Translation is a key determinant controlling the fate of cytoplasmic long non-coding RNAs

Sara ANDJUS[1,5], Ugo SZACHNOWSKI[2,5], Nicolas VOGT[2], Isabelle HATIN[3], Chris PAPADOPOULOS[4], Anne LOPES[4], Olivier NAMY[3], Maxime WERY[2,6,7] and Antonin MORILLON[2,6,7]

[1] ncRNA, epigenetic and genome fluidity, Institut Curie, PSL University, Sorbonne Université, CNRS UMR3244, 26 rue d'Ulm, F-75248 Paris Cedex 05, France

[2] ncRNA, epigenetic and genome fluidity, Institut Curie, Sorbonne Université, CNRS UMR3244, 26 rue d'Ulm, F-75248 Paris Cedex 05, France

[3] Genomics, Structure and Translation, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette cedex, France

[4] Molecular Bio-informatics, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette cedex, France

[5] Co-first authors

[6] Corresponding authors

[7] Co-last authors

Running title: Translation of yeast cytoplasmic lncRNAs

Summary

Despite predicted to lack coding potential, cytoplasmic long non-coding (lnc)RNAs can associate with ribosomes, resulting in some cases into the production of functional peptides. However, the biological and mechanistic relevance of this pervasive lncRNAs translation remains poorly studied. In yeast, cytoplasmic Xrn1-sensitive lncRNAs (XUTs) are targeted by the Nonsense-Mediated mRNA Decay (NMD), suggesting a translation-dependent degradation process. Here, we report that XUTs are translated, which impacts their abundance. We show that XUTs globally accumulate upon translation elongation inhibition, but not when initial ribosome loading is impaired. Translation also affects XUTs independently of NMD, by interfering with their decapping. Ribo-Seq confirmed ribosomes binding to XUTs and identified actively translated small ORFs in their 5'-proximal region. Mechanistic analyses revealed that their NMD-sensitivity depends on the 3'-untranslated region length. Finally, we detected the peptide derived from the translation of an NMD-sensitive XUT reporter in NMD-competent cells. Our work highlights the role of translation in the metabolism of XUTs, which could contribute to expose genetic novelty to the natural selection, while NMD restricts their expression.

Keywords: lncRNA/Xrn1/NMD/translation

INTRODUCTION

Long non-coding (lnc)RNAs constitute a class of transcripts that arise from the pervasive transcription of eukaryotic genomes (Jarroux et al., 2017). Even if the debate on their functional significance is still open (Ponting and Haerty, 2022), some of them are now recognized as important RNA regulators involved in multiple cellular functions (Kopp and Mendell, 2018; Statello et al., 2021; Yao et al., 2019). Consistent with their functional importance, their expression appears to be precisely controlled (Djebali et al., 2012; Lorenzi et al., 2021). Furthermore, the abnormal expression of lncRNAs is associated with human diseases, including cancers (Renganathan and Felley-Bosco, 2017; Saha et al., 2017; Schmitt and Chang, 2016). However, these evidence remain marginal and full mechanistic description is still required to understand the *raison d'être* of lncRNAs in cells.

By definition, lncRNAs have been predicted to lack coding potential. However, this assumption has been challenged by several independent observations, showing that cytoplasmic lncRNAs can associate with ribosomes (Carlevaro-Fita et al., 2016; Ingolia et al., 2014; Ingolia et al., 2011; van Heesch et al., 2014). In fact, ribosome profiling (Ribo-Seq) analyses have revealed small open reading frames (smORFs) on lncRNAs (Aspden et al., 2014; Chen et al., 2020; Ingolia et al., 2014; Smith et al., 2014), the translation of which resulting, in some cases, into the production of functional peptides (D'Lima et al., 2017; Matsumoto et al., 2017; Slavoff et al., 2013; van Heesch et al., 2019; Zanet et al., 2015).

Beside these examples of functional lncRNA-derived peptides, which remain a minority to date, the biological relevance of this 'pervasive' translation of lncRNAs remains unclear. In this regard, an emerging view in the field proposes that lncRNAs could constitute a reservoir of rapidly evolving smORFs in which the cell can get to explore the potential of genetic novelty and produce novel peptides (Ruiz-Orera et al., 2014). If beneficial, lncRNA-derived peptides could be selected, thereby contributing to the emergence of novel protein-coding genes through the evolutionary process known as *de novo* gene birth (Blevins et al., 2021; Carvunis et al., 2012; McLysaght and Hurst, 2016; Papadopoulos et al., 2021; Schmitz et al., 2018; Van Oss and Carvunis, 2019; Zhao et al., 2014).

In the budding yeast *Saccharomyces cerevisiae*, the idea that cytoplasmic lncRNAs are also pervasively translated is mainly supported by their sensitivity to the Nonsense-Mediated mRNA Decay (NMD). NMD is a conserved translation-dependent RNA decay pathway known to target mRNAs bearing premature stop codons (Losson and Lacroute, 1979), although such 'aberrant' transcripts represent only one type of NMD substrates (for review, see (Andjus et al., 2021)). Actually, most yeast cytoplasmic lncRNAs, previously annotated as Xrn1-sensitive Unstable Transcripts (XUTs) due to their extensive degradation by the cytoplasmic 5'-exonuclease Xrn1 (Van Dijk et al., 2011), turned out to be NMD substrates (Malabat et al., 2015; Wery et al., 2016).

For yeast mRNAs, the length of the long 3' untranslated region (UTR) downstream of the termination codon is known to be critical for NMD activation (Amrani et al., 2004; Celik et al., 2017; Muhlrad and Parker, 1999). Mechanistically, the interaction between the poly(A) binding protein (Pab1) and the eukaryotic release factors eRF1/eRF3, which normally promotes efficient translation termination, is impeded by the long 3' UTR. Instead, this favors the recruitment of the NMD core factor Upf1 by eRF1/eRF3, leading to the formation of an NMD complex at the level of the terminating ribosome.

Consistent with the view that XUTs would be translated and that this would determine their degradation by NMD, the analysis of pioneer Ribo-Seq data obtained in Upf1-lacking yeast cells revealed ribosome footprints in the 5' region of some NMD-sensitive XUTs, followed by a long ribosome-free 3' region (Smith *et al.*, 2014; Wery *et al.*, 2016). However, the coverage of this unique early dataset was not sufficient to allow a robust systematic identification of actively translated smORFs within XUTs, nor to unveil the equilibrium between their translation and their decay. Furthermore, the biological and mechanistic relevance of XUTs translation remained unknown.

Here, we investigated the impact of translation on the fate of cytoplasmic lncRNAs, using XUTs as a paradigm. We found that NMD-sensitive XUTs rapidly accumulate in wild-type (WT) yeast cells treated with translation elongation inhibitors. Besides NMD, our data indicate that translation can also affect XUTs decay in an NMD-independent manner, by interfering with their decapping. In contrast to

the effect of the translation elongation inhibitors, we found that XUTs levels remain unchanged in stress conditions associated to global inhibition of translation initiation, suggesting a mechanism where the elongating ribosomes protect XUTs from the decay factors while they are translated. Ribo-Seq analyses confirmed that a substantial fraction of XUTs is actually bound by ribosomes, and identified actively translated smORFs in the 5' proximal portion of XUTs. Mechanistic analyses on a candidate XUT demonstrated that its NMD-sensitivity depends on the length of its 3' UTR. Finally, we show that a peptide can be produced from an NMD-sensitive lncRNA reporter in WT cells, suggesting that despite the 'cryptic' nature of the transcript, its translation can result into a detectable product.

Altogether, our data support a model where translation occupies a central role in the metabolism of cytoplasmic lncRNAs, a rapid binding by ribosomes probably being the default route as they reach the cytoplasm. We propose that these translation events allow lncRNA-derived peptides to be exposed to the natural selection, while NMD ensures that the transcripts they originate from are efficiently eliminated.

RESULTS

Translation determines the decay of cytoplasmic NMD-sensitive lncRNAs

The NMD-sensitivity of XUTs suggests that translation determines their decay. Thus, we anticipated that inhibiting translation would result in the accumulation of NMD-sensitive XUTs. To explore this idea, we treated exponentially growing WT cells with cycloheximide (CHX), a translation elongation inhibitor which binds the E site of the ribosome, preventing tRNA release and ribosome translocation (Garreau de Loubresse et al., 2014). Samples were collected at different time points after addition of the drug, then total RNA was extracted and analyzed by Northern blot. We observed that *XUT1678* and *XUT0741*, two NMD-sensitive XUTs that we previously characterized (Wery *et al.*, 2016), accumulate as soon as 5-10 min after CHX addition (Figure 1A). This effect is reversible, as the levels of both lncRNAs decreased after washing the CHX-treated cells and returning them to growth in fresh medium without CHX (Figure 1B). We noted that the 5' ITS1 fragment, a well-known physiological target of Xrn1 (Stevens et al., 1991), did not accumulate in CHX-treated WT cells (Figure 1A), indicating that CHX does not block the activity of Xrn1. In addition, we found that anisomycin (ANS), which also inhibits translation elongation but at a different stage than CHX (Figure S1A), led to a similar accumulation of *XUT1678* and *XUT0741* in WT cells (Figure 1C), reinforcing our hypothesis of a general translation-dependent lncRNAs decay process.

These data were extended at the genome-wide level using RNA-Seq, showing that the majority of NMD-sensitive XUTs significantly accumulate (fold-change >2, *P*-value <0.05) in WT cells treated with CHX or ANS (Figure 1D-E; see also Figure S1B and Table S1). In contrast, CHX and ANS only had a moderate effect on Cryptic Unstable Transcripts (CUTs), which are degraded in the nucleus by the Exosome (Neil et al., 2009; Wyers et al., 2005; Xu et al., 2009), indicating that translation primarily impacts cytoplasmic transcripts (Figure S1C).

The observation that NMD-sensitive XUTs rapidly accumulate in WT cells following inhibition of translation elongation is consistent with the idea that translation determines the degradation of cytoplasmic NMD-sensitive lncRNAs.

Translation can also affect lncRNAs decay independently of NMD

While NMD targets most XUTs, about 30% of them remain NMD-insensitive (Wery *et al.*, 2016). We asked whether these cytoplasmic transcripts that escape NMD also react to translation elongation inhibition.

Our RNA-Seq data showed that most NMD-insensitive XUTs significantly accumulate in CHX- and ANS-treated WT cells (Figure 2A-B, see also Figure S2A and Table S1). This indicates that translation can affect XUTs decay independently of NMD.

To further explore this idea, we performed RNA-Seq in *upf1Δ* cells, treated or not with CHX. This analysis revealed that NMD inactivation and CHX have a synergic effect on NMD-sensitive XUTs (Figure 2B), their global levels being significantly higher in the CHX-treated *upf1Δ* cells compared to the untreated *upf1Δ* cells ($P = 3.53e^{-100}$, Wilcoxon rank-sum test) or the CHX-treated WT cells ($P = 1.41e^{-27}$, Wilcoxon rank-sum test). Similar observations were made with ANS (see Figure S2A). Importantly, this synergy between NMD inactivation and CHX- or ANS-induced translation elongation inhibition was only observed for the NMD-sensitive XUTs, but not for the NMD-insensitive ones (Figure 2B, see also Figure S2A).

These observations raise the question of the mechanism by which translation could affect XUTs independently of NMD. In a previous work from the Parker's lab, CHX has been proposed to interfere with the decapping of the *MFA2* mRNA, leading to its stabilization (Beelman and Parker, 1994). This led us to assess whether this could also be the case for the CHX-sensitive XUTs.

To determine the capping status of the XUTs that accumulate upon CHX treatment, we performed RNA-Seq using the same RNA extracts as above, but including a treatment with the Terminator 5'-phosphate-dependent exonuclease, which degrades RNAs with 5'-monophosphate ends but not those with an intact m^7G cap (Figure 2C). This allowed us to show that 517 (35%) of the XUTs that accumulate in CHX-treated WT cells are Terminator-resistant, indicating that they accumulate as capped RNAs (Figure 2D; see also Table S1).

Decapping and NMD are functionally linked (He and Jacobson, 2015; Parker, 2012). According to the current models, the recruitment of the NMD core factors precedes the recruitment of the decapping machinery (Dehecq et al., 2018). As NMD depends on translation, one could imagine that NMD is less efficient in CHX-treated cells, which would in turn negatively impact the recruitment of the decapping factors. To explore how NMD inactivation affects the decapping of XUTs, we assessed the Terminator-sensitivity of XUTs in *upf1Δ* cells. Unexpectedly, we found that most NMD-sensitive XUTs that accumulate in the *upf1Δ* mutant are decapped, as shown by their global sensitivity to the Terminator exonuclease (Figure S2B-C). In fact, only 149 XUTs significantly accumulate in the *upf1* mutant (*upf1Δ*/WT ratio >2, *P*-value < 0.05) following Terminator digestion (Figure S2B; see also Table S1).

Thus, since most XUTs are efficiently decapped in the absence of Upf1, their accumulation in the NMD mutant is unlikely to reflect a decapping defect, but rather the disability of Xrn1 to access them. In addition, since the effect of CHX on the decapping of XUTs is more important than the effect of NMD inactivation, we conclude that the elongating ribosomes could directly interfere with the decapping of a fraction of XUTs, independently of NMD.

Cytoplasmic lncRNAs levels remain globally unchanged upon stress-induced inhibition of translation initiation

The data described above show that treating WT cells with CHX or ANS results into the accumulation of most XUTs. At the molecular level, these drugs act by freezing elongating ribosomes on their RNA substrates, a property which is widely exploited in Ribo-Seq analyses (Lareau et al., 2014; Wu et al., 2019).

Interestingly, mRNA degradation is known to occur co-translationally (Hu et al., 2009), and several reports have shown that the physical presence of ribosomes on an mRNA can interfere with its co-translational degradation by Xrn1 (Pelechano et al., 2015; Serdar et al., 2016). This led us to investigate whether the accumulation of XUTs observed in the presence of CHX or ANS could reflect a

protective effect of the ribosomes themselves, forming a physical obstacle that would block Xrn1 (Figure 3A). If correct, this model predicts that XUTs should not accumulate when ribosomes are not pre-loaded on the transcripts, *i.e.* in conditions where translation initiation is inhibited (Figure 3A).

In a recent study, the Tollervey's lab showed that the stress response induced by glucose starvation or heat-shock is associated to a global translational inhibition and rapid displacement of translation initiation factors (Bresson et al., 2020). We investigated how these stresses impact XUTs levels.

Firstly, we analyzed the effect of glucose deprivation by RNA-Seq using WT cells grown in glucose-containing medium and then shifted for 16 min in glycerol- and ethanol-containing medium (see Figure S3A-C). Strikingly, in contrast to the effect of translation elongation inhibition (CHX), we observed that XUTs globally do not accumulate upon glucose depletion (Figure 3B). In fact, only 61 were significantly up-regulated in the stress condition (fold-change >2, *P*-value < 0.05; see Table S1). Secondly, a re-analysis of published RNA-Seq data obtained in heat-shock conditions (Bresson *et al.*, 2020) showed that this stress does not lead to a global accumulation of XUTs neither (see Figure S3D and Table S1). Thus, the effects of glucose depletion and heat-shock are in sharp contrast with the effect of CHX (Figure 3B; see also Table S1).

Interestingly, we noted that the sensitivity of XUTs to CHX was significantly reduced following glucose depletion, though not totally (Figure 3C), suggesting that ribosome loading is indeed strongly reduced in this stress condition, though not fully abolished.

Altogether, these data suggest that the stabilization of XUTs observed in CHX-treated WT cells is mediated by the elongating ribosomes, which once bound on the XUTs, protect them from degradation.

Translational Landscape of yeast lncRNAs

A previous Ribo-Seq analysis in *upf1Δ* yeast cells revealed 47 smORFs on a 43 lncRNAs, providing a first proof-of-concept that lncRNAs can also be bound by ribosomes in *S. cerevisiae* (Smith

et al., 2014). However, it was not sufficient to get a robust and extensive identification of actively translated smORFs on 'cryptic' transcripts such as XUTs.

In order to define a more comprehensive translational landscape of yeast lncRNAs, we performed a new Ribo-Seq experiment in WT and *upf1Δ* cells. For each genetic background, we produced two datasets: one in native conditions (*i.e*. no treatment with translation inhibitor), a second using cells treated with CHX (Figure 4A).

As a first approach to analyze our Ribo-Seq data, we pooled them and searched for smORFs (≥ 5 codons, starting with an AUG codon) using the Ribotricer method, which directly assesses the 3-nt periodicity of Ribo-Seq data to identify actively translated ORFs (Choudhary et al., 2020). This led to the identification of 1560 translated smORFs on 748 XUTs (Figure 4A; see list 1 in Table S2). We then repeated the same procedure, but separating the conditions. This produced a refined list of 1270 smORFs from 633 XUTs, translated in at least one condition (Figure 4A-B; see list 2 in Table S2). Applying an additional coverage threshold (≥ 10 reads/smORF in at least one condition) restricted the list of 825 smORFs for 475 XUTs (Figure 4A; see list 3 in Table S2), which corresponds to the most robust candidates within the set of translated smORFs/XUTs, showing the highest levels of translation and being translated in at least one condition. However, since the translation of lncRNAs could also be transient and occur at low levels, we decided to use the second list of 633 translated XUTs as a compromise for the descriptive analysis below. Figure 4C shows a metagene view of the Ribo-Seq signals for these XUTs in the four conditions. A similar metagene analysis for the other XUTs (not detected as translated) revealed that the signals are globally lower, suggesting that our analysis captured the XUTs that display the highest levels of translation (Figure S4A).

First of all, 510 and 123 of these 633 XUTs are NMD-sensitive and NMD-insensitive, respectively (Figure 4D; see also Table S2). Notably, 297 of these 633 XUTs are detected as translated in native condition, essentially in the *upf1Δ* mutant (Figure 4B). As one could expect, combining NMD inactivation and CHX treatment strongly increases the number of XUTs identified as translated

(502/633, including 118 XUTs detected only in CHX-treated *upf1Δ* cells; see Figure 4B). Cumulatively, 411 XUTs were detected as translated in at least two datasets (Figure 4B).

The smORFs detected on XUTs display a median size of 87 nt (Figure S4B). We noted that for half of the XUTs (311/633), ribotricer detected more than one smORF per transcript (Figure S4C). This could reflect the potential of several smORFs on a same XUT to attract the translation machinery, and/or the existence of distinct isoforms for a same XUT, displaying different boundaries and possibly encompassing different smORFs. Interestingly, for 75% of the translated XUTs, the smORF showing the highest Ribo-Seq signal corresponds to one of the three first smORFs predicted in the XUTs sequence (Figure S4D). This is in line with the metagene analysis showing that ribosomes preferentially bind the 5'-proximal region of the translated XUTs (Figure 4C). Finally, we observed that the size of the 3' UTR is significantly higher for the NMD-sensitive XUTs than for the NMD-insensitive ones (median = 733 nt *vs* 236 nt; $P = 1.63e^{-26}$, Wilcoxon rank-sum test; see Figure 4E), suggesting that as for mRNAs, the length of the 3' UTR is a critical determinant for degradation by NMD (Celik *et al.*, 2017; Muhlrad and Parker, 1999).

Together, these data show that a substantial fraction of XUTs carry smORFs that are actively translated, and that the NMD-sensitive XUTs display a longer 3' UTR than the XUTs which escape NMD.

The NMD-sensitivity of *XUT0741* depends on its long 3'-UTR

The observation that the 3' UTR is significantly longer for the NMD-sensitive XUTs compared to the NMD-insensitive ones suggests that it might also constitute a key determinant of the NMD-sensitivity for XUTs, as for mRNAs (Celik et al., 2017; Muhlrad and Parker, 1999). We therefore investigated this hypothesis, using the NMD-sensitive *XUT0741* as a model candidate.

XUT0741 belongs to the top list of translated XUTs, with a single 5'-proximal smORF (15 codons) detected by each of our different analyses (see lists 1-3 in Table S2; see also Figure S5A). This smORF is followed by a 1.3 kb long 3' UTR, with multiple stop codons in the same frame (Figure 5A). To explore the role of the 3' UTR as a *cis* element determining its NMD-sensitivity, we designed six

mutants of *XUT0741* by mutating several of these in-frame stop codons, to progressively lengthen the smORF and consequently shorten the 3' UTR (Figure 5A; see sequences in Supplementary File 1). These mutant alleles were integrated at the genomic locus in WT and *upf1Δ* strains. Their expression and NMD-sensitivity were then assessed by strand-specific RT-qPCR.

Our data show that the abundance of the XUT in WT cells and its NMD-sensitivity remain unchanged in the three first mutants (Figure 5B; see also Figure S5B). However, as the 3' UTR is shortened to 298 nt in the *xut0741-d* mutant (which is in the range of 3' UTR size for NMD-insensitive XUTs; see Figure 4E), we observed a significant accumulation of the mutated transcript, correlating with a significant decrease of its sensitivity to NMD (Figure 5B; see also Figure S5B). Further shortening the 3' UTR in mutants *–e* and *–f* did not aggravate these effects (Figure 5B). Note that the mutations introduced in *XUT0741* do not affect the NMD-sensitivity of another XUT (Figure S5B).

Thus, changing the length of the coding region relative to the 3' UTR not only modifies the abundance of *XUT0741* in WT cells, but also its NMD-sensitivity. To discriminate whether the later depends on the length of the ORF or of the 3' UTR, we constructed a chimera combining the extended ORF of 'NMD-resistant' *xut0741-d* to the long 3' UTR of the native *XUT0741* (Figure 5C). The fate of this chimera was then analyzed by Northern blot. As expected, the corresponding RNA was longer than the native XUT (Figure 5D). Notably, if the chimera was detected in WT cells, its levels increased by 3-fold in the *upf1Δ* context (Figure 5D), indicating that it is NMD-sensitive. More precisely, the quantifications we performed from four independent experiments demonstrated that the chimera displays the same NMD-sensitivity as the native XUT (Figure 5E). We therefore conclude that the NMD-sensitivity of the XUT is determined by its long 3' UTR.

Translation of an NMD-sensitive lncRNA produces a peptide in NMD-competent WT cells

All the observations described above contribute to highlight that translation occupies a critical place in the metabolism of cytoplasmic lncRNAs. This led us to ask whether peptides could be produced as these lncRNAs are targeted to NMD, possibly during a single (so-called pioneer) round of translation

(note that for simplicity, we will systematically use the term 'peptide' to refer to the product of the translation of a lncRNA, regardless its size).

Conceptually, the fact that NMD is triggered as translation terminates makes it possible for a peptide to be produced and released. To explore whether this could occur with yeast NMD-sensitive lncRNAs, we took advantage of the *xut0741-b* mutant (Figure 5A), which displays the same NMD-sensitivity as the native XUT (Figure 5B), but encodes a larger peptide easier to detect by Western blot. We decided to use this mutant as an NMD-sensitive lncRNA reporter, following the insertion a C-terminal 3FLAG tag (Figure 6A; see sequence in Supplementary File 1). We controlled that the insertion of the 3FLAG tag does not affect the NMD-sensitivity of the transcript (Figure 6B). Importantly, despite the very low abundance of the transcript in WT cells, at the protein level we observed by Western blot a clear band at the expected size, demonstrating that the encoded peptide is produced (Figure 6C, lane 3), with an increase in the *upf1Δ* context (Figure 6C, lane 4). These results provide the proof-of-principle evidence that a peptide can be produced from an NMD-sensitive transcript in WT yeast cells.

To gain further insight into the relationship between translation and NMD-sensitivity of XUTs, we designed a construct where the translation of our NMD-sensitive lncRNA reporter is blocked in *cis*, using a short stem-loop (SL) element, previously shown to inhibit translation initiation (Beelman and Parker, 1994; Muhlrad et al., 1995). This SL was inserted into our reporter, 11 nt upstream from the translation start site (Figure 6A; see sequence in Supplementary File 1). A Western blot showed that the production of the peptide is completely lost upon SL insertion, indicating that the transcript is not translated anymore in this context (Figure 6C, lanes 5-6). Notably, at the RNA level, this loss of translation correlates with a dramatic and significant reduction of the sensitivity of the XUT to both CHX and NMD (Figure 6D). In contrast, the sensitivity to CHX and NMD of other NMD-sensitive XUTs used as controls remained unaffected (Figure S6A-B).

In conclusion, our data show that translation of an NMD-sensitive lncRNA can give rise to a peptide, as the transcript is efficiently targeted to NMD in WT cells. Furthermore, our mechanistic analysis confirms that the CHX- and NMD-sensitivity of XUTs reflects an active translation process.

DISCUSSION

Since their discovery, lncRNAs have been considered as transcripts devoid of coding potential, escaping translation. However, accumulating experimental evidence lead us to re-evaluate this assumption. In fact, lncRNAs co-purify with polysomes in different models, including yeast (Smith *et al.*, 2014) and human cells (Carlevaro-Fita *et al.*, 2016; Douka et al., 2021; van Heesch *et al.*, 2014). In addition, high-throughput sequencing of ribosome-bound fragments using Ribo-Seq or related approaches has uncovered smORFs within lncRNAs (Aspden *et al.*, 2014; Chen *et al.*, 2020; Douka *et al.*, 2021; Ingolia *et al.*, 2014; Ingolia *et al.*, 2011). Finally, several studies reported the identification of peptides resulting from the translation of smORFs carried on lncRNAs (Chen *et al.*, 2020; D'Lima *et al.*, 2017; Douka *et al.*, 2021; Matsumoto *et al.*, 2017; Slavoff *et al.*, 2013; Zanet *et al.*, 2015).

In yeast, lncRNAs expression is restricted by the extensive action of nuclear and cytoplasmic RNA decay machineries (Tisseur et al., 2011), including the 5'-exoribonuclease Xrn1 which degrades a class of cytoplasmic lncRNAs referred to as XUTs (Van Dijk *et al.*, 2011). We and others previously reported that most of them are targeted by the translation-dependent NMD pathway, suggesting that XUTs are translated and that translation controls their degradation (Malabat *et al.*, 2015; Wery *et al.*, 2016).

Here, we report several observations supporting this hypothesis. We showed that the majority of XUTs accumulate in WT cells treated with CHX or ANS (Figure 1), two drugs known to inhibit translation elongation but *via* different modes of action (Garreau de Loubresse *et al.*, 2014). Using Ribo-Seq, we showed that a substantial fraction of XUTs are actually bound by ribosomes, and we identified actively translated smORFs which are mainly found in the 5'-proximal region of XUTs (Figure 4). Mechanistic analyses at the level of a candidate XUT showed that its sensitivity to NMD is determined by the length of the 3' UTR downstream of the translated smORF (Figure 5). Finally, we showed that a detectable peptide is produced from an NMD-sensitive lncRNA reporter in WT cells, as the transcript is targeted to NMD (Figure 6).

The fact that NMD-sensitive XUTs accumulate in the presence of a translation elongation inhibitor reinforces our model of a translation-dependent decay process. However, the underlying molecular mechanism appear to be more complex than anticipated, as the accumulation of XUTs observed upon CHX/ANS treatment cannot be solely explained by the inability of the cell to trigger NMD when translation is inhibited. Firstly, NMD-insensitive XUTs (which account for 30% of XUTs) also accumulate in presence of CHX or ANS (Figure 2A-B). Secondly, stress conditions associated to global translation initiation inhibition do not recapitulate the stabilization effect of the translation elongation inhibitors on XUTs (Figure 3B-C; see also Figure S3D). Thirdly, blocking elongating ribosomes with CHX interferes with the decapping of 35% of XUTs, which accumulate as capped RNAs in CHX-treated cells (Figure 2D), while most of the XUTs that accumulate upon NMD inactivation are decapped (Figure S2B-C). Together, these observations lead us to propose that while XUTs are translated, they would be protected by the elongating ribosomes sterically blocking the decapping enzyme Dcp2 and/or Xrn1, independently of NMD. This model extends beyond mRNAs the idea that translating ribosomes can protect any transcripts from the degradation (Bresson *et al.*, 2020; Chan et al., 2018).

Yet, several points remain to be clarified. Among them, the observation that the decapping of a fraction of XUTs is affected upon translation elongation inhibition raises the question of the difference between the XUTs that accumulate as capped or decapped in CHX-treated cells. By analogy with the model described above, we propose that ribosomes could sterically block Dcp2 when the translated smORF is close from the transcript start site (TSS). Additional mechanistic analyses are required to validate this hypothesis. Nonetheless, the fact that 65% of XUTs are efficiently decapped in CHX-treated cells rules out the idea that CHX would act as a global inhibitor of decapping.

The observation that most XUTs accumulate as decapped RNAs in *upf1Δ* cells was unexpected. On one hand, this shows that decapping remains efficient in the absence of NMD. On the other hand, how to explain that XUTs accumulate and escape Xrn1 in this context, if they are decapped? Again, we could envisage a ribosome-mediated protection, in this case involving the terminating ribosome which would sterically block Xrn1, but not Dcp2 (unless for very short, TSS-proximal smORFs, for which the

terminating ribosome would remain close enough from the TSS to interfere with the decapping machinery). In this regard, it has been shown that the ATPase activity of Upf1 is required for efficient ribosome release at the level of the stop codon; consequently, the inability to remove the terminating ribosome when this activity is lost impedes mRNA degradation by Xrn1, leading to the accumulation of 3' mRNA decay fragments (Serdar *et al.*, 2016). Future work should decipher whether the stabilization of XUTs observed in absence of functional NMD involves a similar molecular mechanism.

Our Ribo-Seq analysis allowed us to identify actively translated smORFs for 38% of annotated XUTs, including 510 NMD-sensitive XUTs and 123 NMD-insensitive XUTs (Figure 4D), considerably extending the repertoire of translated lncRNAs in yeast (Smith *et al.*, 2014; Wery *et al.*, 2016). These data point out that NMD insensitivity does not imply lack of translation, and that the translational landscape of yeast lncRNAs extends beyond the scope of NMD. This is consistent with the observation that translation elongation inhibition also impacts the decay of most NMD-insensitive XUTs (Figure 2).

The number of smORFs/XUTs detected as translated depends on the stringency of the approach used to analyze the Ribo-Seq signals (Figure 4A), which is in line with the idea that lncRNAs translation is transient and therefore more difficult to detect in comparison to canonical mRNAs translation (Wacholder et al., 2021). Besides the global low abundance of XUTs even in conditions where they are stabilized (NMD inactivation, CHX treatment), we imagine that the translation of many of their smORFs remains labile, probably reflecting the fact that they are rapidly and continuously evolving. Furthermore, perhaps some constraints associated to canonical mRNA translation could be relaxed in the context of lncRNAs translation to maximize the range of possibilities when exploring the potential of genetic novelty, which would be interesting from an evolutionary point of view. But the corollary is therefore a difficulty for us to detect such non-canonical translation events using pipelines that use the marks of canonical translation (*e.g*. use of an AUG initiator codon, predominance of one phase *vs* the two others). The field is therefore in need of dedicated approaches and computational tools to reveal the exhaustive landscape of lncRNAs translation.

Together with the observation that the NMD-sensitive XUTs display a longer 3' UTR than the NMD-insensitive ones, the mechanistic analysis on the *XUT0741* candidate highlights the critical role of the 3' UTR in determining the NMD-sensitivity of XUTs, as for mRNAs (Celik *et al.*, 2017; Muhlrad and Parker, 1999). However, even in the last mutant of *XUT0741* (where the 3' UTR is shortened to 91 nt), the NMD-sensitivity is not fully abolished (Figure 5B; see also Figure S5B). One possibility to explain that is the existence of an alternative smORF, unaffected in our mutants. Supporting this hypothesis, we observed low Ribo-seq signals upstream from the detected smORF, overlapping the annotated TSS of *XUT0741* (Figure S5A). Interestingly, *XUT0741* TSS corresponds to the 'G' of an 'ATG' triplet, followed by 14 codons before the first in-frame stop codon (see sequences in Supplemental File 1). The production of multiple RNA isoforms production from the same transcription unit is common in yeast (Pelechano et al., 2013), and we can imagine that any 5'-extended isoforms of *XUT0741* would encompass this ATG and therefore carry this alternative smORF. Additional mechanistic analyses combined to RNA isoforms profiling would be required to confirm this hypothesis. Nonetheless, the complexity of the yeast transcriptome, with the existence of multiple RNA isoforms displaying different boundaries, might possibly explain the detection of several smORFs per XUT and should be kept in mind when investigating how the position of smORFs relative to its annotated extremities can impact the fate of a XUT.

One important finding of our work is that the translation of an NMD-sensitive lncRNA reporter gives rise to a peptide detectable in a WT context, where NMD is functional. From a conceptual point of view, the idea that a peptide can be produced from an NMD substrate is plausible, since NMD is activated as translation terminates at the level of a 'normal' stop codon (this is the position of this codon within the transcript which is sensed as 'abnormal'). However, the fate of this peptide has not been characterized in detail so far and remains largely obscure. On one side, a study in yeast proposed that Upf1 stimulates the proteasome-dependent degradation of the truncated translation product derived from an NMD-sensitive mRNA carrying nonsense mutation (Kuroha et al., 2009), consistent with the classical view that such products might be deleterious for the cell and should be eliminated.

On the other side, a study in mammalian cells revealed that the pioneer round of the translation which targets mRNAs with premature stop codons to NMD can in the same time produce antigenic peptides for the MHC class I pathway (Apcher et al., 2011). In this context, the observation we made here using our tagged NMD-sensitive reporter provides the proof-of-principle that translation of an NMD-sensitive transcript can give rise to a peptide which can exist into the cell, even if the transcript it originates from is targeted to the degradation. This first observation paves the way towards the future characterization of the yeast peptidome, searching for native peptides derived from the translation of XUTs.

Overall, our data lead us to propose that translation of 5'-proximal smORFs is a general feature of the cytoplasmic lncRNAs, modulating their cellular abundance (Figure 6E). While they are translated, the presence of elongating ribosomes would protect them from the decay factors. Then, as translation of these smORFs terminates far away from the poly(A) tail, the NMD factors would be recruited by the eukaryotic release factors (eRF1/eRF3) to the terminating ribosome, triggering NMD. In the same time, the peptides that have been produced would be exposed to the natural selection, possibly contributing to the emergence of *de novo* protein-coding genes.

De novo gene birth has been associated with adaptation to environmental stress (Arendsee et al., 2014), and NMD is known to be repressed under a variety of stress conditions (Gardner, 2008; Mendell et al., 2004). It is therefore tempting to speculate that despite the cell has evolved efficient pathways to degrade lncRNAs and restrict their expression, these pathways can be down-regulated under some specific conditions (*e.g*. stress) to sample the peptide potential hosted in these lncRNAs.

An important corollary of our model is that lncRNA-derived peptides are unlikely to be functional yet. Consequently, their loss is not expected to confer a phenotype. However, their overexpression might confer a selective advantage. This thought highlights the importance of addressing the question of the functionality of lncRNAs by considering approaches based on gain-of-function (Rodriguez-Lopez et al., 2022; Vakirlis et al., 2020).

In conclusion, our work contributes to point out that translation plays a major role in the post-transcriptional metabolism of cytoplasmic lncRNAs, and that their definition as 'non-coding' is probably not appropriate to describe their actual status. Rather, they might be viewed as transcripts oscillating between the 'coding' and 'non-coding' worlds, assessing the potential of genetic novelty *via* the production of novel peptide, which if beneficial for the cell, might be selected to give rise to novel protein-coding genes.

EXPERIMENTAL PROCEDURES

Yeast strains and media

The strains used in this study are listed in Table S3. Mutants were constructed by transformation and were all verified by PCR on genomic DNA (see above).

Yeast cells were grown to mid-log phase (OD_{600} 0.5) at 30°C in Yeast Extract-Peptone-Dextrose (YPD) medium or Complete Synthetic Medium (CSM), with 2% glucose. In the glucose starvation experiments, glucose was replaced glycerol and ethanol.

5-Fluoroorotic acid (5-FOA) was used at a final concentration of 1 g/L on solid CSM plates. G418 (Geneticin; Gibco) was used at a final concentration of 100 µg/ml on solid YPD plates. CHX (Sigma) and ANS (Sigma) were used at a final concentration of 100 µg/ml.

Construction of xut0741 mutants

The *xut0741-a*, *-b*, *-d* and *-f* alleles, flanked by NaeI sites, were produced as synthetic gBlocks DNA fragments (IDT – Integrated DNA Technologies), and then cloned between the KpnI and XbaI sites of the pAM376 backbone vector (Szachnowski et al., 2019), giving the pAM594, pAM596, pAM598 and pAM600 vectors, respectively. The *xut0741-c* and *xut0741-e* mutants were constructed by site-directed mutagenesis from *xut0741-d* and then cloned into the same backbone vector, giving the pAM724 and pAM723 vectors, respectively. The sequence of each alleles was verified by Sanger sequencing and is available in Supplemental File 1. The mutant alleles were excised from the pCRII vector using NaeI digestion and transformed into the YAM2831 (where the *XUT0741*/*ADH2* locus has been deleted by *URA3*). After 1 day of growth on non-selective medium, transformants were replicated on CSM + 5-FOA plates and incubated at 30°C for 4-5 days. The proper integration of the mutant alleles was confirmed by PCR on genomic DNA using oligonucleotide AMO3350-3351. *UPF1* was deleted subsequently by transformation with the product of a PCR on YAM202 (*upf1Δ::kanMX6*) genomic DNA with oligonucleotides AMO2710-2711. The transformants were selected on YPD + G418 plates and *UPF1* deletion was verified by PCR on genomic DNA using oligonucleotides AMO190-2712.

The chimera-encoding plasmid (pAM726) was produced in two steps. Firstly, the 3'-UTR of the native *XUT0741* was amplified by PCR on YAM1 genomic DNA using oligonucleotides AMO3471-3382, and then cloned between the KpnI and XbaI sites of a pCRII-TOPO backbone, giving the pAM725 vector. Secondly, the sequence corresponding to the 5'-UTR and ORF of the *xut0741-d* mutant was amplified by PCR on YAM2854 genomic DNA using oligonucleotides AMO3379-3497, and then cloned between the KpnI and EcoRI sites of pAM725, giving the pAM726 vector. The sequence of the chimera allele was verified by Sanger sequencing (see Supplemental File 1). Plasmid digestion, transformation in YAM2831 cells, transformants selection and screening, as well as *UPF1* deletion, were as described above.

C-terminal 3FLAG tagging of *xut0741-b* was performed using an 'overlap extension PCR' strategy. A first amplicon was produced by PCR on YAM2853 genomic DNA using oligonucleotides AMO3379-3530. A second amplicon was produced by PCR on the same DNA using oligonucleotides AMO3382-3531. After purification on agarose gel, the two amplicons (displaying a 28-bp overlap) were mixed and used as DNA templates for PCR using oligonucleotides AMO3379-3382. The final full PCR product was then digested by KpnI and XbaI and cloned in the same backbone vector as the other *xut0741* mutants, giving the pAM728 plasmid (see Supplemental File 1 for insert sequence). All subsequent steps were as above.

The stem-loop (GATCCCGCGGTTCGCCGCGG), previously shown to inhibit *MFA2* mRNA translation (Beelman and Parker, 1994), was inserted into the 3FLAG-tagged *xut0741-d* allele using a similar 'overlap extension PCR' strategy, involving the overlapping oligonucleotides AMO3550 (for the 5' amplicon) and AMO3549 (for the 3' amplicon), ultimately giving the pAM741 plasmid (insert sequence available in Supplemental File 1). All subsequent steps were as above.

Total RNA extraction

Total RNA was extracted from exponentially growing cells (OD_{600} 0.5) using standard hot phenol procedure. Extracted RNA was ethanol-precipitated, resuspended in nuclease-free H_2O

(Ambion) and quantified using a NanoDrop 2000c spectrophotometer and/or a Qubit fluorometer with the Qubit RNA HS Assay kit (Life Technologies).

Northern blot

10 µg of total RNA were separated on denaturing 1.2% agarose gel and then transferred to Hybond-XL nylon membrane (GE Healthcare). ^{32}P-labelled oligonucleotides (listed in Table S4) were hybridized overnight at 42°C in ULTRAhyb®-Oligo hybridization buffer (Ambion). After hybridization, membranes were washed twice in 2X SSC/0.1% SDS for 15 minutes at 25°C, and once in 0.1X SSC/0.1% SDS for 15 minutes at 25°C. Membranes were exposed to Storage Phosphor screens. Signal was detected using a Typhoon Trio PhosphorImager and analyzed with the version 10.1 of the ImageQuant TL sofware (Cytiva).

Strand-specific RT-qPCR

Strand-specific RT-qPCR experiments were performed from three biological replicates, as previously described (Wery et al., 2018a). The oligonucleotides used are listed in Table S4.

Total RNA-Seq

For each strain/condition, total RNA-Seq was performed from two biological replicates. For each sample, 1 µg of total RNA was mixed with 2 µl of diluted ERCC RNA spike-in mix (1:100 dilution in nuclease-free H_2O; Invitrogen). Ribosomal (r)RNAs were depleted using the Ribominus Eukaryote v2 kit (Ambion). Alternatively, 1.5 µg of total RNA was mixed with 3 µl of diluted ERCC RNA spike-in mix and then digested for 1h at 30°C with 1 unit of Terminator 5´-Phosphate-Dependent Exonuclease (Epicentre) in 1X Reaction Buffer A containing 10 units of SUPERase-In RNase inhibitor (Invitrogen). After phenol/chloroform extraction, Terminator-digested RNA was precipitated with ethanol, and then resuspended in nuclease-free H_2O.

Libraries were prepared from the rRNA-depleted or Terminator-digested RNAs using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina) and the IDT for Illumina – TruSeq RNA UD indexes (Illumina). Paired-end sequencing (2 x 50 nt) was performed on a NovaSeq 6000 system (Illumina).

Total-Seq data processing and analysis

Reads were trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore) and mapped on the S288C reference genome (R64-2-1, including the 2-micron plasmid), with addition of either ERCC RNA spike-in sequences or the *Schizosaccharomyces pombe* genome (ASM294v2, for the heat-shock dataset) using version 2.2.0 of Hisat (Kim et al., 2019), with default parameters and a maximum size for introns of 5000. All subsequent analyses used uniquely mapped reads.

Gene counts were computed using version 2.0.0 of featureCounts (Liao et al., 2014), and then normalized using the estimateSizefactorsForMatrix function from the DESeq2 package (Love et al., 2014). Tag densities were obtained as: normalized gene count/gene length.

For all the RNA-Seq data produced in this study, normalization on the ERCC RNA spike-in signal was used in a first time to control that snoRNAs expression is not affected in the mutant/condition analyzed, and snoRNA counts were then used for normalization, as previously described (Wery *et al.*, 2016; Wery et al., 2018b).

For the heat-shock dataset (retrieved from the NCBI Gene Expression Omnibus using accession number GSE148166), gene counts were normalized on the *S. pombe* spike-in RNA, as snoRNAs levels were abnormally low in both stressed and control cells, probably due to differences in the library preparation protocol (Bresson *et al.*, 2020).

Differential expression analyses were performed with DESeq2 (Love *et al.*, 2014).

Ribo-Seq libraries preparation

Ribo-Seq analysis was performed from two biological replicates YAM1 (WT) and YAM202 (*upf1Δ*) cells, grown to mid-log phase (OD_{600} 0,5) at 30°C in YPD, then treated or not for 15 minutes with CHX (100 µg/ml, final concentration). For each sample, 250 ml of cells were harvested by centrifugation at room temperature and directly frozen in liquid nitrogen after supernatant removal.

Cells were lysed in 1X lysis buffer (10 mM Tris-HCl pH 7.4, 100 mM NaCl, 30 mM $MgCl_2$) supplemented by 2X cOmplete Protease Inhibitor Cocktail (Roche) and ribosome protected fragments (RPFs) were prepared as previously described (Baudin-Baillieu et al., 2016), with the following modifications. Polysomes were purified on sucrose cushion then digested with RNase I (Ambion, 5 units/UA_{260}). Biotinylated oligonucleotides (IDT - Integrated DNA Technologies) used for ribo-depletion are listed in Table S4.

Libraries were then prepared from 10 ng of RPFs using the D-Plex Small RNA-Seq kit for Illumina (Diagenode) and the D-Plex Unique Dual Indexes for Illumina – set A (Diagenode). The RPFs were diluted in a final volume of 8 µl before the addition of 2 µl of Dephoshorylation Buffer, 5 µl of Crowding Buffer and 0.5 µl of Dephosphorylation Reagent. The samples were incubated for 15 minutes at 37°C. RNA tailing was performed by adding 1.5 µl of Small Tailing Master Mix (1 µl of Small Tailing Buffer + 0.5 µl of Small Tailing Reagent) to the dephosphorylated RNAs, and incubating the samples for 40 minutes at 37°C. The samples were transferred on ice for 2 minutes before the addition of 1 µl of the Reverse Transcription Primer (RTPH). The samples were denaturated for 10 minutes at 70°C and then cooled down to 25°C at a 0.5°C/sec rate. A Reverse Transcription Master Mix (RTMM) was prepared by mixing 5 µl of Reverse Transcription Buffer and 1 µl of Reverse Transcription Reagent; 6 µl of this mix were added to the samples, which were then incubated for 15 minutes at 25°C. After adding 2 µl of Small Template Switch Oligo, the samples were incubated for 120 minutes at 42°C, then heated for 10 minutes at 70°C and finally kept at 4°C. For the PCR amplification, 20 µl of D-Plex Primer UDI and 50 µl of PCR Master Mix were added, then the following program was run: initial denaturation at 98°C for 30 seconds; 10 cycles including 15 seconds at 98°C followed by 1 minute at 72°C; final incubation of 10 minutes at 72°C; hold at 4°C. The libraries were then purified using the Monarch PCR & DNA Cleanup

Kit (NEB), using a 5:1 ratio of Binding Buffer: Sample. Purified DNA was eluted in 50 µl of nuclease-free H_2O (Ambion). A second cleanup of the libraries was performed using 1 volume of AMPure XB beads (Beckman). Libraries were eluted in 20 µl of nuclease-free H_2O (Ambion), and then quantified using the Qubit dsDNA HS assay (Invitrogen). Finally, the size and the molarity of each library were determined using a High Sensitivity D1000 ScreenTape in a 4200 TapeStation (Agilent Technologies).

Single-end sequencing (50 nt) of the libraries was performed on a NovaSeq 6000 system (Illumina).

Detection of translated XUTs/smORFs using Ribotricer

Unique molecular identifiers (UMI) were extracted using umi_tools (Smith et al., 2017), and then used to discard PCR duplicates. Reads were trimmed using cutadapt v2.10 (Martin, 2011), and then mapped using Hisat v2.0.0 (Kim *et al.*, 2019), as above. Reads mapping on rRNA were discarded. Subsequent analyses only used uniquely mapped reads with a size comprised between 25 and 35 nt.

Ribotricer 1.3.1 was used to extract translated ORFs (minimum length of 15 nt) based on *S. cerevisiae* genome annotation (including XUTs), using ATG as the start codon and a phase-score cutoff of 0.318, as recommended by the authors (Choudhary *et al.*, 2020). The phasing of Ribo-Seq data was also controlled independently (see Figure S7). List 1 of translated XUTs was obtained after pooling the bam files from all conditions. List 2 was obtained by analyzing each condition separately, polling the bam files from the two biological replicates. List 3 was obtained from list 2, upon application of a coverage filter (at least 10 reads per translated smORF).

Protein extraction and Western blot

Protein extracts were prepared from exponentially growing cells, using a standard method based on cell lysis with glass beads in 'IP' buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 20% glycerol), supplemented with 0.05% NP40, 0.5X cOmplete Protease Inhibitor Cocktail (Roche) and 1 mM AEBSF.

40 µg of total extracts were separated on NuPAGE 4-12% Bis-Tris gel (Invitrogen) in 1X NuPAGE MOPS SDS running buffer (Invitrogen), and then transferred on a nitrocellulose membrane using iBlot 2 Transfer Stack system (Invitrogen), with program '0'.

The FLAG-tagged peptide and Pgk1 were detected using the anti-FLAG M2 (Sigma) and anti-Pgk1 22C5D8 (abcam) monoclonal antibodies, revealed using the SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) and the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific), respectively, with a ChemiDoc Imaging System (BioRad).

DATA ACCESSIBILITY

Raw sequences generated in this work have been deposited to the NCBI Gene Expression Omnibus and can be accessed using accession number GSE203283. Genome browsers for visualization of processed data will be publicly accessible as soon as the manuscript will be accepted for publication.

ACKNOWLEDGEMENTS

We would like to thank Aaron Wacholder and Anne-Ruxandra Carvunis for sharing unpublished results and fruitful discussions. We are also grateful to all members of our labs for discussions. High-throughput sequencing was performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the ITMO-Cancer Aviesan (Plan Cancer III) and by the SiRIC-Curie program (SiRIC Grant INCa-DGOS-465 and INCa-DGOSInserm_12554). Data management, quality control and primary analysis were performed by the Bioinformatics platform of the Institut Curie. This work has benefited from the ANR "DNA-life" (ANR-15-CE12-0007) grant and the ERC "DARK" consolidator grant allocated to Antonin Morillon.

AUTHOR CONTRIBUTIONS

M.W., I.H., O.N. & A.M. designed experiments. S.A., N.V., I.H. & M.W. performed experiments. U.S., C.P. & A.L. performed bioinformatics analyses. S.A., U.S., I.H., C.P., A.L. & M.W. analyzed the data. M.W. & A.M. designed the project. M.W. & A.M. supervised the project. M.W. wrote the manuscript, with input from all authors. A.M. acquired funding.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

FIGURE LEGENDS

Figure 1. NMD-sensitive lncRNAs accumulate upon translation inhibition.

A. WT (YAM1) cells were grown to mid-log phase in rich (YPD) medium at 30°C. CHX was then added at a final concentration of 100 µg/ml, and samples were collected at different time points. Untreated *xrn1Δ* (YAM6) and *upf1Δ* (YAM202) cells, grown under the same conditions, were used as controls. Total RNA was extracted and analyzed by Northern blot. *XUT1678* (and the overlapping *SUT768*), *XUT0741*, the 5' ITS1 fragment (as well as the 20S rRNA precursor it derives from) and *scR1* (loading control) were detected using ^{32}P-labelled AMO1595, AMO1762, AMO496 and AMO1482 oligonucleotides, respectively.

B. WT (YAM1), *xrn1Δ* (YAM6) and *upf1Δ* (YAM202) cells were grown as above. CHX was then added to the WT cells for 15 minutes (100 µg/ml, final concentration). The CHX-treated cells were then washed with fresh pre-heated YPD medium and re-incubated at 30°C. Samples of washed cells were collected after 15, 30, 60 and 120 minutes. Total RNA was extracted and analyzed by Northern blot as described above.

C. Same as Figure 1A using ANS (100 µg/ml final concentration) instead of CHX.

D. Total RNA-Seq was performed using total RNA extracted from exponentially growing WT (YAM1) cells (grown as above) treated for 15 minutes with CHX (100 µg/ml, final concentration) or with an equal volume of DMSO (control). The scatter plot shows the RNA-Seq signals (tag densities, $\log_2$ scale) for the NMD-sensitive XUTs, mRNAs (light grey dots) and snoRNAs (black dots) in CHX-treated and control WT cells. The significantly up-regulated (CHX/control fold-change >2, *P*-value <0.05) and unaffected NMD-sensitive XUTs are represented as red and dark grey dots, respectively.

E. Venn diagram showing the number of NMD-sensitive XUTs that accumulate in CHX- and/or ANS-treated WT cells.

Figure 2. Translation can also impact XUTs independently of NMD.

A. Total RNA-Seq was performed in WT (YAM1) and *upf1Δ* (YAM202) cells, with or without treatment with CHX (15 minutes, 100 µg/ml final concentration) or ANS (30 minutes, 100 µg/ml final concentration). Densities were computed for NMD-sensitive and NMD-insensitive XUTs. The sensitivity to NMD and/or CHX/ANS of each transcript is shown as an heatmap of the fold-change (log_2 scale) relative to the corresponding control WT cells (treated for the same time with an equal volume of DMSO).

B. Same as above. The data are presented as densities (tag/nt, log_2 scale) for NMD-sensitive and NMD-insensitive XUTs in control (DMSO) or CHX-treated WT (YAM1) and *upf1Δ* (YAM202) cells. *** *P*-value < 0.001; ns, not significant upon two-sided Wilcoxon rank-sum test (adjusted for multiple testing with the Benjamini–Hochberg procedure).

C. Schematic representation of the action of the Terminator 5'-phosphate-dependent exonuclease, which degrades RNAs that are decapped (grey), but not those with an intact m^7G cap (red).

D. Total RNA-Seq was performed using the same RNA extracts as in Figure 1D, including a treatment with the Terminator 5'-phosphate-dependent exonuclease before the preparation of the libraries. The data are presented as in Figure 1D, the red dots representing the 517 CHX-sensitive XUTs that are still detected as significantly up-regulated in CHX-treated WT cells (CHX/control fold-change >2, *P*-value <0.05) upon Terminator treatment. The other XUTs (Terminator-sensitive) are represented as dark grey dots.

Figure 3. XUTs do not globally accumulate in stress conditions associated to translation initiation inhibition.

A. Schematic interpretation of the effect of CHX-mediated inhibition of translation elongation (left) and of stress-induced inhibition of translation initiation (glucose starvation, right) on Xrn1-dependent degradation of XUTs (red). The red arrow on the XUT represents a smORF.

B. Total RNA-Seq was performed in WT (YAM1), *xrn1Δ* (YAM6) and *upf1Δ* (YAM202) grown in CSM. WT cells grown in the same conditions and then submitted to a CHX treatment or glucose starvation (-Glu)

were also included. Densities (tag/nt) were computed for the 1321 XUTs significantly up-regulated in the *xrn1* mutant grown in this CSM (see Figure S3B), which were then separated according to their sensitivity to NMD (see Figure S3C). The sensitivity of each of these XUTs to CHX and glucose starvation is presented as an heatmap of the fold-change (log_2 scale). As an indication, the sensitivity of these XUT to Xrn1 (*xrn1Δ*/WT) and NMD (*upf1Δ*/WT) is also presented.

C. Box-plot showing the RNA-Seq signals (densities, tag/nt, log_2 scale) for the same set of XUTs as in B (1321), in WT cells grown in CSM with glucose (control) or undergoing glucose starvation (- Glucose), followed by a treatment with CHX or mock (DMSO) - see experimental scheme in panel A. *** *P*-value < 0.001; ns, not significant upon two-sided Wilcoxon rank-sum test (adjusted for multiple testing with the Benjamini–Hochberg procedure).

Figure 4. Translational landscape of XUTs.

A. Experimental scheme. Ribo-Seq libraries were prepared from biological duplicates of WT and *upf1Δ* cells grown in native conditions or treated for 15 minutes with CHX (100 µg/ml final concentration). SmORFs (≥ 5 codons, starting with an AUG) were detected using the ribotricer software (Choudhary *et al.*, 2020), pooling all conditions together (list 1) or analyzing them separately (list 2). A third list was produced from list 2 upon application of a signal threshold (≥ 10 reads/smORF). See lists in Table S2.

B. Venn diagram showing the number of XUTs detected as translated by Ribotricer (list 2) in each of the indicated conditions. See also Table S2.

C. Metagene of Ribo-Seq signals along the 633 translated XUTs (list 2). For each condition, the densities (tag/nt, log_2) along the XUTs +/- 200 nt were piled up, then the average signal was plotted. The shading surrounding each line denotes the 95% confidence interval.

D. Heatmap view of the Ribo-Seq signals (densities, tag/nt) from positions -50 to +150 relative to the AUG codon of the smORF showing the highest signal for the 510 NMD-sensitive and 123 NMD-insensitive XUTs detected as translated. A separate heatmap is shown for each condition.

E. Box-plot showing the size of the 3' UTR for the 510 NMD-sensitive and 123 NMD-insensitive XUTs detected as translated. When several translated smORFs have been identified for a same XUT, the size of the 3' UTR was computed using the smORF showing the highest Ribo-Seq signal. The *P*-value obtained upon two-sided Wilcoxon rank-sum test is indicated.

Figure 5. The NMD-sensitivity of *XUT0741* depends on its long 3' UTR.

A. Schematic representation of the native and mutant alleles of *XUT0741*. The transcript and the coding region are represented as a red line and a blue arrow, respectively. The red bars represent the stop codons that are in the same frame as the smORF. The sequence of the smORF in the native *XUT0741* is indicated. The length of the coding region and of the 3'UTR is shown beside each allele.

B. WT and *upf1Δ* cells expressing the different alleles of *XUT0741* were grown to mid-log phase, at 30°C, in YPD medium. After total RNA extraction, the levels of each transcript were assessed by strand-specific RT-qPCR, and then normalized on *scR1*. The grey bars correspond to the levels of the different alleles of *XUT0741* in WT cells (*y*-axis on the left), the level of the native XUT being set to 1. The black bars represent the NMD-sensitivity of each allele (*y*-axis on the right), calculated as the ratio between the mean levels in the *upf1* mutant and the mean levels in the WT strain. Mean and SD values were calculated from three independent biological replicates. ** $P < 0.01$; *** $P < 0.001$; ns, not significant upon t-test.

C. Schematic representation of the chimera construct, combining the 5' UTR and extended coding region of the *xut0741-d* allele to the long 3' UTR of the native *XUT0741*. Same representation as in A.

D. WT and *upf1Δ* cells expressing the native *XUT0741*, the *xut0741-d* allele and the chimera were grown as described above. Total RNA was extracted and analyzed by Northern blot. The different alleles of *XUT0741* and *scR1* (loading control) were detected using ^{32}P-labelled AMO1762 and AMO1482 oligonucleotides, respectively. The star indicates an uncharacterized RNA species that might correspond to a transcriptional isoform or processing product of the chimera.

E. Quantification of the signals from Northern blot. Mean and SEM values were calculated from four independent biological replicates. ** $P < 0.01$; ns, not significant upon t-test.

Figure 6. Detection of a translation product derived from NMD-sensitive XUT reporter in WT cells.

A. Schematic representation of the tagged *xut0741-b* alleles, using the same color code as in Figure 5A.

B. WT and *upf1Δ* cells expressing the native *XUT0741* or the *xut0741-b* allele fused to a C-terminal 3FLAG tag (*xut0741-b-FLAG*) were grown to mid-log phase, at 30°C, in YPD medium. Total RNA was extracted and analyzed by Northern blot. *XUT0741* and *scR1* were detected as described above.

C. WT and *upf1Δ* cells expressing the native *XUT0741*, the *xut0741-b-FLAG* or the *SL-xut0741-b-FLAG* alleles were grown as above. Protein extracts (40 µg) were separated by poly-acrylamide gel electrophoresis and then transferred to a nitrocellulose membrane. The size of the protein ladder bands is indicated on the left of the panel. Pgk1 was used as a loading control.

D. WT and *upf1Δ* cells expressing the *xut0741-b-FLAG* or the *SL-xut0741-b-FLAG* alleles were grown as to mid-log phase, at 30°C, in YPD medium. In addition, a sample of the WT cells expressing each allele was treated with CHX (100 µg/ml, final concentration), for 15 minutes. After total RNA extraction, the levels of the corresponding transcript were assessed by strand-specific RT-qPCR, and then normalized on *scR1*. The sensitivity of *xut0741-b-FLAG* (black bars) and *SL-xut0741-b-FLAG* (white bars) to CHX (left) and NMD (right) was calculated as the ratio between the RNA levels in CHX-treated *vs* untreated WT cells, and in *upf1Δ vs* WT cells, respectively. *** $P < 0.001$ upon t-test.

E. Model. Translation of 5' proximal smORF (red arrow) modulates the abundance of cytoplasmic lncRNAs. As they are translated, lncRNAs are protected from the degradation by the ribosomes. Then, as translation terminates far away from the poly(A) tail, NMD is activated, leading to the degradation of the transcript. In the same time, the peptide that has been produced can be exposed to the natural selection and possibly contributes to the progressive emergence of novel genes. The left part of the cartoon illustrates the idea that in the absence of translation, lncRNAs can also be efficiently degraded, independently of NMD. See main text for additional details.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. NMD-sensitive lncRNAs accumulate upon translation inhibition.

A. Schematic representation of the eukaryotic translation elongation cycle. The steps specifically inhibited by CHX and ANS are highlighted. The codons on the mRNA, the tRNAs and the amino acids (aa) are represented as rectangles, loops and circles, respectively (a color code is used to show the codon/tRNA/aa correspondence). The E, P and A sites of the ribosome are indicated.

B. Total RNA-Seq was performed using total RNA extracted from exponentially growing WT (YAM1) cells (grown as above) treated for 30 minutes with ANS (100 µg/ml, final concentration) or with an equal volume of DMSO (control). The scatter plot shows the RNA-Seq signals (tag densities, $\log_2$ scale) for the NMD-sensitive XUTs, mRNAs (light grey dots) and snoRNAs (black dots) in ANS-treated and control WT cells. The significantly up-regulated (ANS/control fold-change >2, *P*-value <0.05) and unaffected NMD-sensitive XUTs are represented as red and dark grey dots, respectively.

C. Sensitivity of NMD-sensitive XUTs and CUTs to CHX. The box-plot shows the global sensitivity to CHX of NMD-sensitive XUTs and 'strict' CUTs, in WT cells (CHX/control ratio of RNA-Seq signals). The 'strict' CUTs correspond to a subgroup of CUTs (621) that do not overlap XUTs.

Figure S2. Translation can also impact XUTs independently of NMD.

A. Total RNA-Seq was performed in WT (YAM1) and *upf1Δ* (YAM202) cells treated for 30 minutes with ANS (100 µg/ml, final concentration) or an equal volume of DMSO. The box-plot shows the densities (tag/not, $\log_2$) computed for the NMD-sensitive and NMD-insensitive XUTs. *** *P*-value < 0.001; ns, not significant upon two-sided Wilcoxon rank-sum test (adjusted for multiple testing with the Benjamini–Hochberg procedure).

B. Total RNA-Seq was performed using total RNA extracts from WT (YAM1) and *upf1Δ* (YAM202) cells, including a treatment with the Terminator 5'-phosphate-dependent exonuclease (which digests decapped RNAs) before the preparation of the libraries. The data are presented as a scatter plot

showing the RNA-Seq signals (tag densities, log_2 scale) for the NMD-sensitive XUTs, mRNAs (light grey) dots and snoRNAs (black dots). The red dots represent the 149 NMD-sensitive XUTs that are still detected as significantly up-regulated in *upf1Δ* cells (*upf1Δ*/WT fold-change >2, *P*-value <0.05) upon Terminator treatment. The other XUTs (Terminator-sensitive) are represented as dark grey dots.

C. Box-plot of the *upf1Δ*/WT fold-change for the NMD-sensitive XUTs computed using RNA-Seq data obtained from libraries prepared using total RNA extracts submitted to rRNA depletion (Ribo-) or Terminator digestion (Terminator).

Figure S3. XUTs do not globally accumulate in stress conditions associated to translation initiation inhibition.

A. XUTs landscape in CSM medium. Total RNA-Seq was performed in WT (YAM1) and *xrn1Δ* (YAM6) cells grown to mid-log phase in Complete Synthetic Medium (CSM). Densities (tag/not, log_2) were computed for XUTs, mRNAs (light grey dots) and snoRNAs (black dots). The 1321 XUTs up-regulated in the *xrn1* mutant (*xrn1Δ*/WT fold-change >2, *P*-value <0.05) are highlighted in red. The dark grey dots correspond to the other XUTs, the expression of which is not significantly affected.

B. Landscape of NMD-sensitive XUTs in CSM medium. Same as above, using WT (YAM1) and *upf1Δ* (YAM202) cells grown in CSM. The red dots represent the 779 XUTs defined as NMD-sensitive in this condition (*upf1Δ*/WT fold-change >2, *P*-value <0.05).

C. Experimental scheme. WT (YAM1) cells were grown to mid-log phase in CSM with glucose as carbon source, and then shifted for 16 minutes in CSM where glucose has been replaced by glycerol and ethanol (glucose starvation). In parallel, control cells were maintained for the same time in glucose-containing CSM. CHX (100 µg/ml final concentration) or an equal volume of DMSO (Mock) was then added to each sample. Cells were harvested after 15 minutes of treatment, then total RNA was extracted. Note that the CSM medium used here is different from the rich medium (YPD) that was originally used to annotate XUTs, so that we had to re-define the XUTs landscape in CSM (see above).

D. Analysis of published RNA-Seq data obtained in WT cells grown in CSM and then shifted for 16 min at 42°C (Bresson *et al.*, 2020). Densities (tag/nt) were computed for the 1335 XUTs expressed in CSM (see A), which were then separated according their sensitivity to NMD (see B). The sensitivity of each of these XUTs to heat-shock is presented as an heatmap of the fold-change (log_2 scale), relative to the control (unstressed) cells.

Figure S4. Translational landscape of XUTs.

A. Metagene of Ribo-Seq signals along the 1031 XUTs that were not detected as translated upon analysis using the Ribotricer method, separating the different conditions (*i.e.* XUTs excluded from list 2). For each condition, the densities (tag/nt, log_2) along the XUTs +/- 200 nt were piled up, then the average signal was plotted. The shading surrounding each line denotes the 95% confidence interval.

B. Box-plot representation of the size of the 1270 translated smORFs of XUTs (list 2). The mean and median values are indicated.

C. Histogram showing the number of translated smORFs per XUTs (for the 1270 smORFs and 633 XUTs of list 2).

D. Pie chart showing for the 633 translated XUTs (list 2) the position of the smORF with the highest Ribo-Seq signal relative to all the smORFs predicted across the XUT sequence (≥ 5 codons, starting with an AUG).

Figure S5. The NMD-sensitivity of *XUT0741* depends on its long 3' UTR.

A. Snapshot of Ribo-Seq signals across *XUT0741* in WT and *upf1Δ* cells, with or without CHX treatment. For each condition, the signals (tag/nt) obtained for the two biological replicates were added. *XUT0741* is depicted as a red line. The blue arrow represents the single smORF detected as actively translated in our analysis.

B. WT and *upf1Δ* cells expressing the different alleles of *XUT0741* (see Figure 5A) were grown to mid-log phase, at 30°C, in YPD medium. Total RNA was extracted and then analyzed by Northern blot.

XUT0741, *XUT1678* (and the overlapping *SUT768*) and *scR1* (loading control) were detected using ^{32}P-labelled AMO1762, AMO1595 and AMO1482 oligonucleotides, respectively.

Figure S6. Detection of a translation product derived from NMD-sensitive XUT reporter in WT cells.

A-B. WT and *upf1Δ* cells expressing the *xut0741-b-FLAG* or the *SL-xut0741-b-FLAG* alleles were grown as to mid-log phase, at 30°C, in YPD medium. In addition, a sample of the WT cells expressing each allele was treated with CHX (100 µg/ml, final concentration), for 15 minutes. After total RNA extraction, the levels of the NMD-sensitive *XUT1092* (A) and the NMD-sensitive *XUT1186* (B) were assessed by strand-specific RT-qPCR, and then normalized on *scR1*. The sensitivity of each XUT to CHX and NMD was calculated as the ratio between the normalized RNA levels in CHX-treated *vs* untreated WT cells, and in *upf1Δ vs* WT cells, respectively, for the *xut0741-b-FLAG* (black bars) and *SL-xut0741-b-FLAG* (white bars) backgrounds. ns, not significant upon t-test.

Figure S7. Phasing of Ribo-Seq data.

For each dataset, the P-site of the different k-mers (25-mers to 35-mers) was predicted with RiboWaltz (Lauria et al., 2018). As a quality control, for each k-mer, we calculated for the protein-coding genes the fraction of reads that are in-frame with the expected ORF (mentioned as P0). D930T01-02 : WT – native conditions, replicates 1-2 ; D930T03-04 : WT – CHX, replicates 1-2 ; D930T05-06 : *upf1Δ* – native conditions, replicates 1-2 ; D930T07-08 : *upf1Δ* – CHX, replicates 1-2.

REFERENCES

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118. 10.1038/nature03060.

Andjus, S., Morillon, A., and Wery, M. (2021). From Yeast to Mammals, the Nonsense-Mediated mRNA Decay as a Master Regulator of Long Non-Coding RNAs Functional Trajectory. Noncoding RNA *7*. 10.3390/ncrna7030044.

Apcher, S., Daskalogianni, C., Lejeune, F., Manoury, B., Imhoos, G., Heslop, L., and Fahraeus, R. (2011). Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proceedings of the National Academy of Sciences of the United States of America *108*, 11572-11577. 10.1073/pnas.1104104108.

Arendsee, Z.W., Li, L., and Wurtele, E.S. (2014). Coming of age: orphan genes in plants. Trends Plant Sci *19*, 698-708. 10.1016/j.tplants.2014.07.003.

Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A., Brocard, M., and Couso, J.P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. eLife *3*, e03528. 10.7554/eLife.03528.

Baudin-Baillieu, A., Hatin, I., Legendre, R., and Namy, O. (2016). Translation Analysis at the Genome Scale by Ribosome Profiling. Methods Mol Biol *1361*, 105-124. 10.1007/978-1-4939-3079-1_7.

Beelman, C.A., and Parker, R. (1994). Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast MFA2 mRNA. The Journal of biological chemistry *269*, 9687-9692.

Blevins, W.R., Ruiz-Orera, J., Messeguer, X., Blasco-Moreno, B., Villanueva-Canas, J.L., Espinar, L., Diez, J., Carey, L.B., and Alba, M.M. (2021). Uncovering de novo gene birth in yeast using deep transcriptomics. Nat Commun *12*, 604. 10.1038/s41467-021-20911-3.

Bresson, S., Shchepachev, V., Spanos, C., Turowski, T.W., Rappsilber, J., and Tollervey, D. (2020). Stress-Induced Translation Inhibition through Rapid Displacement of Scanning Initiation Factors. Molecular cell *80*, 470-484 e478. 10.1016/j.molcel.2020.09.021.

Carlevaro-Fita, J., Rahim, A., Guigo, R., Vardy, L.A., and Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. Rna *22*, 867-882. 10.1261/rna.053561.115.

Carvunis, A.R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and de novo gene birth. Nature *487*, 370-374. 10.1038/nature11184.

Celik, A., Baker, R., He, F., and Jacobson, A. (2017). High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. Rna *23*, 735-748. 10.1261/rna.060541.116.

Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. eLife *7*. 10.7554/eLife.32536.

Chen, J., Brunner, A.D., Cogan, J.Z., Nunez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., and Weissman, J.S. (2020). Pervasive functional translation of noncanonical human open reading frames. Science *367*, 1140-1146. 10.1126/science.aay0262.

Choudhary, S., Li, W., and A, D.S. (2020). Accurate detection of short and long active ORFs using Ribo-seq data. Bioinformatics *36*, 2053-2059. 10.1093/bioinformatics/btz878.

D'Lima, N.G., Ma, J., Winkler, L., Chu, Q., Loh, K.H., Corpuz, E.O., Budnik, B.A., Lykke-Andersen, J., Saghatelian, A., and Slavoff, S.A. (2017). A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol *13*, 174-180. 10.1038/nchembio.2249.

Dehecq, M., Decourty, L., Namane, A., Proux, C., Kanaan, J., Le Hir, H., Jacquier, A., and Saveanu, C. (2018). Nonsense-mediated mRNA decay involves two distinct Upf1-bound complexes. EMBO J *37*. 10.15252/embj.201899278.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101-108. 10.1038/nature11233.

Douka, K., Birds, I., Wang, D., Kosteletos, A., Clayton, S., Byford, A., Vasconcelos, E.J.R., O'Connell, M.J., Deuchars, J., Whitehouse, A., and Aspden, J.L. (2021). Cytoplasmic long noncoding RNAs are differentially regulated and translated during human neuronal differentiation. Rna *27*, 1082-1101. 10.1261/rna.078782.121.

Gardner, L.B. (2008). Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. Mol Cell Biol *28*, 3729-3741. 10.1128/MCB.02284-07.

Garreau de Loubresse, N., Prokhorova, I., Holtkamp, W., Rodnina, M.V., Yusupova, G., and Yusupov, M. (2014). Structural basis for the inhibition of the eukaryotic ribosome. Nature *513*, 517-522. 10.1038/nature13737.

He, F., and Jacobson, A. (2015). Control of mRNA decapping by positive and negative regulatory elements in the Dcp2 C-terminal domain. Rna *21*, 1633-1647. 10.1261/rna.052449.115.

Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Coller, J. (2009). Co-translational mRNA decay in Saccharomyces cerevisiae. Nature *461*, 225-229. 10.1038/nature08265.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell reports *8*, 1365-1379. 10.1016/j.celrep.2014.07.045.

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. 10.1016/j.cell.2011.10.002.

Jarroux, J., Morillon, A., and Pinskaya, M. (2017). History, Discovery, and Classification of lncRNAs. Adv Exp Med Biol *1008*, 1-46. 10.1007/978-981-10-5203-3_1.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907-915. 10.1038/s41587-019-0201-4.

Kopp, F., and Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell *172*, 393-407. 10.1016/j.cell.2018.01.011.

Kuroha, K., Tatematsu, T., and Inada, T. (2009). Upf1 stimulates degradation of the product derived from aberrant messenger RNA containing a specific nonsense mutation by the proteasome. EMBO Rep *10*, 1265-1271. 10.1038/embor.2009.200.

Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. eLife *3*, e01257. 10.7554/eLife.01257.

Lauria, F., Tebaldi, T., Bernabo, P., Groen, E.J.N., Gillingwater, T.H., and Viero, G. (2018). riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. PLoS Comput Biol *14*, e1006169. 10.1371/journal.pcbi.1006169.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930. 10.1093/bioinformatics/btt656.

Lorenzi, L., Chiu, H.S., Avila Cobos, F., Gross, S., Volders, P.J., Cannoodt, R., Nuytens, J., Vanderheyden, K., Anckaert, J., Lefever, S., et al. (2021). The RNA Atlas expands the catalog of human non-coding RNAs. Nat Biotechnol *39*, 1453-1465. 10.1038/s41587-021-00936-1.

Losson, R., and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. Proceedings of the National Academy of Sciences of the United States of America *76*, 5134-5137. 10.1073/pnas.76.10.5134.

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

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. eLife *4*, e06722. 10.7554/eLife.06722.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10-12. 10.14806/ej.17.1.200.

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K.I., Clohessy, J.G., and Pandolfi, P.P. (2017). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature *541*, 228-232. 10.1038/nature21034.

McLysaght, A., and Hurst, L.D. (2016). Open questions in the study of de novo genes: what, how and why. Nat Rev Genet *17*, 567-578. 10.1038/nrg.2016.78.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet *36*, 1073-1078. 10.1038/ng1429.

Muhlrad, D., Decker, C.J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. Mol Cell Biol *15*, 2145-2156.

Muhlrad, D., and Parker, R. (1999). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. Rna *5*, 1299-1307.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038-1042.

Papadopoulos, C., Callebaut, I., Gelly, J.C., Hatin, I., Namy, O., Renard, M., Lespinet, O., and Lopes, A. (2021). Intergenic ORFs as elementary structural modules of de novo gene birth and protein evolution. Genome Res. 10.1101/gr.275638.121.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. Genetics *191*, 671-702. 10.1534/genetics.111.137265.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127-131. 10.1038/nature12121.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400-1412. 10.1016/j.cell.2015.05.008.

Ponting, C.P., and Haerty, W. (2022). Genome-Wide Analysis of Human Long Noncoding RNAs: A Provocative Review. Annu Rev Genomics Hum Genet. 10.1146/annurev-genom-112921-123710.

Renganathan, A., and Felley-Bosco, E. (2017). Long Noncoding RNAs in Cancer and Therapeutic Potential. Adv Exp Med Biol *1008*, 199-222. 10.1007/978-981-10-5203-3_7.

Rodriguez-Lopez, M., Anver, S., Cotobal, C., Kamrad, S., Malecki, M., Correia-Melo, C., Hoti, M., Townsend, S., Marguerat, S., Pong, S.K., et al. (2022). Functional profiling of long intergenic non-coding RNAs in fission yeast. eLife *11*. 10.7554/eLife.76000.

Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M. (2014). Long non-coding RNAs as a source of new peptides. eLife *3*, e03523. 10.7554/eLife.03523.

Saha, P., Verma, S., Pathak, R.U., and Mishra, R.K. (2017). Long Noncoding RNAs in Mammalian Development and Diseases. Adv Exp Med Biol *1008*, 155-198. 10.1007/978-981-10-5203-3_6.

Schmitt, A.M., and Chang, H.Y. (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell *29*, 452-463. 10.1016/j.ccell.2016.03.010.

Schmitz, J.F., Ullrich, K.K., and Bornberg-Bauer, E. (2018). Incipient de novo genes can evolve from frozen accidents that escaped rapid transcript turnover. Nat Ecol Evol *2*, 1626-1632. 10.1038/s41559-018-0639-7.

Serdar, L.D., Whiteside, D.L., and Baker, K.E. (2016). ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. Nat Commun *7*, 14021. 10.1038/ncomms14021.

Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol *9*, 59-64. 10.1038/nchembio.1120.

Smith, J.E., Alvarez-Dominguez, J.R., Kline, N., Huynh, N.J., Geisler, S., Hu, W., Coller, J., and Baker, K.E. (2014). Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. Cell reports *7*, 1858-1866. 10.1016/j.celrep.2014.05.023.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res *27*, 491-499. 10.1101/gr.209601.116.

Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol *22*, 96-118. 10.1038/s41580-020-00315-9.

Stevens, A., Hsu, C.L., Isham, K.R., and Larimer, F.W. (1991). Fragments of the internal transcribed spacer 1 of pre-rRNA accumulate in Saccharomyces cerevisiae lacking 5'----3' exoribonuclease 1. J Bacteriol *173*, 7024-7028.

Szachnowski, U., Andjus, S., Foretek, D., Morillon, A., and Wery, M. (2019). Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome. Life Sci Alliance *2*, e201900407. 10.26508/lsa.201900407.

Tisseur, M., Kwapisz, M., and Morillon, A. (2011). Pervasive transcription - Lessons from yeast. Biochimie *93*, 1889-1896. 10.1016/j.biochi.2011.07.001.

Vakirlis, N., Acar, O., Hsu, B., Castilho Coelho, N., Van Oss, S.B., Wacholder, A., Medetgul-Ernar, K., Bowman, R.W., 2nd, Hines, C.P., Iannotta, J., et al. (2020). De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. Nat Commun *11*, 781. 10.1038/s41467-020-14500-z.

Van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non coding RNA in yeast. Nature *475*, 114-117.

van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P.B., de Bruijn, E., Hao, W., MacInnes, A.W., Cuppen, E., and Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol *15*, R6. 10.1186/gb-2014-15-1-r6.

van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.L., et al. (2019). The Translational Landscape of the Human Heart. Cell *178*, 242-260 e229. 10.1016/j.cell.2019.05.010.

Van Oss, S.B., and Carvunis, A.R. (2019). De novo gene birth. PLoS Genet *15*, e1008160. 10.1371/journal.pgen.1008160.

Wacholder, A., Acar, O., and Carvunis, A.-R. (2021). A reference translatome map reveals two modes of protein evolution. bioRxiv, 2021.2007.2017.452746. 10.1101/2021.07.17.452746.

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.S., Gautheret, D., and Morillon, A. (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Molecular cell *61*, 379-392. 10.1016/j.molcel.2015.12.020.

Wery, M., Gautier, C., Descrimes, M., Yoda, M., Migeot, V., Hermand, D., and Morillon, A. (2018a). Bases of antisense lncRNA-associated regulation of gene expression in fission yeast. PLoS Genet *14*, e1007465. 10.1371/journal.pgen.1007465.

Wery, M., Gautier, C., Descrimes, M., Yoda, M., Vennin-Rendos, H., Migeot, V., Gautheret, D., Hermand, D., and Morillon, A. (2018b). Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. Rna *24*, 196-208. 10.1261/rna.063446.117.

Wu, C.C., Zinshteyn, B., Wehner, K.A., and Green, R. (2019). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. Molecular cell *73*, 959-970 e955. 10.1016/j.molcel.2018.12.009.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725-737.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

Yao, R.W., Wang, Y., and Chen, L.L. (2019). Cellular functions of long noncoding RNAs. Nat Cell Biol *21*, 542-551. 10.1038/s41556-019-0311-8.

Zanet, J., Benrabah, E., Li, T., Pelissier-Monier, A., Chanut-Delalande, H., Ronsin, B., Bellen, H.J., Payre, F., and Plaza, S. (2015). Pri sORF peptides induce selective proteasome-mediated protein processing. Science *349*, 1356-1358. 10.1126/science.aac5677.

Zhao, L., Saelao, P., Jones, C.D., and Begun, D.J. (2014). Origin and spread of de novo genes in Drosophila melanogaster populations. Science *343*, 769-772. 10.1126/science.1248286.

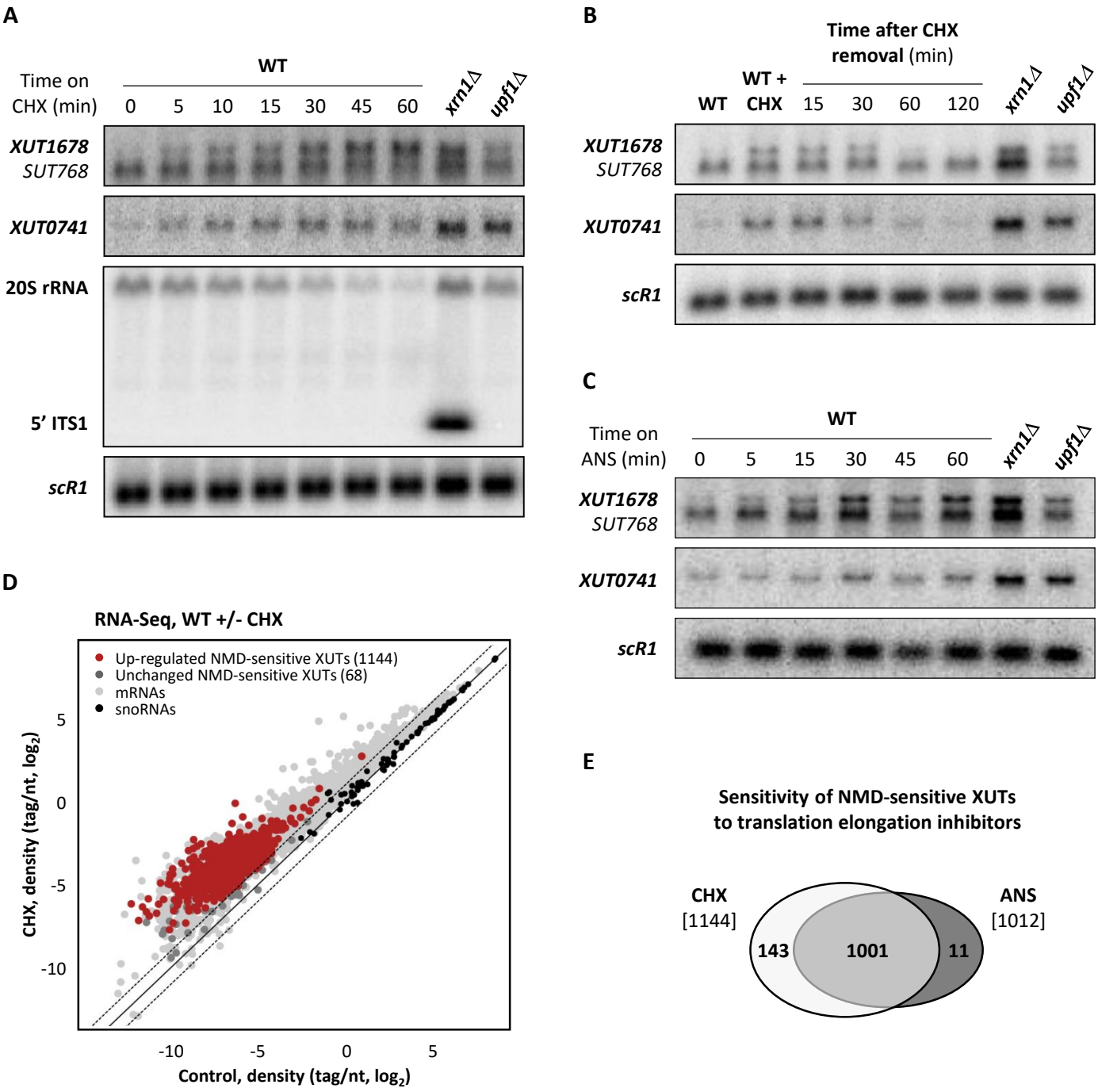
A
Time on CHX (min)
WT
0 5 10 15 30 45 60
xrn1Δ
upf1Δ
XUT1678
SUT768
XUT0741
20S rRNA
5' ITS1
scR1
B
Time after CHX removal (min)
WT
WT + CHX
15 30 60 120
xrn1Δ
upf1Δ
XUT1678
SUT768
XUT0741
scR1
C
Time on ANS (min)
WT
0 5 15 30 45 60
xrn1Δ
upf1Δ
XUT1678
SUT768
XUT0741
scR1
D
RNA-Seq, WT +/- CHX
Up-regulated NMD-sensitive XUTs (1144)
Unchanged NMD-sensitive XUTs (68)
mRNAs
snoRNAs
CHX, density (tag/nt, log2)
Control, density (tag/nt, log2)
5
0
-5
-10
-10 -5 0 5
E
Sensitivity of NMD-sensitive XUTs to translation elongation inhibitors
CHX
[1144]
ANS
[1012]
143
1001
11

Figure 1

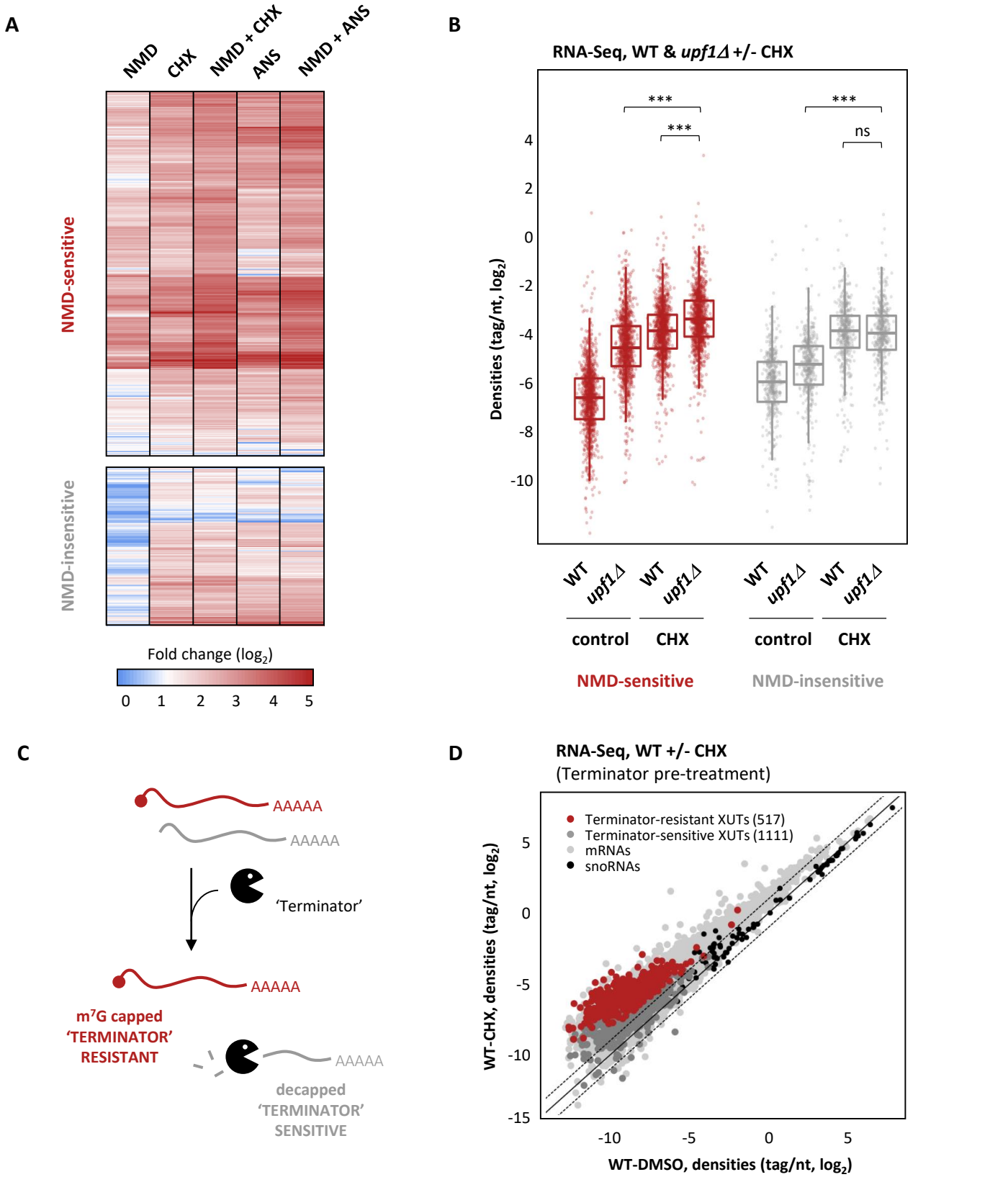
A
NMD
CHX
NMD + CHX
ANS
NMD + ANS
NMD-sensitive
NMD-insensitive
Fold change (log2)
0 1 2 3 4 5
B
RNA-Seq, WT & upf1Δ +/- CHX
***
***
***
ns
Densities (tag/nt, log2)
4 2 0 -2 -4 -6 -8 -10
WT upf1Δ WT upf1Δ WT upf1Δ WT upf1Δ
control CHX control CHX
NMD-sensitive
NMD-insensitive
C
AAAAA
AAAAA
'Terminator'
AAAAA
m7G capped
'TERMINATOR'
RESISTANT
AAAAA
decapped
'TERMINATOR'
SENSITIVE
D
RNA-Seq, WT +/- CHX
(Terminator pre-treatment)
Terminator-resistant XUTs (517)
Terminator-sensitive XUTs (1111)
mRNAs
snoRNAs
WT-CHX, densities (tag/nt, log2)
5 0 -5 -10 -15
-10 -5 0 5
WT-DMSO, densities (tag/nt, log2)

Figure 2

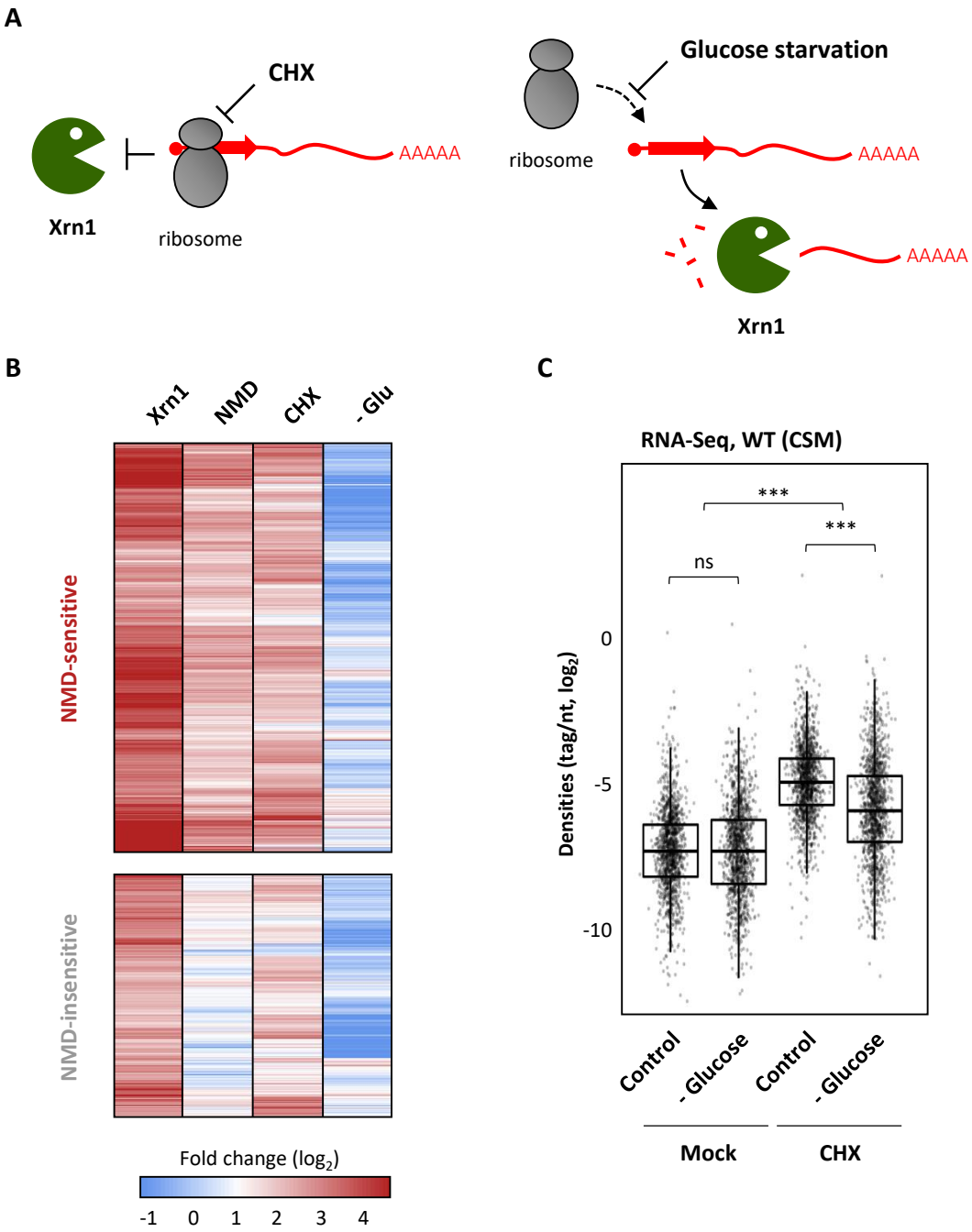
A
CHX
Xrn1
ribosome
AAAAA
Glucose starvation
ribosome
AAAAA
AAAAA
Xrn1
B
Xrn1
NMD
CHX
- Glu
NMD-sensitive
NMD-insensitive
Fold change (log2)
-1 0 1 2 3 4
C
RNA-Seq, WT (CSM)
***
***
ns
Densities (tag/nt, log2)
0
-5
-10
Control
- Glucose
Control
- Glucose
Mock
CHX

Figure 3

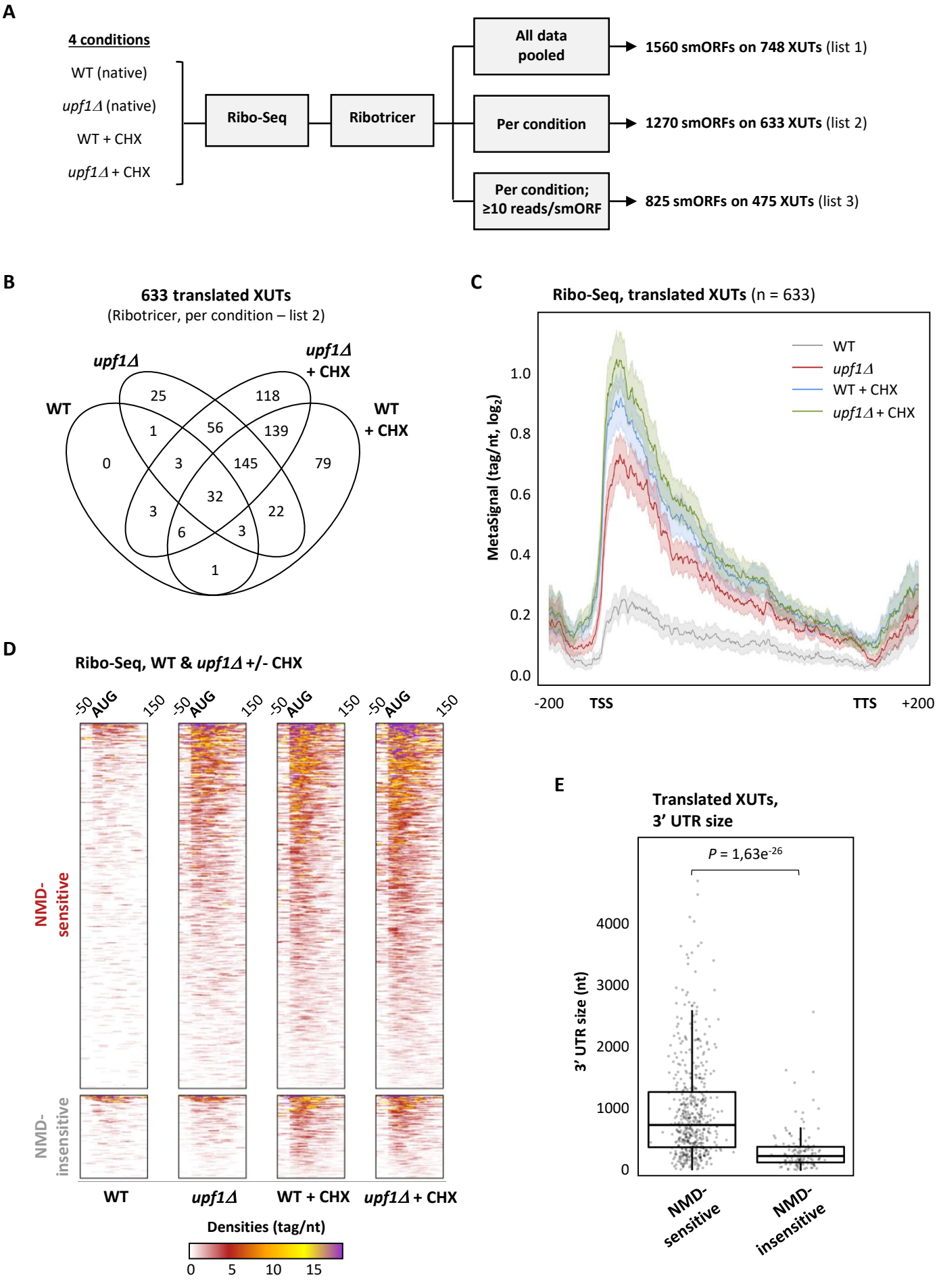
A
4 conditions
WT (native)
upf1Δ (native)
WT + CHX
upf1Δ + CHX
Ribo-Seq
Ribotricer
All data pooled
1560 smORFs on 748 XUTs (list 1)
Per condition
1270 smORFs on 633 XUTs (list 2)
Per condition; ≥10 reads/smORF
825 smORFs on 475 XUTs (list 3)
B
633 translated XUTs
(Ribotricer, per condition – list 2)
upf1Δ
upf1Δ + CHX
WT
WT + CHX
25
118
1
56
139
0
3
145
79
32
3
22
6
3
1
C
Ribo-Seq, translated XUTs (n = 633)
WT
upf1Δ
WT + CHX
upf1Δ + CHX
MetaSignal (tag/nt, log2)
1.0
0.8
0.6
0.4
0.2
0.0
-200
TSS
TTS
+200
D
Ribo-Seq, WT & upf1Δ +/- CHX
-50
AUG
150
NMD-sensitive
NMD-insensitive
WT
upf1Δ
WT + CHX
upf1Δ + CHX
Densities (tag/nt)
0
5
10
15
E
Translated XUTs,
3' UTR size
P = 1,63e-26
3' UTR size (nt)
4000
3000
2000
1000
0
NMD-sensitive
NMD-insensitive

Figure 4

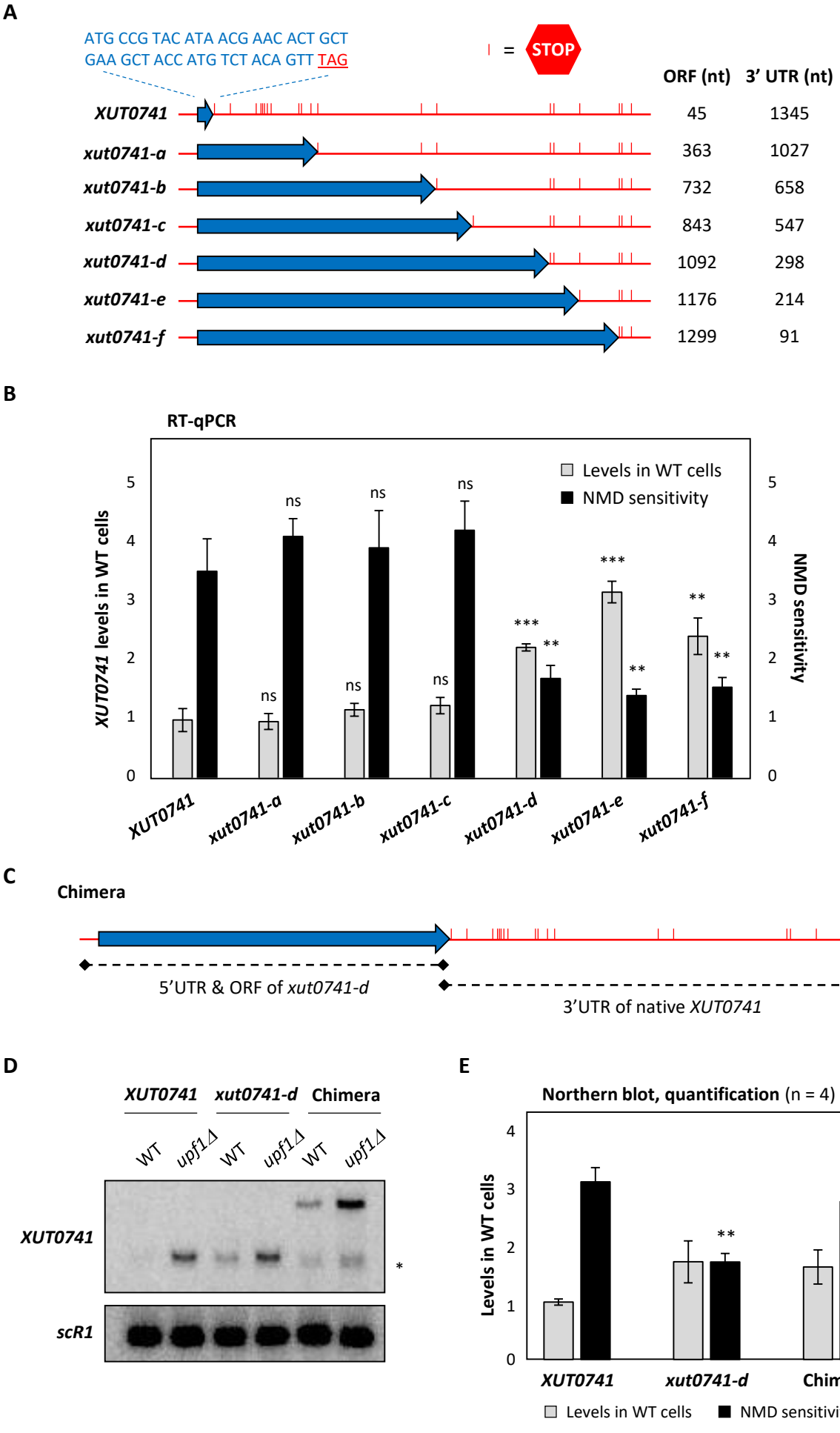
A
ATG CCG TAC ATA ACG AAC ACT GCT
GAA GCT ACC ATG TCT ACA GTT TAG
= STOP
ORF (nt)
3' UTR (nt)
XUT0741 45 1345
xut0741-a 363 1027
xut0741-b 732 658
xut0741-c 843 547
xut0741-d 1092 298
xut0741-e 1176 214
xut0741-f 1299 91
B
RT-qPCR
Levels in WT cells
NMD sensitivity
XUT0741 levels in WT cells
NMD sensitivity
ns
***
**
XUT0741
xut0741-a
xut0741-b
xut0741-c
xut0741-d
xut0741-e
xut0741-f
C
Chimera
5'UTR & ORF of xut0741-d
3'UTR of native XUT0741
D
XUT0741
xut0741-d
Chimera
WT
upf1Δ
XUT0741
scR1
*
E
Northern blot, quantification (n = 4)
Levels in WT cells
NMD sensitivity
ns
**
XUT0741
xut0741-d
Chimera
Levels in WT cells
NMD sensitivity

Figure 5

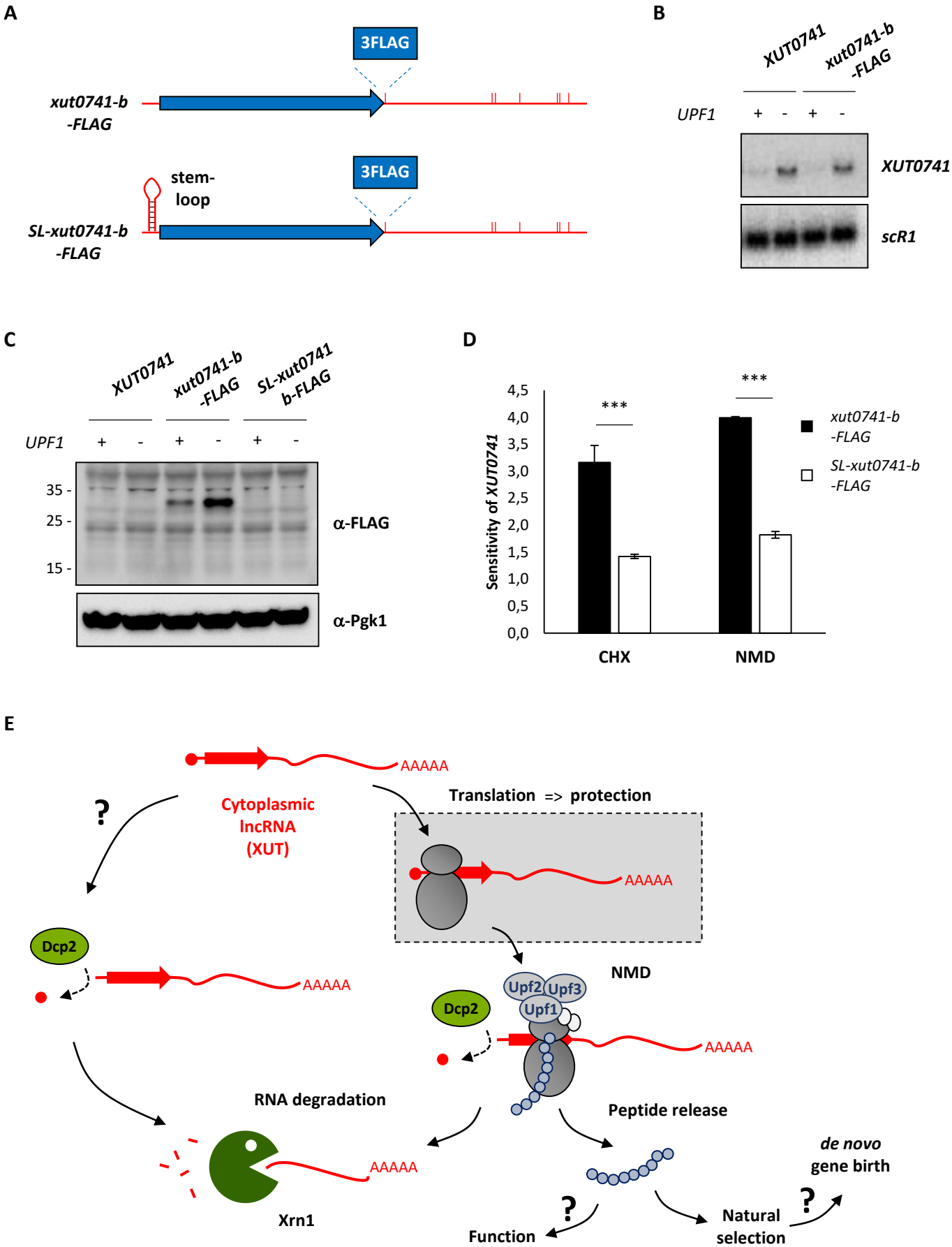
A
3FLAG
xut0741-b
-FLAG
stem-
loop
3FLAG
SL-xut0741-b
-FLAG
B
XUT0741
xut0741-b
-FLAG
UPF1
+ - + -
XUT0741
scR1
C
XUT0741
xut0741-b
-FLAG
SL-xut0741
b-FLAG
UPF1
+ - + - + -
35 -
25 -
15 -
α-FLAG
α-Pgk1
D
Sensitivity of XUT0741
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
***
***
xut0741-b
-FLAG
SL-xut0741-b
-FLAG
CHX
NMD
E
AAAAA
Cytoplasmic
lncRNA
(XUT)
?
Translation => protection
AAAAA
Dcp2
AAAAA
NMD
Upf2
Upf3
Upf1
Dcp2
AAAAA
RNA degradation
Peptide release
de novo
gene birth
AAAAA
Xrn1
?
Function
Natural
selection
?

Figure 6