs associate with polyribosomes to extents similar to mRNA and that they encode small open reading frames bound by ribosomes. Consistent with their translation, we observe a significant percentage of these RNAs are sensitive to nonsense-mediated RNA decay (NMD), a translation-dependent process. Similarly, we calculate that a subset of mammalian lncRNA are sensitive to NMD indicating that these transcripts are also substrates for translation. Together, our data expand the coding capacity of the yeast genome beyond the current annotation and suggest expression of dozens of short polypeptides from transcripts previously predicted to lack coding potential.

RESULTS

Hundreds of unannotated and previously unclassified RNA transcripts are expressed in S. cerevisiae

We performed genome-wide gene expression analysis using RNA-Seq to generate a global map of transcripts expressed in yeast. Whole-cell, steady-state RNA from wild-type cells was ribosomal RNA-depleted and used to construct strand-specific cDNA libraries that were analyzed by Illumina HiSeq to produce ~11–22 million uniquely mapped sequence reads (Table S1 and Figure S1A). Reads mapping to annotated features of the Ensembl sacCer2

Saccharomyces genome confirmed expression of 5066 protein-coding mRNA and classic noncoding RNA transcripts (ncRNA; e.g. snRNA, snoRNA). The remainder of reads mapped to unique and unannotated loci (see Supplemental Experimental Procedures; Roberts et al., 2011) revealing expression of 1146 transcripts with a length greater than 200 nt, herein referred to as unannotated RNAs (uRNAs; Table S2). A number of uRNAs are expressed from loci corresponding to transcripts previously described by our group as DCP2-sensitive, long noncoding RNAs (Geisler et al., 2012) or RNAs previously described as either stable unannotated transcripts (SUTs; Xu et al., 2009) or RNAs targeted for degradation by the ribonucleases RRP6 or XRN1 (CUTs; Xu et al., 2009, or XUTs; van Dijk et al., 2011, respectively; Figure 1A; see Supplemental Experimental Procedures). The remainder of uRNAs (~800) lack previous classification and include transcripts expressed from intergenic regions of the genome and antisense to annotated protein-coding genes.

A majority of uRNAs associate with polyribosomes akin to mRNA

The yeast genome has been exhaustively annotated for protein coding capacity, and the uRNAs we identified by RNA-Seq are predicted to lack protein-coding potential. Recent studies, however, have uncovered unexpected associations between predicted noncoding RNA and the translation machinery, leading us to directly assess whether uRNAs in yeast are, in fact, noncoding. To evaluate the translational status of uRNAs, we used polyribosome analysis to enrich translation complexes and their associated RNA by sedimentation of cell lysates through sucrose gradients. Gradient fractions corresponding to polysomes were pooled (Figure 1B) and isolated RNA analyzed by RNA-seq to provide a genome-wide view of polyribosome-associated RNA (i.e. Polysome-Seq). The ~23 million mapped reads (Table S1 and Figure S1B) were compared to RNA-Seq data generated from total RNA to generate a Translatability Score (TS) representing the relative ratio of polysome association for every cellular transcript.

As anticipated, classic ncRNAs were generally excluded from polyribosomes as represented by low Translatability Scores (Figure 1C&D; mean=0.24 +/− 0.19 SD). In contrast, protein-coding mRNAs spanned a large range of translatability reflecting differences in translation efficiency as well as different rates of co-translational degradation (mean=1.12 +/− 0.49 SD; Hu et al., 2009). Importantly, 98.98% of mRNA exhibited a Translatability Score greater than the mean score for classic ncRNA, demonstrating that polyribosome analysis provides an effective biochemical method to characterize the association of RNAs with the translation machinery. Analysis of the Translatability Score for uRNAs revealed a wide range of association with the translation machinery similar to that of mRNAs (mean=0.98 +/− 0.79 SD; Figure 1C&D), although with a distinct distribution pattern that cannot simply be attributed to differences in RNA length (Figure S1C). Critically, >95% of uRNAs have a translatability score greater than the mean for classic ncRNA highlighting a significant distinction between well characterized non-coding RNAs and transcripts predicted to be non-protein coding. These data reveal that uRNAs demonstrate a varying degree of association with ribosomes and provide preliminary evidence that many uRNAs in yeast engage the translation machinery.

Ribosome profiling reveals short ORFs within uRNAs

We performed ribosomal profiling to corroborate the association of uRNA with the translation machinery and define - at nucleotide resolution - the nature of interaction between each uRNA and 80S ribosomes (Ingolia et al., 2009; Figure 2A). To minimize recovery of non-ribosome-bound, nuclease-protected RNA fragments that can arise by this procedure (Guttman et al., 2013), RNase-digested cell lysates were subject to sucrose gradient centrifugation and the broadened 80S gradient fractions resulting from collapse of polyribosomes were exclusively selected (Figure 2B). High-throughput sequencing of 80S-bound material derived from wild-type cells generated ~3–6 million mapped non-ribosomal RNA reads (Table S1). Analysis of ribosome-protected fragments revealed >50% of uRNAs detected in this analysis (185 of 331) bound ribosomes at levels ≥10% of expressed transcript levels (see Supplemental Experimental Procedures). Importantly, when reads generated by ribosome profiling were compared to RNA-Seq reads of fragmented total RNA prepared in parallel, the resulting Footprinting Score correlated strongly with Translatability Scores calculated from Polysome-Seq (Figure S2A).

In addition to validating a large fraction of uRNAs in yeast are ribosome bound, analysis of the distribution of ribosome-protected fragments along uRNAs revealed two striking observations. First, the coverage area of fragments aligning to individual uRNAs was small, suggesting that these transcripts encode polypeptides of limited size (Figure 2C). Indeed, the average size of nuclease-protected regions on uRNAs was 365 nt, significantly smaller than annotated yeast coding regions (CDS) which average 1344 nt (Figure S2B). Second, the distribution of ribosome-protected fragments mapped predominantly proximal to the uRNA 5′ end, consistent with the scanning model of translation for mRNAs (Figure 2C; Kozak, 1989). Moreover, the distribution of 80S-protected RNA resulted, in some cases, in long regions of downstream RNA that does not appear to associate with ribosomes (discussed below).

To further resolve the protein coding potential for uRNAs based on ribosome profiling, we analyzed nuclease-protected fragments exactly 28 nt in length for their ability to predict periodicity - fragments which align to a single reading frame due to the 3 nt translocation of the ribosome along the RNA *in vivo* (Ingolia et al., 2009). Analysis of 28 nt reads mapping to annotated protein-coding genes demonstrated that >70% corresponded to the +1 frame position (Figure 2D), confirming codon-triplet phasing and a strong bias towards in-frame footprints as compared to fragmented input RNA. Strikingly, for uRNAs with sufficient 28-mer footprints, 61 of 80 transcripts had footprints mapping predominantly to a single frame (See Supplemental Experimental Procedures; Figure 2E). Moreover, for 53 of these, the ribosome-protected fragments clearly demarcated at least one reading frame flanked by canonical AUG initiation and translation termination codons (e.g. Figure 2F). Metagene analysis of ribosome footprints along mRNAs and uRNAs confirm the annotation of CDS and predicted ORFs, respectively (Figures S2C and S2D). Importantly, ORFs predicted to be encoded within uRNAs are small - between 10 and 100 amino acids - and will be referred to herein as short ORFs (sORFs; Table S3).

Evidence for expression of sORFs encoded within uRNAs

Several pieces of evidence indicated that a number of unannotated transcripts expressed in yeast are polyribosome-associated, enriched for 80S ribosome binding within a subregion of the transcript, and harbor translocating ribosomes seemingly engaged in protein synthesis. Inspection of sORF-containing uRNA expression indicated that these transcripts are present at levels equivalent to many mRNAs encoding short polypeptides (Figure S3), suggesting that the putative protein products encoded by uRNAs may be present at physiologically relevant levels and play important biological roles in the cell. To verify sORFs predicted by ribosome profiling can be translated *in vivo*, we epitope-tagged three individual sORFs at their chromosomal locus by homologous recombination (Longtine et al., 1998; Figure 3A). Polypeptide of the expected size was detected from one of these and was dependent upon insertion of the epitope in the correct predicted reading frame (Figure 3B), demonstrating sORF translation under endogenous conditions.

To avoid alteration of the genomic locus downstream of the sORF that occurs as a consequence of chromosomal gene tagging, we cloned DNA encoding five intergenic uRNAs and inserted sequences encoding an epitope tag precisely upstream of the predicted stop codon (Figure 3C). Using this approach, we observed peptide products from two predicted sORFs (Figure 3D). Importantly, uRNA transcription is driven by endogenous promoter elements within the cloned DNA and expressed transcripts harbor native leader and 3′ UTR sequences. These data provide clear evidence for *in vivo* translation of sORFs from uRNA predicted to lack protein coding potential.

sORFs are conserved within fungal species

As a means to evaluate if polypeptides encoded by sORFs have biological significance, we examined the level of evolutionary conservation within yeast. Importantly, 10 species spanning >100 million years of evolution across 12 distinct clades were evaluated (Kurtzman and Robnett, 2003), with the expectation that conservation of peptides amid such significant genetic divergence is indicative of selective pressure to maintain sORF expression. Comparison of peptide sequences predicted from uRNAs revealed that 39 sORFs exhibited varying levels of conservation within closely related species (with 20 sORFs displaying conservation between >1 species; Figure 3E and Table S4). Homologs for 6 of the most conserved polypeptides were detected within at least one fungal species outside of the *Saccharomyces sensu stricto* genus, with three of these found in strains predicted to diverge from *S. cerevisiae* over 100 million years ago. Importantly, 12 sORFs exhibited a bias towards synonymous mutation, demonstrating conservation at the level of peptide sequence that is not a consequence of conserved nucleotide sequence elements (Table S4; Zhang et al., 2006). Finally, sORFs for 14 uRNAs are encoded within conserved genomic regions identified by phastCons (Table S4; Siepel et al., 2005). Together, these data reveal evolutionary pressure to maintain expression of a subset of sORFs within yeast species and argue that the encoded polypeptides play important biological functions in the cell.

Numerous uRNAs are targets of nonsense-mediated RNA decay

Our mapping of ribosome-protected fragments revealed that the region of 80S coverage on many uRNAs was limited and concentrated proximal to the transcript 5′ end. Moreover, for a number of uRNAs, the predicted sORF was followed downstream by an extended stretch of unprotected RNA. Based on observations in yeast and metazoa implicating 3′ UTR length in targeting mRNA to rapid decay by the nonsense-mediated mRNA decay pathway (NMD; Muhlrad and Parker, 1999; Singh et al., 2008), we hypothesized that a subset of yeast uRNAs might also be targeted by NMD. Importantly, sensitivity of uRNAs to NMD would serve to provide additional evidence that these transcripts engage actively translocating ribosomes since NMD is strictly a translation-dependent process (Maquat, 2004).

To determine whether uRNA are sensitive to the NMD pathway, we performed RNA-Seq on steady-state RNA isolated from cells deficient in the NMD pathway (due to deletion of *UPF1* encoding a key component of the NMD machinery; Leeds et al., 1991; Table S1). Comparison of RNA levels between wildtype and *upf1*Δ cells revealed 192 of 1146 uRNAs (16.8%) increased in abundance ≥2-fold in the absence of NMD (see Supplemental Experimental Procedures; Figure 4A&B), several of which were verified experimentally by Northern blot analysis (Figure 4C&S4A). While increased steady-state abundance in the absence of UPF1 does not differentiate direct versus indirect substrates of the NMD pathway, we found that NMD-sensitive uRNAs associated with polyribosomes to a similar extent as that observed for NMD-sensitive protein-coding mRNA (Figure S4B), and demonstrated dramatically higher average Translatability Scores compared to NMD-insensitive uRNAs (Figure S4C). Moreover, for individual transcripts, increased ribosome footprints were observed for NMD-sensitive uRNAs in the absence of UPF1, including *ICR1*, a characterized noncoding transcript previously shown to be sensitive to NMD (Toesca et al., 2011; Figure 4D, Table S1). The sensitivity of a subset of uRNAs to NMD and enhanced ribosome association in the absence of UPF1 provides further support that numerous uRNAs in yeast encode short open reading frames engaged by actively translating ribosomes.

We observed that the average length of RNA protected by ribosome footprints, although short, was not significantly different among uRNAs that were NMD-sensitive vs insensitive. In contrast, the length of RNA downstream of the ribosome-protected region was significantly longer for NMD-sensitive transcripts compared to those that did not respond to inactivation of the NMD pathway (891 nt +/− 64 SEM versus 287 nt +/− 50 SEM; Figure 4E). These findings are consistent with the observation that mRNAs in yeast with 3′ UTR lengths greater than 300 nt are efficiently targeted to NMD (Kebaara and Atkin, 2009), and provide a mechanistic explanation by which only a subset of ribosome-associated uRNAs are sensitive to NMD.

Sensitivity of lncRNA to NMD indicates translation of ‘noncoding’ transcripts in mammals

As a means to evaluate whether predicted non-protein coding transcripts in higher eukaryotes also encode sORFs that are translated, we evaluated recent genome-wide gene expression and UPF1 protein binding data gathered from mouse embryonic stem cells (mESC; Hurt et al., 2013). Our analysis identified 519 annotated mRNAs whose expression

increased >1.5-fold in cells inhibited for NMD versus control cells (of 13,043 expressed protein-coding genes; ~4%), many of which correspond to previously characterized NMD targets (Hurt et al., 2013). Strikingly, 46 transcripts classified as lncRNAs also increased >1.5-fold upon inhibition of NMD (of 265 lncRNA; Figure 4F). Consistent with these transcripts being direct targets for NMD, UPF1 binding sites were enriched 9.6-fold on these RNAs over NMD-insensitive lncRNAs (Figure S4D&E). These data provide evidence that a number of mammalian lncRNAs lacking predicted protein-coding potential are engaged in active translation.

In addition to a similar proportion of uRNAs and lncRNAs being sensitive to perturbations in the NMD pathway (16.8% and 17.4% in yeast and mESC, respectively), we observed that ribosome footprints are enriched specifically on NMD-sensitive lncRNAs upon NMD inhibition compared to NMD-insensitive transcripts, and that these 80S ribosome-protected fragments map proximal to the transcript 5′ end (Figure S4F). Based on the observation that 3′ UTR length also plays a role in targeting transcripts to NMD in mammalian cells (Singh et al., 2008), unprotected RNA downstream of putative coding regions within lncRNAs likely contributes to the sensitivity of these RNAs to NMD, and suggests a common mechanism by which such transcripts are subject to regulation by this cellular RNA surveillance pathway.

DISCUSSION

Our analysis of the global landscape of expressed transcripts in yeast revealed hundreds of previously uncharacterized RNAs that do not map to annotated, protein-coding gene loci. We show by a number of means, including polyribosome analysis, ribosome profiling, and NMD sensitivity, that many of these unannotated transcripts are associated and/or actively engaged with translating ribosomes. Moreover, periodicity observed for a subset of ribosome-protected fragments facilitated precise demarcation of open reading frames utilized by the translation machinery *in vivo*, providing heightened evidence for translation of defined short polypeptides encoded within a number of yeast uRNAs.

We demonstrate that a significant fraction of yeast uRNAs are sensitive to NMD, a translation-dependent surveillance pathway generally described to target mRNA. Moreover, analysis of published genome-wide expression data in mESC cells revealed a similar percentage of mammalian lncRNAs are also sensitive to NMD. Targeting of individual or subsets of predicted non-coding RNA to NMD has been previously observed in various organisms including yeast (Thompson and Parker, 2007; Toesca et al., 2011), plants (Kurihara et al., 2009) and human cells (Tani et al., 2013), and these data argue that predicted non-coding RNAs are present in the cell cytoplasm and, contrary to expectations, engage the translation machinery. Our ribosome profiling data extend these observations and not only predict short protein-coding sequences within these transcripts but also reveal extended regions of RNA downstream of predicted sORFs that are unprotected by 80S ribosomes. Importantly, these ribosome-free regions mimic long 3′ UTRs that commonly target mRNA to NMD in yeast and metazoa and provide a mechanistic explanation for how uRNAs (and lncRNAs) are targeted by this specialized decay pathway.

While sensitivity to NMD provides compelling evidence supporting translation of sORFs encoded within uRNAs, we note that the accelerated degradation of transcripts targeted by NMD would reduce steady-state levels of these uRNAs and effectively dampen expression of any polypeptide encoded by the predicted sORF. Biologically, the sensitivity of uRNAs to NMD may serve to ensure that these transcripts maintain a primary role as functional RNA molecules, either in the nucleus as regulators of transcriptional events or in the cytoplasm as modulators of mRNA and/or protein function. Alternatively, the degradation of uRNAs by NMD may provide a unique means to regulate sORF expression, allowing robust accumulation of small polypeptides under conditions when NMD efficiency is reduced or inactivated (Huang and Wilkinson, 2012).

At present we have demonstrated expression of polypeptides from two conserved yeast sORFs; however, a biological function for these and other predicted sORF translation products remains unclear. Notwithstanding, roles for small polypeptides in cellular function is well documented (Andrews and Rothnagel, 2014). In yeast, mating pheromones are 12 and 13 amino acids in length, and the large ribosomal protein L41 required for 25S rRNA folding is 25 amino acids. Systematic analysis of annotated yeast mRNAs encoding small ORFs (<100 codons) revealed dozens important for cell growth under various conditions (Kastenmayer et al., 2006). Recently, functional small peptides have been found expressed from predicted non-protein coding RNAs in flies (Galindo et al., 2007; Magny et al., 2013) and zebrafish (Pauli et al., 2014), and short polypeptides derived from lncRNAs have been detected in human cells (Slavoff et al., 2012). We predict, therefore, that a number of sORFs identified in this study will express peptide products with important biological roles in yeast.

Transcriptome analysis of polyribosome-associated RNA revealed a large percentage of uRNAs associated with polysomes, similar to that observed for mRNAs. The distribution of uRNA association with polysomes was, however, clearly distinct from that of mRNA. We attribute this difference to functional heterogeneity within the class of uRNAs as compared to mRNAs. In contrast to mRNAs whose primary role is as templates for protein synthesis, uRNAs identified in our analysis include transcripts which we demonstrate are translated and ones for which there is limited association with the translational machinery. It will be of interest to evaluate uRNAs with low Translatability Scores for function as RNA regulators and accumulation in various compartments within the cell. Indeed, visualization of several uRNAs of this type using single molecule fluorescence *in situ* hybridization suggests that these transcripts are enriched in the nucleus (data not shown). A likely role for these uRNAs is as regulators of gene expression through chromatin modification or influencing transcriptional events.

As a model eukaryote, *S. cerevisiae* has been the focus of extensive gene expression analysis and the proverbial guinea pig for many large scale genomic and transcriptomic experimental studies. Because of this attention, the yeast genome has been described in exquisite detail and is currently annotated to express 6380 protein-coding transcripts. As technologies measuring gene expression at finer resolution are developed or honed, previously undetected transcripts will continue to be uncovered. Our work adds to a number of recent studies identifying expression of RNA transcripts in yeast predicted to lack protein coding potential. Although a majority of these RNAs (and similar non-coding RNAs in metazoa) lack

characterized function in the cell, we show here that a number encode predicted sORFs exploited by the translational machinery for the expression of small polypeptides, some of which demonstrate evolutionary conservation. Our present findings reveal additional protein coding capacity within the yeast genome, but it will not be unexpected to learn that the remarkable complexity that continues to be uncovered in this single-celled eukaryote will also be found hidden in the genomes of other, more complex organisms, including humans.

EXPERIMENTAL PROCEDURES

Yeast culture and standard methods

Cells were grown under standard conditions, unless otherwise noted. Yeast strains, plasmids, and oligonucleotides are listed in Table S5. RNA isolation, northern and western blot analyses were performed as previously described (Geisler et al., 2012). Epitope-tagged sORFs were generated using homologous recombination (Longtine et al., 1998) or standard molecular cloning strategies.

Total RNA Library Preparation

5 μg of DNase I-treated whole-cell RNA was depleted of rRNA using Epicentre Human/Mouse/Rat RiboZero rRNA Removal Kit. Strand-specific, random-primed cDNA libraries were generated by the CWRU Genome and Transcriptome Sequencing Core using the Epicentre ScriptSeq v2 RNA-Seq Library Preparation Kit.

Polysome-associated RNA Library Preparation

Yeast whole-cell lysates were subjected to polyribosome analysis on a 15–45% (w/w) sucrose gradient. RNA was extracted from fractions containing polyribosomes and pooled. 5 μg of RNA was used to prepare libraries as described above.

Ribosome Profiling Library Preparation

Isolation and sequencing of ribosome-protected RNA fragments was performed based on the described protocol (Ingolia et al., 2012), with modifications as described in Supplemental Experimental Procedures. For fragmented total RNA libraries, whole-cell RNA was purified, DNase-treated, and rRNA depleted as for the total RNA library preparation. RNA was fragmented with base as described (Ingolia, 2010), and 26–34 nt fragments gel-purified and used for library preparation.

RNA Sequencing

cDNA libraries were sequenced on the Illumina HiSeq platform. Details of sequencing data analysis can be found in Supplemental Experimental Procedures.

Analysis of mESC Data

RNA-Seq, Ribo-Seq, and CLIP-Seq data generated by Hurt et al., 2013 were downloaded from the Gene Expression Omnibus (GSE41785). Details of data analysis can be found in Supplemental Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Tim Nilsen for critical evaluation of the manuscript, and members of the Baker and Coller labs for helpful insight into this work. This research was funded by the National Institute of General Medical Sciences (GM080465 for JC and GM095621 for KEB) and the National Science Foundation (NSF1253788 for KEB).

References

Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet. 2014; 15:193–204. [PubMed: 24514441]

Bánfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE Jr, Kundaje A, Gunawardena HP, Yu Y, Xie L, et al. Long noncoding RNAs are rarely translated in two human cell lines. Genome Res. 2012; 22:1646–1657. [PubMed: 22955977]

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science. 2012; 335:552–557. [PubMed: 22194413]

Chew GL, Pauli A, Rinn JL, Regev A, Schier AF, Valen E. Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. Development. 2013; 140:2828–2834. [PubMed: 23698349]

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM. A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA. 2006; 103:5320–5325. [PubMed: 16569694]

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012; 22:1775–1789. [PubMed: 22955988]

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biology. 2007; 5:1052–1062.

Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol. 2013; 14:699–712. [PubMed: 24105322]

Geisler S, Lojek L, Khalil AM, Baker KE, Coller J. Decapping of long noncoding RNAs regulates inducible genes. Mol Cell. 2012; 45:279–291. [PubMed: 22226051]

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009; 458:223–227. [PubMed: 19182780]

Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell. 2013; 154:240–251. [PubMed: 23810193]

Hu W, Sweet TJ, Chamnongpol S, Baker KE, Coller J. Co-translational mRNA decay in *Saccharomyces cerevisiae*. Nature. 2009; 461:225–229. [PubMed: 19701183]

Huang L, Wilkinson MF. Regulation of nonsense-mediated mRNA decay. Wiley Interdiscip Rev RNA. 2012; 3:807–828. [PubMed: 23027648]

Hurt JA, Robertson AD, Burge CB. Global analyses of UPF1 binding and function reveals expanded scope of nonsense-mediated mRNA decay. Genome Res. 2013; 10:1636–1650. [PubMed: 23766421]

Ingolia NT. Genome-wide translational profiling by ribosome footprinting. Methods Enzymol. 2010; 470:119–142. [PubMed: 20946809]

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. Nature Protocols. 2012; 7:1534–1550.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009; 324:218–223. [PubMed: 19213877]

Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011; 147:789–802. [PubMed: 22056041]

Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au W-C, Yang H, Carter CD, Wheeler D, Davis RW, Boeke JD, et al. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. Genome Res. 2006; 16:365–373. [PubMed: 16510898]

Kebaara BW, Atkin AL. Long 3′-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in *Saccharomyces cerevisiae*. Nucleic Acids Res. 2009; 37:2771–2778. [PubMed: 19270062]

Kozak M. The scanning model for translation: an update. J Cell Biol. 1989; 108:229–241. [PubMed: 2645293]

Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A, et al. Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. Proc. Natl. Acad. Sci. USA. 2009; 106:2453–2458.

Kurtzman CP, Robnett CJ. Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res. 2003; 3:417–432. [PubMed: 12748053]

Leeds P, Peltz SW, Jacobson A, Culbertson MR. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translation termination codon. Genes Dev. 1991; 5:2303–2314. [PubMed: 1748286]

Longtine MS, McKenzie A III, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Yeast. 1998; 14:953–961. [PubMed: 9717241]

Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, Couso JP. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science. 2013; 341:1116–1120. [PubMed: 23970561]

Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. 2004; 5:89–99. [PubMed: 15040442]

Muhlrad D, Parker R. Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance. RNA. 1999; 5:1299–1307. [PubMed: 10573121]

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008; 320:1344–1349. [PubMed: 18451266]

Pauli A, Norris MS, Valen E, Chew GL, Gagnon JA, Zimmerman S, Mitchell A, Ma J, Dubrulle J, Reyon D, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science. 2014; 343:1248636. [PubMed: 24407481]

Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics. 2011; 27:2325–2329. [PubMed: 21697122]

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005; 15:1034–1050. [PubMed: 16024819]

Singh G, Rebbapragada I, Lykke-Andersen J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol. 2008; 6:e111. [PubMed: 18447585]

Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nature Chemical Biology. 2012; 9:59–64.

Tani H, Torimura M, Akimitsu N. The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. PLoS One. 2013; 8:e55684. [PubMed: 23383264]

Thompson DM, Parker R. Cytoplasmic decay of intergenic transcripts in *Saccharomyces cerevisiae*. Mol Cell Biol. 2007; 27:92–101. [PubMed: 17074811]

Toesca I, Nery CR, Fernandez CF, Sayani S, Chanfreau GF. Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. PLoS Genet. 2011; 7:e1002163. [PubMed: 21738494]

van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Né P, Loeillet S, et al. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature. 2011; 475:114–117. [PubMed: 21697827]

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol. 2014; 15:R6. [PubMed: 24393600]

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009; 457:1033–1037. [PubMed: 19169243]

Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Geno Prot Bioinfo. 2006; 4:259–263.

HIGHLIGHTS

- Unannotated yeast RNAs predicted to lack coding capacity associate with polyribosomes
- Ribosome profiling reveals novel, short open-reading frames 10–96 codons in size
- Many yeast noncoding RNAs are sensitive to nonsense-mediated RNA decay
- Sensitivity of noncoding RNA to NMD is also observed in mammals

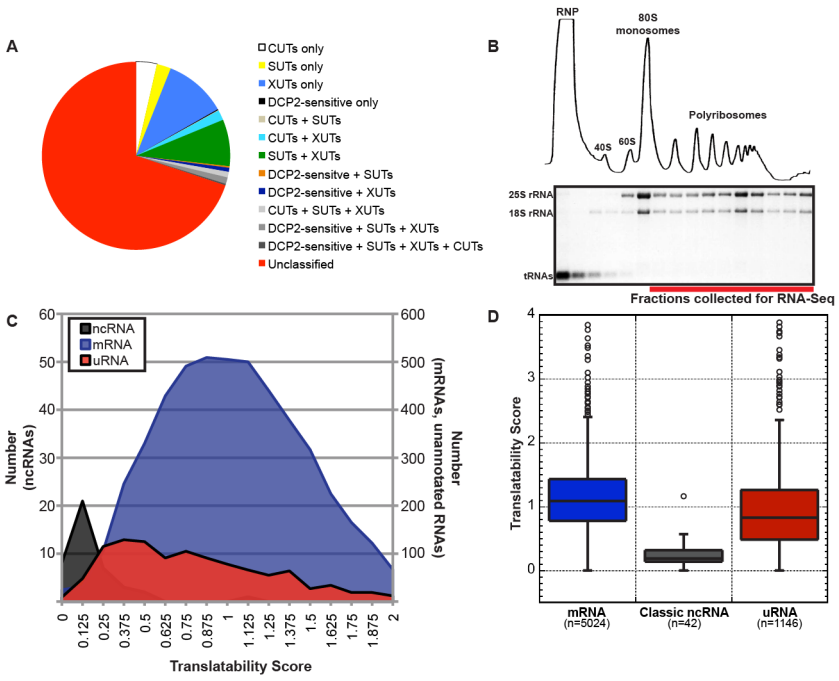

Figure 1. Yeast uRNAs co-sediment with polyribosomes

A) Overlap between uRNAs identified in this study and SUTs, XUTs, CUTs, and DCP2-sensitive lncRNAs (see Supplemental Experimental Procedures).

B) Polysome analysis of yeast cell lysates. Top - UV trace after sedimentation through sucrose gradients. Bottom - ethidium bromide stain of RNA isolated from each gradient fraction. RNA for Polysome-seq pooled from fractions indicated.

C) Translatability Score ($FPKM_{polysomes}/FPKM_{steady\text{-}state}$ based on averages of expression from two biological replicates) for characterized ncRNAs, mRNAs, and uRNAs.

D) Distribution of Translatability Scores as in (C) for each class of RNA. Box includes 25–75 percentiles; whiskers indicate +/− 1.5 IQRs, with outliers indicated by circles.

See also Figure S1, Tables S1–S2.

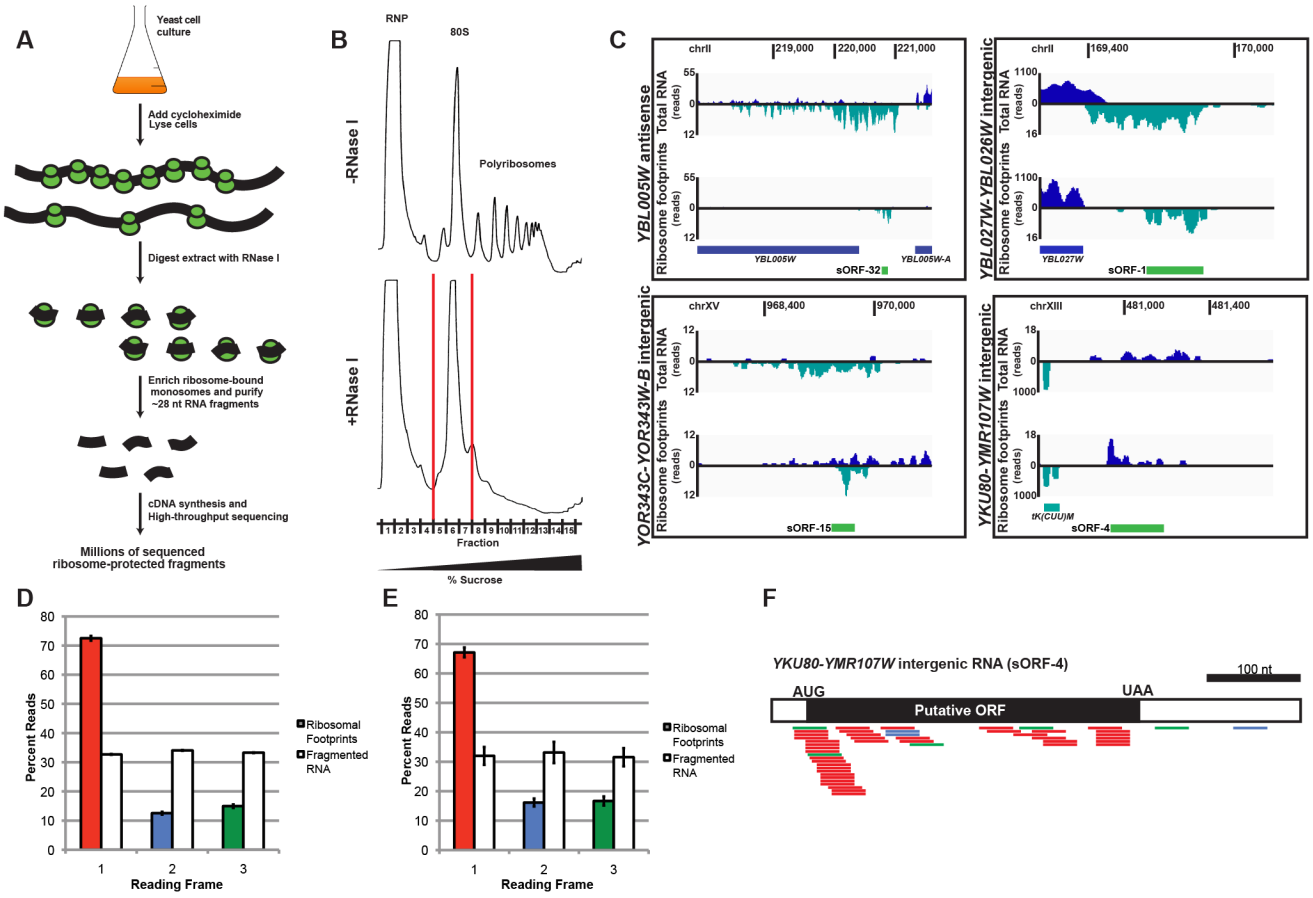

Figure 2. Ribosome profiling provides evidence for translation of uRNAs

A) Schematic of ribosome profiling protocol.

B) Representative UV trace of polyribosome gradients from cell lysates without (−) or with (+) RNase I treatment. Fractions encompassing the collapsed 80S peak following RNase I-treatment collected for analysis are indicated.

C) RNA-Seq and ribosome footprints for sample uRNAs. Watson strand (navy); Crick strand (teal). Annotated genes (navy or teal bars) and putative sORFs delineated by ribosome footprints (green bars) are indicated.

D) Fraction of 28 nt ribosome-protected fragments (RPFs) mapping to each of 3 frames for annotated mRNAs. Two biological replicates of each WT and *upf1*Δ ribosome footprints were analyzed as 4 independent samples and single replicates of each WT and *upf1*Δ fragmented RNA were analyzed as 2 independent samples. Data are mean +/− SEM.

E) Fraction of 28 nt RPFs mapping to each of 3 frames for the 61 uRNAs demonstrating ribosome phasing (where ≥50% of RPFs mapped to a single frame). For each uRNA, the +1 frame was retrospectively classified. Each uRNA was teated as a single replicate; data shown as mean +/− SEM.

F) 28 nt RFPs mapping to *YKU80-YMR107W* intergenic uRNA demonstrate phasing and delineate an ORF within AUG start and UAA stop codons. RPFs colored based on frame to which they map as in (D) and (E).

See also Figure S2, Table S1.

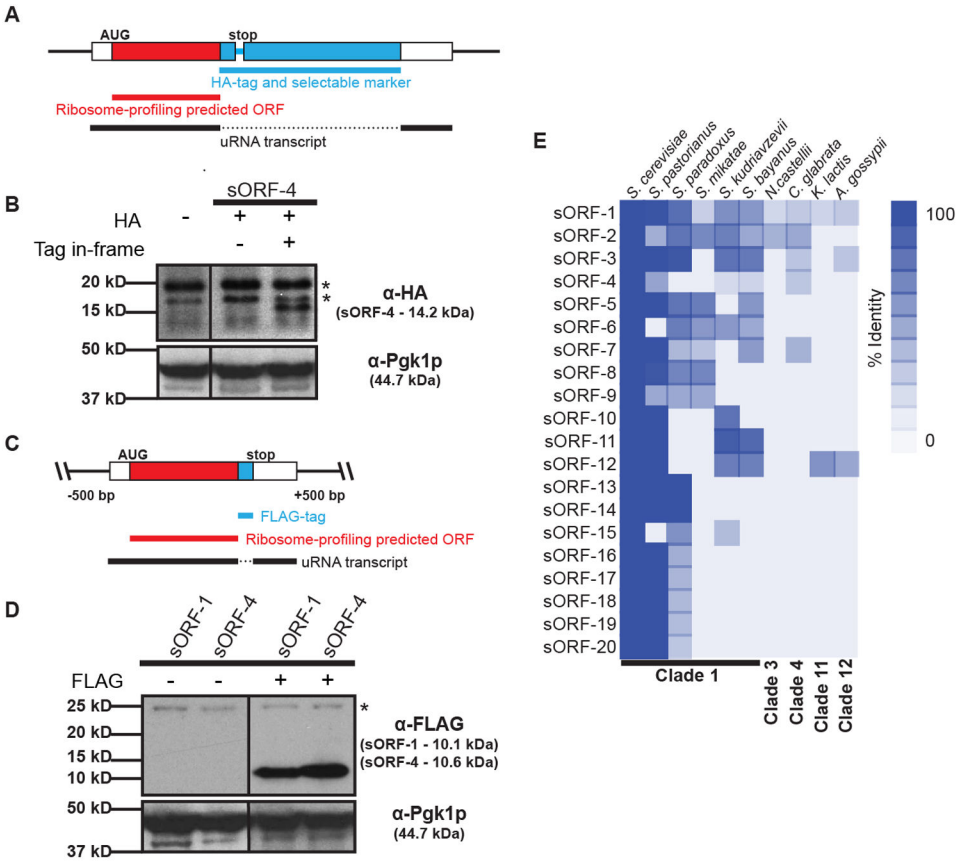

Figure 3. Evidence for expression and conservation of sORFs

A) Epitope tagging of putative sORFs at their endogenous chromosomal locus by homologous recombination. Solid black line represents uRNA defined by RNA-Seq.

B) Western blot analysis detects the translation product of chromosomally-tagged sORF-4. Signal is specific to in-frame tag and corresponds to molecular weight for the chimeric peptide. Asterisk indicates a non-specific signal. Pgk1p serves as loading control.

C) Genomic DNA flanking mapped uRNAs was cloned and the putative sORF epitope tagged at its C-terminus. Solid black line represents uRNA defined by RNA-Seq.

D) Western blot detects translation of yeast sORF-1 and sORF-4. Signal is specific for epitope-tagged sORF and corresponds to expected molecular weight for each chimeric peptide. Asterisk indicates a non-specific signal. Pgk1p serves as loading control.

E) Conservation of sORFs among divergent yeast species. Putative peptides encoded by sORFs were identified in other yeast species based on 6-frame translation using TBLASTN. Percent identical residues relative to full-length putative peptide indicated. Top 20 most conserved candidates shown.

See also Figure S3, Tables S3–S4.

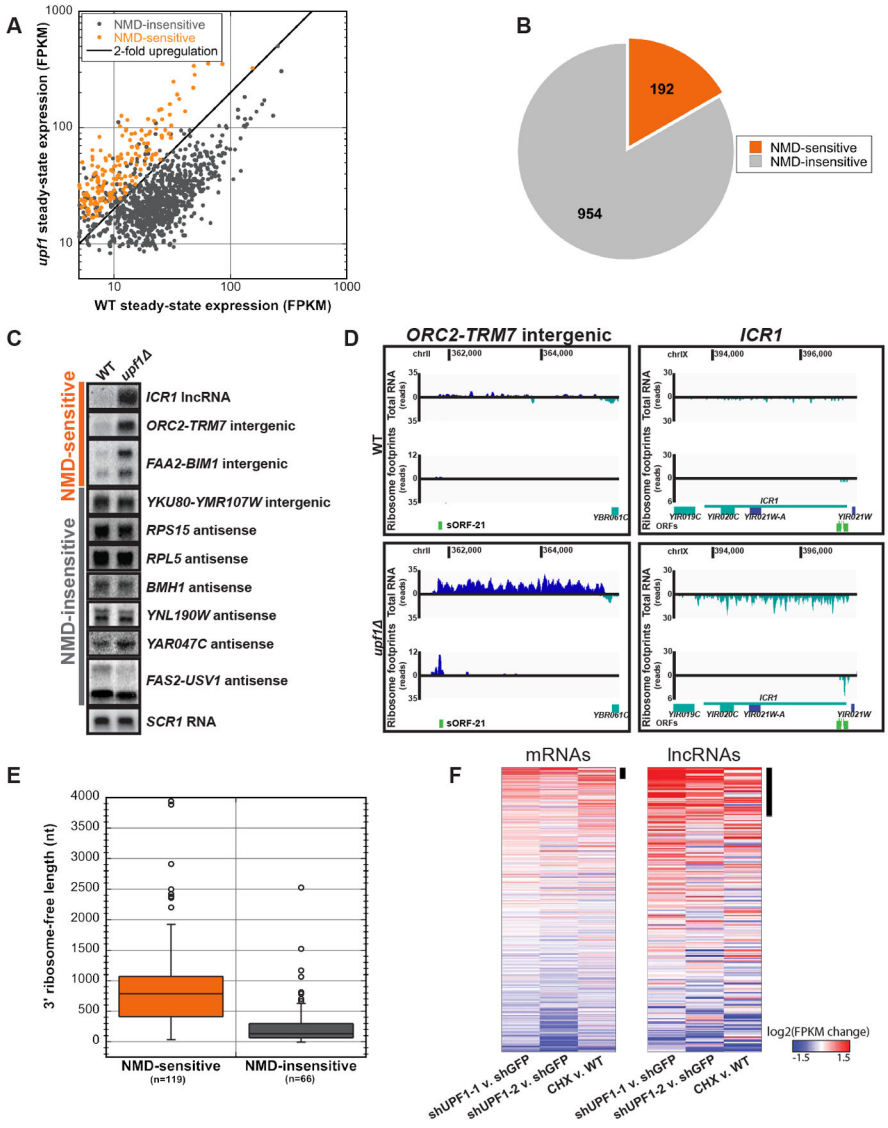

Figure 4. uRNAs are subject to translation-dependent nonsense-mediated RNA decay

A) uRNA expression levels (FPKM) in WT versus *upf1*Δ measured by RNA-seq reveals sensitivity to NMD. NMD-sensitive uRNAs exhibit ≥2-fold increase in steady-state levels in *upf1*Δ (statistically significant at an FDR<0.05 by Cuffdiff analysis; orange).

B) Fraction of uRNAs showing sensitivity to NMD.

C) Northern blot analysis of steady state RNA from WT and *upf1*Δ cells shows uRNAs and lncRNA *ICR1* predicted by RNA-Seq to be regulated by NMD. Representative *SCR1* loading control is shown.

D) Sequence coverage for NMD-sensitive uRNA or lncRNA *ICR1* in WT (top) or NMD-deficient (*upf1*Δ) cells (bottom). Data as in Figure 2C.

E) Length distribution of downstream ribosome-free regions for NMD-sensitive and -insensitive uRNAs. Box includes 25–75 percentiles; whiskers indicate +/− 1.5 IQRs, with outliers indicated by circles.

F) Change in mRNA and lncRNA expresssion in each of three NMD inhibition experiments in mESCs (shRNA UPF1-1, shRNA UPF1-2, and cycloheximide [CHX] treatment; Hurt et al, 2013). Changes are log2 expression (FPKM) ratios over control, averaged over 2 replicates. Potential NMD targets, defined as genes derepressed >1.5-fold, are highlighted (black bar).

See also Figure S4.