Molecular Cell
Supplemental Information

Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure

Maxime Wery, Marc Descrimes, Nicolas Vogt, Anne-Sophie Dallongeville, Daniel Gautheret, and Antonin Morillon
A

|        | Galactose | Glucose |
|--------|-----------|---------|
| NRD1   | +         | -       |
| P_{GAL1::NRD1} | +     | -       |
| NRD1   | +         | -       |
| P_{GAL1::NRD1} | +     | -       |
| XRN1   | +         | -       |
| NEL025C| -L        | -S      |
|        | ND ND ND ND | 1 1 5 296 |

B

|        | WT | xrn1Δ | pap1-1 | xrn1Δ | pap1-1 | rrp6Δ | pap1-1 | rrp6Δ |
|--------|----|-------|--------|-------|--------|-------|--------|-------|
| Temp. (°C) | 25 | 37    | 25     | 37    | 25     | 37    | 25     | 37    |

- XUT1678
- SUT768
- RTL
- scR1
Figure S2 – Wery et al.
Distance between annotated 5' coordinates of overlapping SUTs & XUTs (nt)

Distance between annotated 3' coordinates of overlapping SUTs & XUTs (nt)

|         | 5'   | 3'   |
|---------|------|------|
| Median  | 2 nt | 29 nt|
| Average | 10 nt| 55 nt|
Figure S4 – Wery et al.
Figure S5 – Wery et al.
Figure S6 – Wery et al.
Figure S7 – Wery et al.
Supplemental figures legends

Figure S1. Synthesis of overlapping SUT768/XUT1678 does not involve the fail-safe termination complex NRD1 but requires the Pap1 poly(A)-polymerase; related to Figure 1.

A. NEL025C detection by Northern blot using the same membrane and probe shown in Figure 1C, with longer exposure. Indicated values correspond to NEL025C levels quantified by RT-qPCR and normalized on scr1. Normalized NEL025C level for wild-type cells shifted in glucose-containing medium was set to 1. ND: not determined.

B. XUT1678 accumulation depends on the Pap1 poly(A)-polymerase. YAM92 (WT), YAM97 (xrn1Δ), YAM1364 (pap1-1), YAM1416 (pap1-1 xrn1Δ), YAM2456 (rrp6Δ) and YAM2457 (pap1-1 rrp6Δ) cells were grown to mid-log phase in YPD medium at 25°C, then shifted at 37°C for 1 hour.

Figure S2. Comprehensive XUT landscape in S. cerevisiae; related to Figure 2.

A. Experimental strategy overview. Total RNA was extracted from exponentially growing WT and xrn1Δ cells of different laboratory strains of S. cerevisiae. After depletion of rRNA, strand-specific RNA-Seq libraries were constructed and paired-end sequenced. Reads were uniquely mapped on the S288C reference genome with a tolerance of 3 mismatches per read to take into account the existence of SNPs between the different strains. Transcripts were assembled using the ZINAR segmentation protocol, and then the most robust XUTs were selected using a min. 2-fold enrichment in the xrn1Δ mutant and a significant P-value upon differential expression analysis. In parallel, XUTs TSS were re-annotated using CAGE-Seq data obtained from decapping-deficient cells.

B. Abundance of XUTs in WT and xrn1Δ cells of haploid S288C, W303 and SK1, and diploid SK1 strains. Total RNA-Seq signals for XUTs in the WT and xrn1Δ of each background are shown as a heatmap (the color turns from yellow to red as the RNA-Seq signal increases). Each horizontal line corresponds to one XUT. Six different subsets of XUTs were defined: differentially expressed in only
one background (a); common to the two SK1 strains (c); common to S228C and W303 (d); common to three haploid strains (e); common to all backgrounds (f); all other possible combinations (b).

C. Overlap between the 1681 XUTs of S288C, the 1046 Dcp2-sensitive XUTs and the 101 Dcp2-sensitive IncRNAs (Geisler et al., 2012).

D. Size distribution of uniquely mapped CAGE-Seq reads from WT and dcp2-7 samples.

E. Distance between the TSS of the 625 XUTs re-annotated using CAGE-Seq and their former 5’ coordinates obtained from segmentation of RNA-Seq signals.

F-H. Distribution of CAGE-Seq (this work) signals in WT (solid grey lines) and dcp2-7 (solid black lines) cells shifted for 2 hours at 37°C and TSS-sequencing/L5p signals (Malabat et al., 2015) in WT (dashed grey lines) and upf1Δ (dashed black lines) cells around TSS of mRNAs (F), SUTs (G) and XUTs (H).

Figure S3. XUTs overlapping SUTs display 3’-extensions; related to Figure 3.

Distance between the 5’ (x axis) and 3’ (y axis) annotated coordinates of overlapping SUTs and XUTs (with ≥ 75% of overlap, n=292) was determined. Each grey dot corresponds to a pair of SUT/XUT. The intersection between the horizontal and vertical black bars indicates the median value for each axis, and the length of the bars corresponds to the distance between the first and third quartiles.

Figure S4. Sensitivity of XUTs to decapping and NMD; related to Figure 4.

A. Snapshot of total RNA-Seq signals for XUT1678 in xrn1Δ, dcp2-7 and upf1Δ cells. RNA-Seq signals for WT (YAM1), xrn1Δ (YAM6) and upf1Δ (YAM202) cells, and for WT (YAM1) and dcp2-7 (YAM2283) cells shifted for 2 hours at 37°C were visualized as heatmaps using the VING software (Descrimes et al., 2015). Annotated ORFs, SUT and XUT are represented as blue, grey and red arrows, respectively.

B. Venn diagram showing the number of XUTs of S288C (Xrn1, n=1681) that are sensitive to NMD (Upf1, n=1229) and decapping (Dcp2, n=1046).

C. Overlap between the 1681 XUTs of S288C, the 1229 Upf1-sensitive XUTs and the 1146 uRNAs (Smith et al., 2014).
D. Venn diagram showing the number of S288C XUTs with total RNA-Seq mutant/WT ratio ≥2 in xrn1∆, upf1∆ and upf1∆ xrn1∆. Libraries were previously described (Malabat et al., 2015) and data were retrieved from the NCBI GEO database (accession number GSE64139).

E. Density plot representation of the distribution of total RNA-Seq mutant/WT ratio for XUTs (solid lines) and snoRNAs (dashed lines) in xrn1∆ (red), upf1∆ (blue) and upf1∆ xrn1∆ (black). Same data as above.

F. Snapshot of fragmented RNA (upper panel) and ribosome profiling (lower panel) signals for ARG1 mRNA (+ strand, dark blue) and the antisense XUT1678 (− strand, red) in upf1∆ cells. Libraries were previously described (Smith et al., 2014) and data were retrieved from NCBI BioProject accession number PRJNA245106. ARG1 mRNA and XUT1678 are represented by blue and red lines, respectively, below the two panels. Arrows represent (s)ORFs.

Figure S5. Antisense XUTs form dsRNA in vivo; related to Figure 5.

A. Size and 5’ base distribution of small RNAs produced upon RNAi reconstitution. WT (YAM1730), RNAi+ (YAM1725), xrn1∆ (YAM2271) and xrn1∆ RNAi+ (YAM1982) cells were grown to mid-log phase in YPD at 30°C. Libraries were constructed using purified small RNAs.

B. Proportion of 19-23 nt reads that mapped to rRNA, transposons (TY) and other repeated sequences, ORFs and XUTs was determined for each of the strains described above.

C. Snapshot of the small RNAs along the FAR1/XUT0521 locus. Densities of 19-23 nt uniquely mapped reads are shown for the strains described above. Signals for the + and – strands are shown in the upper and lower panels, respectively. ORFs, XUT and SUT are represented as in Figure 3C. Thin blue lines correspond to UTRs. Blue stars indicate the position of probes used for Northern blot analysis (see below). The snapshot was produced using VING (Descrimes et al., 2015).

D. Northern blot detection of small RNAs. RNAs < 100 nt purified from total RNA of exponentially growing WT (YAM1730), RNAi+ (YAM1725), xrn1∆ (YAM2271) and xrn1∆ RNAi+ (YAM1982) cells were separated on polyacrylamide gel, transferred on nylon membrane and chemically cross-linked. Small
RNAs derived from FAR1, TAT1 and TY1 mRNAs were detected using $^{32}$P-labelled oligonucleotides. snR52 was used as a loading control.

**Figure S6. Formation of dsRNA protects XUTs from NMD; related to Figure 6.**

**A.** Sensitivity of XUT0051 and XUT0521 to Xrn1 and NMD. Tag densities for XUT0051 and XUT0521 in xrn1Δ (YAM1), upf1Δ (YAM202) and their respective WT were computed from the total RNA-Seq data described in Figures S2 and 4A.

**B.** Growth curves. WT (YAM115) and tetOFF::MTR4 (BSY1756) cells were grown in YPD medium to mid-log phase at 30°C, then doxycycline was added at a 10 µg/ml final concentration. Growth was followed by measuring OD$_{600}$. Y-axis corresponds to (OD$_{600}$ at T$_x$)/(OD$_{600}$ at T$_0$), in log scale.

**C-D.** Metagene representation of Mtr4 binding along mRNAs (C) and XUTs (D). The upper and lower parts of each panel show the signals along the considered transcript class (mRNA or XUT) and along its antisense strand, respectively. TSS and TTS correspond to transcription start site and transcription termination site of the considered transcript, respectively. x-axis is in virtual nt. Solid and dashed lines correspond to specific Mtr4 binding signals and background signals obtained with an untagged strain, respectively. Libraries were previously described (Tuck and Tollervey, 2013) and data were retrieved from the NCBI GEO database (accession number GSE46742).

**E.** WT (YAM1), xrn1Δ (YAM6) and upf1Δ (YAM202) cells transformed with the pMD2 empty vector (black) or expressing the anti-complementary transcript of solo XUT1150 in trans (grey) were grown in CSM-U to mid-log phase at 30°C. Anti-complementary transcript of XUT1150 was quantified by strand-specific RT-qPCR from total RNA and normalized on scR1 (left panel) or detected by Northern blot (right panel) using the same RNA extracts (18S and 25S rRNA detected upon staining of the gel with ethidium bromide serve as loading controls).

**F-G.** WT (YAM1), xrn1Δ (YAM6) and upf1Δ (YAM202) cells transformed with the pMD2 empty vector (black bars) or expressing in trans the anti-complementary transcript (grey bars) of solo XUT1092 (F) and solo XUT1186 (G) were grown in CSM-U to mid-log phase at 30°C. Levels of the anti-
complementary transcript of XUT1092 and XUT1186 were determined and normalized as described above. XUTs levels in the WT strain, in each condition, was set to 1.

Figure S7. Segmentation of total RNA-Seq data using ZINAR; related to Experimental Procedures.

3880 segmentations were performed with ZINAR using different sets of parameters (sliding window size and signal threshold). For each of them, the number of properly detected ORFs and detected SUTs was plotted. Each point corresponds to one segmentation. The selected segmentation is highlighted in red (parameters used were sliding window = 5 nt; threshold = 146,88 tag/nt).
| Name       | Background | Genotype                                                                 | Reference/source                          |
|------------|------------|---------------------------------------------------------------------------|-------------------------------------------|
| YAM1       | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0                                         | Euroscarf                                 |
| YAM6       | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 xrn1::kanMX4                            | Euroscarf                                 |
| YAM92      | W303       | MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100                  | (He et al., 2003)                         |
| YAM97      | W303       | MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100 xrn1::ADE2       | (He et al., 2003)                         |
| YAM115     | BMA64-1A   | MATa ade2-1 his3-11,15 leu2-3,112 trp1Δ ura3-1 can1-100                  | B. Séraphin                               |
| YAM124     | BY4742     | MAT@ his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 rrp6::kanMX4                            | (Egecioglu et al., 2006)                 |
| YAM125     | BY4742     | MAT@ his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 trf4::kanMX4                            | (Egecioglu et al., 2006)                 |
| YAM127     | BY4742     | MAT@ his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 trf5::kanMX4                            | (Egecioglu et al., 2006)                 |
| YAM128     | BY4742     | MAT@ his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 trf4::his3MX6 rrp6::kanMX4              | (Egecioglu et al., 2006)                 |
| YAM129     | BY4742     | MAT@ his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 trf5::his3MX6 trf5::kanMX4              | (Egecioglu et al., 2006)                 |
| YAM143     | W303       | MATa ade2-1 his3-11,5 leu2-3,112 trp1-1 ura3-1 can1-100 HIS3::pGAL1-NRD1  | (Thiebaut et al., 2006)                  |
| YAM199     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ccr4::kanMX4                            | Euroscarf                                 |
| YAM200     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 dom34::kanMX4                           | Euroscarf                                 |
| YAM202     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 upf1::kanMX4                           | Euroscarf                                 |
| YAM203     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 upf2::kanMX4                           | Euroscarf                                 |
| YAM204     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 upf3::kanMX4                           | Euroscarf                                 |
| YAM225     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ism1::kanMX4                           | Euroscarf                                 |
| YAM226     | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ism7::kanMX4                           | Euroscarf                                 |
| YAM985     | W303       | MATa ade2-1 his3-11,5 leu2-3,112 trp1-1 ura3-1 can1-100 HIS3::pGAL1-NRD1  | This work                                 |
| BSY1756/YAM997 | BMA64-1A | MATa ade2-1 his3-11,5 leu2-3,112 trp1Δ2 ura3-1 can1-100 kanMX6-tetOFF::MTR4 | B. Séraphin                               |
| YAM1364    | W303       | MATa ade2-1 his3-11,5 leu2-3,112 trp1-1 ura3-1 can1-100 pap1-1            | D. Libri                                  |
| YAM1416    | W303       | MATa ade2-1 his3-11,5 leu2-3,112 trp1-1 ura3-1 can1-100 pap1-1 xrn1::kanMX4 | This work                                 |
| YAM1725    | W303       | MATcu LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 | (Drinnenberg et al., 2009)               |
| YAM1730    | W303       | MATcu leu2-3,112 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 | (Drinnenberg et al., 2009)               |
| YAM1977    | SK1        | MAT@ his4 leu2 lys2 ura3 trp1 ho::LYS2                                 | A. Nicolas                                |
| YAM1982    | W303       | MATcu LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 xrn1::His3MX6 | This work                                 |
| YAM2067    | SK1        | MATa/@ HIS4/his4 lys2/lys2 ura3/ura3 leu2/leu2 trp1/trp1 ho::LYS2/ho::LYS2 | This work                                 |
| YAM2068    | SK1        | MATa/@ HIS4/his4 lys2/lys2 ura3/ura3 leu2/leu2 trp1/trp1 ho::LYS2/ho::LYS2 xrn1::hphNT1/ | This work                                 |
| YAM2087    | SK1        | MAT@ his4 leu2 lys2 ura3 trp1 ho::LYS2 xrn1::hphNT1                     | This work                                 |
| YAM2271    | W303       | MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3::EGFP(S65T)-KanMX6 can1-100 xrn1::His3MX6 | This work                                 |
| YAM2283    | BY4741     | MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lys2Δ0 dcp2-7::URA3                  | (Wilson et al., 2007)                     |
| YAM2456    | W303       | MATa ade2-1 his3-11,15 leu2-3,112 trp1Δ ura3-1 can1-100 rrp6::URA3      | D. Libri                                  |
| YAM2457    | W303       | MATa ade2-1 his3-11,15 leu2-3,112 trp1Δ ura3-1 can1-100 pap1-1 rrp6::URA3 | D. Libri                                  |
| YAM262/ BY7115 | BY4741 | MATa his3Δ0 leu2Δ0 met- lys? ura3Δ0 dbp2::KanMX6                           | (Cloutier et al., 2012)                   |

Table S1. Yeast strains; related to Experimental Procedures.
## Table S2. Oligonucleotides; related to Experimental Procedures.

| ID     | Sequence 5'-3'                                      | Target                        |
|--------|-----------------------------------------------------|-------------------------------|
| AMO193 | GGCTGTAATGGGCTTCTGG                                  | scrR1 (qPCR)                  |
| AMO253 | GAATCGAGAAGGAGGTGTCCAGGG                             | TY1 mRNA (NB)                 |
| AMO415 | GTGCGGAATAGAGAAACTATCC                               | scrR1 (RT & qPCR)             |
| AMO606 | CCAGAAAGAAAGGCCGGCGTTGGA                             | 18S rRNA (NB)                 |
| AMO1377| GAGGAGAACTTCTAGTATATTCTGTATACC                      | RTX1 XUT (NB)                 |
| AMO1482| ATCCGGGCCGCCCTCCACATC                               | scrR1 (NB)                    |
| AMO1509| CCGTTAGACGTTTCAGCTTCCAAACAGAGAATGTGAGAAGCTTCCACTAAG| TY1 small RNAs (NB)           |
| AMO1595| GGGAAAAGTTGTGCTTACTATTCTGCTGTTTAG                   | XUT1678/SUT768 (NB)           |
| AMO1762| GACAGTGTTCGAGATTTTACGA                               | XUT0741 (NB)                  |
| AMO1788| TGAGCGGTACCGAAGCATC                                 | PGK1 mRNA (NB)                |
| AMO2209| CGATAAGGAACATTGTTCTAATATACAC                        | XUT1678 (NB)                  |
| AMO2363| AAATCTTTCGATTTGTATAGAGATTGTTCAGGC                  | snR52 (NB)                    |
| AMO2517| CAAACAAGATCAAATGCGGCTATCCACACATTTTGTTCTCATATATG    | FAR1 small RNAs (NB)          |
| AMO2519| TGCCAAATGGGCGCTACTACACCGGATATCTTTTCACCAAAACACGAT   | FAR1 small RNAs (NB)          |
| AMO2521| GGATTGCTGCCTGCTCGAGCTTTGCTGTTTGCTCCAGATACCTCTTGT   | TAT1 small RNAs (NB)          |
| AMO2522| TTTGCTTCTTGATACCAATTCCCTCGAGATTTGCTTTGGCATG        | TAT1 small RNAs (NB)          |
| AMO2581| CGAATAAGGCTTTGAGCAGCT                          | NEL025C (NB)                  |
| AMO2697| TGGCGTTGAGCTCTCACAAGA                            | XUT1150 (RT)                  |
| AMO2752| TTCAAGTTAAAATTGGGTA                                | XUT1092 (qPCR)                |
| AMO2753| AGTGACATCTGCCGTGATA                                 | XUT1092 (RT & qPCR)           |
| AMO2774| CTCACAGAGATGACCGAACAA                              | XUT1150 (RT anti- & qPCR)     |
| AMO2775| GCCCTTGGCGTGTATTCA                                  | XUT1150 (qPCR)                |
| AMO2776| TCCAGTGATGTGGACGAGAA                               | XUT1186 (qPCR)                |
| AMO2777| AAGCGTTATGAAGACTC                                | XUT1186 (RT & qPCR)           |

(NB) = Northern blot
Supplemental experimental procedures

Yeast strains and plasmids

Yeast strains used in this study are listed in Table S1. Mutant strains were constructed by transformation or standard meiotic crosses. Anti-complementary sequences to solo XUT1092, XUT1150 and XUT1186 were cloned in the pMD2 multi-copy vector (Van Mullem et al., 2003), giving the pAM448, pAM449 and pAM450 plasmids, respectively, and expressed from a constitutive PGK1 promoter. Absence of mutation was confirmed by sequencing of the inserts.

Total RNA extraction

Total RNA was extracted from exponentially growing cells using standard hot phenol procedure. Extracted RNA was ethanol-precipitated, resuspended in nuclease-free H₂O (Ambion) and quantified using a NanoDrop 2000c spectrophotometer and/or a Qubit fluorometer with the Qubit RNA HS Assay kit (Life Technologies). In addition, RNA quality and integrity was checked by Northern blot and/or analysis in a RNA 6000 Pico chip in a 2100 bioanalyzer (Agilent).

Purification of poly(A)^+ RNA

Poly(A)+ RNA was purified from 500 µg of total RNA using the PolyATtract® mRNA Isolation System IV (Promega). Unbound poly(A)^− and purified poly(A)^+ fractions were concentrated using ethanol precipitation, and then resuspended in nuclease-free H₂O (Ambion).

Northern blot

For long RNA detection, 10 µg of total RNA were separated on denaturing 1.2% agarose gel and then transferred to Hybond™-XL nylon membrane (GE Healthcare). For detection of small RNAs by Northern blot, 5 µg of purified <100 nt RNA were separated on 15% TBE-Urea polyacrylamide gel.
After transfer to a Hybond-NX nylon membrane, small RNAs were chemically cross-linked using EDC (Pall and Hamilton, 2008).

$^{32}$P-labelled oligonucleotides (listed in Table S2) 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 the ImageQuant TL v5.2 sofware (GE Healthcare).

**Strand-specific reverse transcription**

Strand-specific reverse transcription reactions were performed from 1 µg of total RNA, using the SuperScript®II Reverse Transcriptase kit (Invitrogen), in the presence of actinomycin D at a final concentration of 6.25 µg/ml. For each sample, a control without RT was also included. Subsequent quantitative real-time PCR were performed in technical duplicates on a LightCycler® 480 instrument (Roche). Oligonucleotides used are listed in Table S2.

**Total RNA-Seq**

Strand-specific total RNA-Seq libraries were prepared using the TruSeq® Stranded Total RNA kit (Illumina), starting from 125 ng of rRNA-depleted RNA (Ribominus Eukaryote System v2) or alternatively from 125 ng of Terminator-treated RNA. In this case, 5 µg of total RNA were treated for 1 hour at 30°C with 1 Unit of Terminator™ 5’-Phosphate-Dependent Exonuclease (Epicentre), followed by phenol/chloroform extraction and ethanol precipitation. For each strain/condition, total RNA-Seq analysis was performed from two biological replicates. Paired-end sequencing of the libraries was performed on a HiSeq 2500 sequencer (Illumina).

Reads were mapped to the *S. cerevisiae* S288C reference genome (12/12/2011 version, retrieved from SGD, http://www.yeastgenome.org/) using version 2.0.6 of the TopHat software (Kim
et al., 2013), with a tolerance of 3 mismatches per read and a maximum size of 5000 bp for introns. Tags densities were normalized on snoRNAs levels for all subsequent analyses.

**ZINAR segmentation**

Segmentation from paired-end RNA-Seq data was performed using an in-house algorithm referred to as ZINAR, inspired from a previously described method (Van Dijk et al., 2011). Briefly, the uniquely mapped reads from the \textit{xrn1}.\textsubscript{Δ} mutants were pooled. A signal was computed in a strand-specific manner for each nucleotide as the number of times it is included in a read or in the insert between two properly paired reads. The signal was transformed in \log_2 value and smoothed using a sliding window (the size of the window was parameterized, ranging from 5 to 200 nt). All genomic regions that showed a smoothed \log_2 signal value above a threshold were reported as segments (this threshold was also parameterized, ranging from 1.44 to 432). XUTs were defined as the ≥200 nt segments that do not overlap ORF, tRNA or sn(o)RNA, that show a 2-fold enrichment at least in the \textit{xrn1}.\textsubscript{Δ} context compared to the wild-type and with a \textit{P}-value ≤0.05 in a differential expression analysis using DESeq (Anders and Huber, 2010). DESeq provides \textit{P}-values by comparing the two biological replicates for each condition independently, thereby taking into account the variability between replicates. In total, 3880 segmentations with different sliding window size and threshold parameters were tested in parallel, among which we arbitrarily selected the one showing the best compromise between SUT and ORF detection (Figure S7). Parameters corresponding to the selected segmentation were: sliding window size = 5 nt; threshold = 146.88 tag/nt. Importantly, we controlled that all XUTs showed an expression level higher than the common arbitrary threshold of 1 FPKM (Fragment Per Kilobase per Million mapped reads) in each of the two \textit{xrn1}.\textsubscript{Δ} replicates, in at least one of the laboratory strain in which they were annotated. Finally, a XUT was reported as antisense when the overlap with the sense ORF was ≥1 nucleotide (the region of overlap being actually larger since the 5′- or 3′-UTR of the sense mRNA are not considered in the definition).
CAGE-Seq

Construction of CAGE-Seq libraries was performed as described (Takahashi et al., 2012), starting from 5 µg of total RNA extracted from two biological replicates of wild-type (YAM1) and dcp2-7 (YAM2283) cells shifted for 2 hours at 37°C. Libraries single-read sequencing (40 nt read length) was performed on a Genome Analyzer sequencer (Illumina). After demultiplexing using an in-house script, reads were filtered to retain only those containing the EcoP15I restriction site and with a G as a first base of the insert. Adapter sequences were then removed using version 1.2.1 of cutadapt. Finally, reads with 25 to 27 nt insert were selected (expected read length was 26 nt) and mapped to the reference genome using version 0.12.7 of Bowtie (Langmead et al., 2009), with a tolerance of 3 mismatches. Globally, a total of 21 to 28 million of reads were sequenced per library, resulting in 3 to 4 million of uniquely mapped 25-27 nt reads used for subsequent analysis.

Peak-calling

To identify transcription start sites (TSS) using CAGE signals, we developed an in-house peak-calling algorithm. Briefly, the CAGE-Seq signal was scanned within a +/- 200 nt window around the nucleotide of the genome. Two parameters were used to define peaks within the window: for each position of the window, the minimum number a uniquely mapped reads starting at the considered position, in each of the two biological replicates (1, 2, 5, 10, 15 or 20), and the minimum ratio between the signal at this position and the total signal within the window (0.05, 0.1, 0.15 or 0.2). The window was then slided 10, 25 or 50 nt downstream (therefore centered on nucleotide n+10, n+25 or n+50, respectively). Consequently, each position of the genome was encompassed in several sliding windows. In total, we tested 72 combinations of parameters for the wild-type and the dcp2-7 conditions. Parameters selected for the two conditions were: 10 nt interval between two sliding windows, reads/replicate ≥5, signal/total ratio ≥0.15. Finally, a position was defined as a genuine CAGE peak when it passed these criteria for ≥ 90% of the windows encompassing it.
The peaks identified using these parameters were compared with annotated TSS of mRNAs and SUTs for which RNA-Seq signal in the dcp2-7 mutant was \( \geq 1 \) RPKM (Read Per Kilobase per Million mapped reads), which was the threshold under which a transcript was considered as not expressed. We noted that 69% of the peaks detected in the dcp2-7 context were located within the +/- 100 nt surrounding the annotated TSS of these mRNAs and SUTs (annotation of mRNAs was retrieved from the reference genome R63 and imported into our reference genome using Blastn alignments).

For re-annotation of XUTs TSS, we only considered the 1414 XUTs with total RNA-Seq signal \( \geq 1 \) RPKM in the dcp2-7 context. For 625 of them, we detected at least 1 CAGE peak in the vicinity (+/- 100 nt) of the TSS predicted by the segmentation. Consequently, TSS of these XUTs was re-annotated according to the position of the CAGE peak (for XUTs with more than 1 peak, we used the position giving the largest transcript). In addition, there were 15 pairs of tandem XUTs, separated by less than 200 nt, with variation of RNA-Seq signal (in RPKM) \( \leq 25\% \), and for which the peak-calling only detected CAGE-Seq peak(s) for the first XUT. In these cases, the two tandem XUTs were fused into a single, large XUT.

**Comparative analysis of CAGE-Seq and TSS sequencing signals**

Processed wig files containing both positions and normalized signals for TSS identified by TSS sequencing in WT and upf1-1 conditions (Malabat et al., 2015) were retrieved from NCBI GEO database (accession number GSE64139). After RPM normalization, efficiency of CAGE-Seq and TSS sequencing was compared using metagene representations of the +/- 100 nt region around annotated 5’ end of mRNAs, SUTs and XUTs.

**Small RNA-Seq**

Small-RNA libraries were already described (Sinturel et al., 2015). Briefly, they were constructed according to the “Small RNA Sample Preparation Guide” (Illumina), using 10-40 nt small RNAs purified from total RNA with the flashPAGE Fractionator (Ambion) or on 15% TBE-urea
polyacrylamide gel. Qualitative analysis of the purified small RNAs on a Small RNA chip in a 2100 bioanalyzer (Agilent) validated the size selection.

Single-end sequencing (40 nt) of libraries was performed on a Genome Analyzer Ilx (Illumina). Adapter sequences were removed using cutadapt. Reads were then uniquely mapped to the S288C reference genome using the version 0.12.7 of Bowtie (Langmead et al., 2009), with a tolerance of 3 mismatches.

Identification of XUTs showing ribosome footprints

Ribosome profiling and control fragmented RNA raw data were retrieved from the NCBI BioProject database (accession number PRJNA245106) and analyzed as previously described (Smith et al., 2014), with minor modifications. After trimming of the adapter sequence and size selection, reads ≥ 25 nt were mapped using version 0.12.8 of Bowtie (Langmead et al., 2009), with a tolerance of 1 mismatch. Subsequent analysis used uniquely mapped reads only. After FPKM normalization, transcripts with FPKM > 0 in the WT strain and FPKM ≥ 10 in the upf1Δ mutant were filtered out. The footprinting score was calculated for each XUT as the ratio FPKM_{ribosome-footprint}/FPKM_{fragmented-RNA}, as described (Smith et al., 2014). The 275 XUTs showing footprinting score > 0 in the WT and > 0.1 in upf1Δ were selected for further analysis.

Metagenes were produced as described (Sinturel et al., 2015). Signal corresponds to the mean coverage for each virtual nucleotide, normalized on the total number of uniquely mapped reads.

Analysis of polysome-Seq signals

Polysome-Seq and control RNA-Seq raw data were retrieved from NCBI BioProject accession number PRJNA245106 and analyzed as previously described (Smith et al., 2014), with minor modifications. After trimming the last nucleotide at their right end, reads were mapped using version 0.12.8 of Bowtie (Langmead et al., 2009), with a tolerance of 2 mismatches. Subsequent analysis
used uniquely mapped reads only. Transcripts showing an average expression of FPKM < 10 in the 
upf1Δ control RNA-Seq dataset were excluded from the analysis.
Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol 11, R106.

Cloutier, S.C., Ma, W.K., Nguyen, L.T., and Tran, E.J. (2012). The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. J Biol Chem 287, 26155-26166.

Descrimes, M., Zouari, Y.B., Wery, M., Legendre, R., Gautheret, D., and Morillon, A. (2015). VING: a software for visualization of deep sequencing signals. BMC Res Notes 8, 419.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in Budding Yeast. Science 326, 544-550.

Egecioglu, D.E., Henras, A.K., and Chanfreau, G.F. (2006). Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. RNA 12, 26-32.

Geisler, S., Lojek, L., Khalil, A.M., Baker, K.E., and Coller, J. (2012). Decapping of long noncoding RNAs regulates inducible genes. Mol Cell 45, 279-291.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. Mol Cell 12, 1439-1452.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25.

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.

Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. Nat Protoc 3, 1077-1084.

Sinturel, F., Navickas, A., Wery, M., Descrimes, M., Morillon, A., Torcher, C., and Benard, L. (2015). Cytoplasmic Control of Sense-Antisense mRNA Pairs. Cell Rep 12, 1853-1864.

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 Rep 7, 1858-1866.

Takahashi, H., Lassmann, T., Murata, M., and Carninci, P. (2012). 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. Nat Protoc 7, 542-561.

Thiebaut, M., Kisseleva-Romanova, E., Rougemeille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell 23, 853-864.

Tuck, A.C., and Tollervey, D. (2013). A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and IncRNAs. Cell 154, 996-1009.

Van Mullem, V., Wery, M., De Bolle, X., and Vandenhauwe, J. (2003). Construction of a set of Saccharomyces cerevisiae vectors designed for recombinational cloning. Yeast 20, 739-746.

Wilson, M.A., Meaux, S., and van Hoof, A. (2007). A genomic screen in yeast reveals novel aspects of nonstop mRNA metabolism. Genetics 177, 773-784.